

Experimentell design inom samhällsvetenskapen

De senaste decennierna har det blivit allt vanligare inom samhällsvetenskaplig forskning att använda sig av experiment. Utvecklingen har varit så snabb att vissa beskrivit det som en revolution. Tydligast har trenden varit inom nationalekonomin och i statsvetenskapliga topptidskrifter, men olika sorters experimentell design har även tagit plats i utredningsverksamhet och i den politiska debatten. Allt oftare sägs det att politiska reformer måste vara *utvärderingsbara*, vilket i regel implicerar ett genomförande som påminner om experimentets ideal.

Den här texten är avsedd som en ögonöppnare och inspirationskälla. Förhoppningsvis kan den på så vis fylla tre olika syften. För det första är det viktigt att bekanta sig med experimenttrenden, både för att kunna inspireras och för att förhålla sig till den. Den är helt enkelt för viktig för att inte ha en uppfattning om. För det andra kan den logik och det språkbruk som förekommer i experimentlitteraturen underlätta och fördjupa förståelsen av andra sorters förklaringar vi diskuterat på kursen. Det ideala experimentet är då en användbar referenspunkt att jämföra sig med. För det tredje är det fullt möjligt att använda en experimentell design i en C-uppsats. Om den här texten kan så ett litet frö, vilket i något fall mynnar ut i ett experiment, så är mycket vunnet.

Texten innehåller många exempel på både kontrollerade och naturliga experiment. Med kontrollerade experiment avses situationer där forskaren kontrollerar den behandling vars effekt vi studerar. Naturliga experiment syftar på experimentliknande situationer som uppstått utanför forskarens kontroll. Några exempel är hämtade från berömda forskningsartiklar. Andra är kanske mer lagom för vad som kan åstadkommas i en uppsats på kandidat- eller masternivå. Exemplena härrör från vitt skilda områden och berör kvoter, beteendepsykologi, kärnvapen, inbördeskrig, valforskning och mycket annat. Framställningen är intuitiv snarare än teknisk.

Inledningsvis beskriver avsnittet *Experimentets logik* idén bakom experiment och vilka problem de löser. Det är den viktigaste delen av texten. De mest omfattande delarna av texten är avsnitten om kontrollerade och

naturliga experiment, vilka tar upp de viktigaste typerna av experiment samt lämnar ett stort antal exempel från tidigare forskning. Alla exempel bör ses som illustrativa och vi förväntar oss inte att ni memorerar dem. Texten avslutas med ett stycke om den kritik som brukar riktas mot experiment och några ord om hur experiment förhåller sig till andra typer av studier.

För en mer utförlig men fortfarande lättförståelig genomgång av naturliga experiment rekommenderas Thad Dunnings bok *Natural Experiments in the Social Sciences*.

Experimentets logik

Det fundamentala orsaksproblemet

Hur vet vi att vi har kontrollerat för alla relevanta variabler och att dessa kontrollvariabler är korrekt mätta och definierade? Under vilka förutsättningar kan vi utesluta att den beroende variabeln påverkar den oberoende variabeln? Och när kan vi vara säkra på hur lång tid det tar för en effekt att träda i kraft?

Det tråkiga svaret på dessa frågor är förstas att så sällan är fallet. Oavsett antalet fall vi studerar, kan vi nästan vara säkra på att det finns variabler vi inte observerat som påverkar de övriga variablerna. Ofta är det dessutom så att den beroende variabeln påverkar de oberoende variablerna, vilket ytterligare snedvrider resultaten. Det här är inga nya insikter. Redan John Stuart Mill ansåg att hans *Method of Difference* endast gav pålitliga resultat när en variabel skiljde sig åt mellan de två fallen på grund av att den manipulerats av forskaren – genom experiment – eftersom det endast är då vi kan veta att fallen är lika i alla andra avseenden. Samtidigt insåg Mill att inte heller experiment är tillräckliga om de stannar vid att vara en simpel effektstudie. En förklaring är inte fullständig förrän vi vet hur och varför en variabel påverkar en annan.

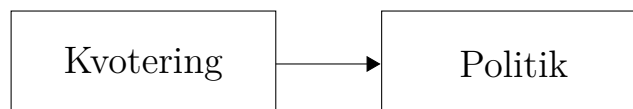
But if we can not artificially produce the phenomenon A, the conclusion that it is the cause of a remains subject to very considerable doubt. Though an invariable, it may not be the unconditional antecedent of A, but may precede it as day precedes night or night day. This uncertainty arises from the impossibility of assuring ourselves that A is the only immediate antecedent common to both the instances. If we could be certain of having ascertained all the invariable antecedents, we might be sure that the unconditional invariable antecedent, or cause, must be found somewhere among them. Unfortunately it is hardly ever possible to ascertain all the antecedents, unless the phenomenon is one which we can produce artificially. Even then, the difficulty is merely lightened, not removed: men knew how to raise water in

pumps long before they adverted to what was really the operating circumstance in the means they employed, namely, the pressure of the atmosphere on the open surface of the water.
- J.S. Mill (1843), p. 481.

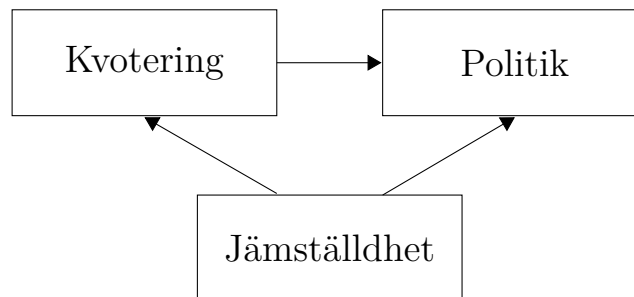
Många av de problem vi stöter på när vi använder icke-experimentell data handlar om att vi har svårt att belägga ett av de fyra orsaksriterierna – isolering. I en multivariat regression försöker vi åstadkomma isolering genom att inkludera relevanta kontrollvariabler, men det finns flera skäl för varför den metoden kan vara otillräcklig. Vi kan sammanfatta varför i tre punkter.

- Vi vet inte vilka alla de möjliga bakomliggande förklaringarna är.
- Även om vi visste det, är det inte självklart hur vi ska mäta dem eller att det är praktiskt möjligt.
- Även om vi kände till och kunde mäta alla bakomliggande variabler, vet vi inte hur vi ska kontrollera för dem. Den linjära och additiva regressionskvationen är bara en möjlighet.

Låt oss ta ett konkret exempel för att illustrera problemet. Många av er har säkert läst *Närvarons politik* av Anne Phillips, vilken handlar om betydelsen av politisk representation. Säg att vi vill studera om kvotering av kvinnor till politiska parlament påverkar politikens innehåll. Ett tillvägagångssätt skulle kunna vara att jämföra politiken i länder som har kvotering med länder som inte har kvotering.



Risken för att fånga spuriösa samband är uppenbar. För det första är det svårt att veta vad vi måste kontrollera för. Jämställdhet? Kvinnors politiska makt? Väljarnas politiska preferenser? Biståndsprojekt som kräver jämn könsfördelning? För det andra, om vi visste, hur kan vi mäta något så abstrakt som kvinnors politiska makt, utan att riskera låg reliabilitet och validitet? För det tredje, om vi kunde mäta, hur vet vi hur sambandet ser ut? Kan vi verkligen anta att alla effekter är *linjära* och *additiva* (oberoende av varandra)?



Idén bakom experiment är att om värdena på x är helt slumpmässiga, som om de vore resultatet av ett tärningsslag, då kan vi utesluta möjligheten att andra variabler har påverkat den oberoende variabeln. Och om ingenting relevant kan påverka den oberoende variabeln, då finns det ingen risk att de samband vi hittar beror på omvänd orsaksriktning (att den beroende variabeln påverkar den oberoende variabeln) eller är spuriösa (andra variabler påverkar både den oberoende och den beroende variabeln).

För att fortsätta med vårt exempel, om vi hade slagit en tärning för att bestämma vilka av världens länder som tillämpar könskvotering, då skulle vi inte ha nyss nämnda isoleringsproblem. Det hade ju varit tärningen som tilldelat kvotering och inte kvinnors politiska makt eller väljarnas politiska preferenser. Det hade inte heller funnits någon risk att den förda politiken påverkade sannolikheten för att kvotering införs (omvänd orsaksriktning).

Världen utvecklas förstås inte via tärningsslag, utan genom sociala och politiska processer. Men ibland ger dessa processer upphov till situationer som i centrala avseenden påminner om experiment. Efter en konstitutionell reform måste ordföranden i en tredjedel av byråden i Indien sedan 1993 vara en kvinna. Vilka byar det gäller är slumpmässigt bestämt. En studie av Chattopadhyay och Duflo (2004) använder det till att studera effekten av kvotering på politikens innehåll. De finner att byar med kvotering investerar mer i politiska projekt som är populära bland kvinnor, såsom tillgång på dricksvatten och olika välfärdsprogram (änkepension, föräldraförsäkring, etc).

Kontrollerade experiment

Det är vanligt att dela upp experimentliknande studier i kontrollerade och naturliga experiment. Ett kontrollerat experiment utformas av forskaren själv vilket ger denna kontroll över vilka analysenheter som utsätts för en behandling (treatment) och hur tilldelningen (assignment) av denna går till. Här har de kontrollerade experimenten delats upp i tre kategorier: laboratorieexperiment, fältexperiment och surveyexperiment.

Laboratorieexperiment

När experimentdeltagarna uppmanas att infinna sig på en plats som är under kontroll av forskaren brukar vi kalla det för ett laboratorieexperiment. Deltagarna vet därför om att de deltar i ett experiment, men däremot vet de kanske inte de vad experimentet går ut på eller vilken del av deras beteende som studeras.

Inom samhällsvetenskapen har det framför allt varit psykologer som genomfört laboratorieexperiment, men de har med tiden blivit vanligare även bland andra samhällsvetare. De vanligaste användningsområdena är att testa antaganden om mänskligt agerande och hypoteser från spelteoretiska resonemang. Trots namnet utförs samhällsvetenskapliga laboratorieexperiment mycket sällan i ett klassiskt kemilabb.

Laboratorieexperimentets styrkor ligger i den fullständiga kontrollen över experimentsituationen. Svagheter i laboratorieexperiment handlar i regel om deras generaliserbarhet. Det allvarligaste problemet i det här avseendet är att det i regel är ett stort steg mellan den uppbyggda experimentsituationen och de verkliga fenomen vi vill uttala oss om. Bara en sådan sak som att deltagarna vet om att de deltar i ett experiment kan radikalt förändra deras beteende. Dessutom är deltagarna ofta få och utgörs inte sällan av universitetets egna studenter.

Anhängare av laboratorieexperiment argumenterar för att fältexperiment (se nedan) inte har större möjligheter att generalisera resultaten, annat än till just den kontext där experimentet ägde rum. De påpekar också gärna att den kritik som riktas mot laboratorieexperiment ofta kan testas med just ytterligare laboratorieexperiment. För ett utförligt försvar av laboratorieexperiment, se Falk och Heckman (2009). En mer kritisk genomgång återfinns i Levitt och List (2007). Nedan ges exempel på kända laboratorieexperiment.

Ultimatumspelet

Ett klassiskt experiment är det så kallade ultimatumspelet, vilket går ut på att två spelare ska fördela en summa pengar emellan sig. Den första spelaren föreslår en fördelning och det är därefter upp till den andre spelaren att godkänna fördelningen. Om den andre spelaren godkänner fördelningen erhåller båda spelarna den föreslagna summan. Om den andre spelaren i stället förkastar fördelningen får ingen av spelarna några pengar alls. Eftersom spelet bara spelas en gång finns det inga skäl att samarbeta inför framtida spel.

Det har visat sig att den andre spelaren ofta förkastar låga bud, trots att det innebär att spelaren inte får några pengar alls i stället för i alla fall en liten summa. Detta brukar tolkas som att människor följer normer om rättvisa och att de bestraffar personer som inte följer dessa normer, även om bestraffningen är kostsam för dem själva.

Milgrams lydnadsexperiment

Ett av de bland samhällsvetare mest berömda laboratorieexperimenten utfördes av psykologen Stanley Milgram (Milgram 1963). Han ville förstå hur så många till synes friska och normala människor kunde utföra mord, tortyr och andra fruktansvärda övergrepp i Nazityskland. Var det möjligt att alla dessa människor helt enkelt bara lydde order från sina överordnade? Hur mycket smärta är vi beredda att åsamka en medmänniska, om vi beordras av en person med stor auktoritet?

Försökspersonerna i Milgrams experiment tilldelades rollen som lärare. Bakom en vägg trodde försökspersonerna att det satt en annan experimentdeltagare i rollen som elev, vilken de träffat under introduktionen till experimentet. Försökspersonerna fick sedan läsa upp ord från en lista för eleven. Om eleven inte kände till det andra ordet i ordparet, uppmanades försökspersonen att ge eleven en elstöt. För varje stöt ökade styrkan, upp till den maximala stöten på 450 volt. På den andra sidan väggen stod i själva verket en bandspelare som reagerade på stötarna med skrik, klagande över hjärtproblem och bedjan om att experimentet skulle avslutas.

Psykologer och psykologstudenter tillfrågades om hur stor andel av deltagarna som skulle ge den maximala stöten på 450 volt. De utgick från andelen sadister i befolkningen och svarade att 1–2 procent av deltagarna skulle göra det. Det visade sig att 26 av 40 deltagare utdelade den maximala stöten. Alla deltagare gick upp till åtminstone 300 volt.

Fältexperiment

Ett fältexperiment äger i stället rum i en naturlig miljö för analysenheterna. Även här är experimentsituationen under forskarens kontroll, men i regel är personerna som observeras inte medvetna om att de deltar i ett experiment. Ovetande försökspersoner och en naturlig experimentmiljö gör det enklare att generalisera resultaten till verkliga situationer.

A fine is a price

De flesta som någon gång arbetat på en förskola är bekanta med problemet att föräldrar hämtar sina barn alldeles för sent. Vid en förskola i Israel tröttnade man på beteendet och bestämde sig för att tillsammans med forskarna Gneezy och Rustichini (2000) utforma ett bestraffningssystem för föräldrar som hämtade sent. För att minska antalet sena hämtningar infördes en böter på 10 israeliska shekels per barn, vilket motsvarar ungefär en tredjedels timlön för en person med medelinkomst.

Följden av detta blev något överraskande att antalet sena hämtningar ökade. Forskarna tolkade resultatet som ett bevis på att ekonomiska incitament kan undantränga andra drivkrafter, så kallad *motivation crowding*.

I detta fall sker detta genom att bötesbeloppet uppfattas som ett pris för att hämta sent, vilket gör att föräldrarna inte längre behöver skämmas eller ta hänsyn till förskolans personal när de kommer för sent. Föräldrarna har trots allt betalat för sig.

Dörr till dörr

Anta att vi vill studera huruvida dörrknackning och andra former av politiskt kampanjande bidrar till ökat valdeltagande. En möjlighet är att fråga partierna om var de har knackat dörr och använda den informationen som oberoende variabel i en regressionsanalys av valdeltagande. Problemet med ett sådant tillvägagångssätt skulle vara den uppenbara risken för såväl spuriösa samband som omvänd orsaksriktning. Hur vet vi att partierna inte knackar dörr där valdeltagandet redan är högt – för att inte slösa tid på personer som inte röstar, eller där det är lågt – för att höja det. Det finns också en risk att partiernas andra skäl för att knacka där de gör, såsom var partimedlemmarna bor eller var de förväntar sig många röster, också är korrelerade med den beroende variabeln. Därför skulle det vara svårt att veta om skillnader mellan områden där det har kampanjats och områden där partiet inte har bedrivit någon valkampanj beror på kampanjen eller att på områdena var olika redan innan kampanjen inleddes.

För att komma runt det problemet har det blivit populärt att använda experiment för att utvärdera effekterna av politiskt kampanjande. De flesta studier har genomförts i USA, men inför valet till Europaparlamentsvalet 2014 utförde jag ett experiment i samarbete med Socialdemokraterna i Södermanland.

Jag slumpade 7 500 hushåll till behandlings- respektive kontrollgrupp. Socialdemokraterna knackade dörr i behandlingsgruppen och lämnade alla hushåll i kontrollgruppen i fred. Efter att ha digitaliserat röstlängderna och jämfört de två grupperna såg vi att dörrknackandet hade höjt valdeltagandet med ca tre procentenheter. Eftersom tilldelningen av behandlingen var slumpmässig kunde vi vara säkra på att en systematisk skillnad mellan grupperna endast kunde bero på dörrknackandet. Grupperna kan förstås också skilja sig åt på grund av slumpen, men det kan vi enkelt hantera genom att beräkna konfidensintervall eller genomföra hypotestest.

Enkätexperiment

Att alternera frågeformuleringarna eller bakgrundsinformationen i en enkät är ett av de enklaste tillvägagångssätten för att utföra ett experiment. Om tilldelningen av olika formuleringar är slumpmässig, kan den inte påverkas av någon bakomliggande variabel. Skillnader i utfall kan därför bara bero på skillnader i frågeformulering (eller annan 'treatment').

Enkätexperimentens största nackdel är att det man uppger i en enkät

inte nödvändigtvis återspeglar ens verkliga attityder eller handlande. Därför bör man vara mycket försiktig med att tolka storleken på olika effekter i enkätexperiment.

Det vanligaste användningsområdet för enkätexperiment är att utvärdera och analysera enkätmetoder. Exempelvis kan vi testa hur intervjuarens kön, ersättning till respondenten eller ordningen på enkätfrågorna påverkar de svar vi får.

Inställning till kärnvapen

Människors inställning till kärnvapen tycks vara väldigt kontextberoende. I ett enkätexperiment lät Press m. fl. (2013) respondenterna läsa en påhittad nyhetsartikel om en konflikt mellan USA och Syrien. Av etiska skäl valde forskarna att informera respondenterna om att nyheten var falsk.

I nyheten beskrevs att al-Qaida har en fabrik i syrien där de producerar kärnvapen och att den amerikanska militären väljer mellan att använda kärnvapen eller konventionella vapen för att slå ut fabriken. I experimentet varierades den information som respondenterna fick i två avseenden. Dels beskrevs effektiviteten av kärnvapen jämfört med konventionella vapen på olika sätt. Dels beskrevs den amerikanska attacken ibland som att den redan hade ägt rum.

Resultaten visade för det första att respondenterna blev positiva till användandet av kärnvapen när de beskrevs som effektivare än andra vapen. Det tolkade författarna som att amerikanernas motstånd till kärnvapen bör förstås som ett konsekvensialistiskt övervägande och inte som ett tabu. De fann också att respondenterna var mer positiva till kärnvapen när beskrivningen handlade om ett anfall som redan skett än ett anfall som planerades. Det skulle kunna tolkas som att en regering kan räkna med att befolkningen kommer att sluta upp bakom deras agerande, även om det på förhand finns en kritisk opinion.

Unmatched count

Inte ens i anonyma enkäter kan man räkna med att respondenterna svarar ärligt på frågor som är känsliga eller kontroversiella. För att komma till rätta med det problemet kan man använda metoder som gör det omöjligt att identifiera vad en enskild respondent har svarat, men där vi kan räkna ut vad personerna har svarat i genomsnitt. Forskning har visat att många blir mer benägna att svara sanningsenligt med den typen av frågor jämfört med när man ställer en direkt fråga, förutsatt att frågan är känslig.

En av dessa metoder kallas *unmatched count*. Metoden går ut på att respondenterna delas in i två grupper. Personerna i den första gruppen får en lista med påståenden och ombeds ange hur stort antal av påståendena som stämmer. Eftersom respondenterna i den andra gruppen får samma

lista, men där ett av alternativen har plockats bort, kan vi genom att jämföra svaren i de två grupperna uppskatta hur stor andel som stämmer i det bortplockade påståendet. Vår bästa uppskattning av hur många som instämmer i det alternativ som endast fanns på den ena gruppens lista, är helt enkelt skillnaden i medelvärde mellan de två grupperna. Om hälften av respondenterna instämmer i påståendet, så kommer den första gruppen i genomsnitt att ange ett halvt påstående mer. Däremot kan vi inte svara på vilka individer som instämde i påståendet (undantaget de som anser att alla påståenden stämmer).

Eftersom det kan finnas slumpmässiga skillnader i hur många som instämde i de fyra första påståendena, måste vi beräkna ett konfidensintervall för skillnaden. Det konfidensintervallet anger hur stor andel som instämmer i det bortplockade påståendet. Eftersom vi jämför antalet påståenden använder vi formeln för konfidensintervall för skillnader mellan 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.

$$\bar{x}_1 - \bar{x}_2 \pm t_{kv} \sqrt{\frac{s_{x_1}^2}{n_{x_1}} + \frac{s_{x_2}^2}{n_{x_2}}} \quad (1)$$

Ekonomern Katherine Coffman (Coffman m. fl. 2013) använde en variant av denna metod för att studera hur stor andel av invånarna i USA som är homosexuella och jämföra det med hur stor andel som uppgav det när de fick en direkt fråga. Skillnaden i storlek mellan de två grupperna var skrämmande. Med den första metoden skattades andelen homosexuella till 19 procent. När hon undersökte samma sak med en direkt fråga var det endast 11 procent som svarade ja, trots att enkäten var anonym.

Naturliga experiment

Ofta är det av praktiska eller etiska skäl inte önskvärt att genomföra den typen av randomisering som kontrollerade experiment förutsätter. I ett naturligt experiment är varken behandlingen eller anvisningen under kontroll av forskaren. Istället används uppkomna situationer som påminner om kontrollerade experiment därför att tilldelningen av en behandling är slumpmässig eller av olika skäl kan betraktas slumpmässig. Naturliga experiment är i strikt mening därför inte experiment (många avser endast kontrollerade experiment när de pratar om experiment).

Det här avsnittet är indelat i tre underavsnitt. Det första behandlar traditionella naturliga experiment, vilket avser situationer där det har förekommit en medvetet slumpmässig tilldelning av en behandling eller där tilldelningen kan betraktas som slumpmässig. Exempel på detta är lotterier och vissa politiska reformer. Den andra handlar om tröskelvärden, eller så kallade

regression discontinuity designs. Dessa upplägg bygger på att det finns en skarp gräns, sådan att analysenheter på den ena sidan gränsen utsetts för en behandling som inte analysenheterna på den andra sidan gränsen får. Om det inte är möjligt för analysenheterna att påverka huruvida de placerar sig precis ovanför eller nedanför tröskelvärdet, kan allokeringen kring tröskeln – och därmed tilldelningen av behandlingen – betraktas som slumpmässig. Den sista delen handlar om hur man kan använda en variabel som instrument för en annan variabel och varför det kan ibland liknas vid ett experiment. Ett bra instrument ska påverka den oberoende variabel vi är intresserade av, men utan att kunna påverkas av den beroende variabeln eller av någon bakomliggande variabel som kan påverka både instrumentet och den beroende variabeln. Dessutom ska instrumentet endast påverka den beroende variabeln genom sin effekt på den oberoende variabeln.

Traditionella naturliga experiment

Med traditionella naturliga experiment avses situationer där det har förekommit en medvetet slumpmässig tilldelning av en behandling eller där tilldelningen kan betraktas som slumpmässig. Det kan handla om reformer som kan betraktas som slumpmässiga, exempelvis för att de har införts vid olika tidpunkter på olika ställen och där det kan anses slumpmässigt när reformen har införts, eller någonting så genuint slumpmässigt som att vinna ett lotteri.

Lotterivinster

Många statsvetare har undersökt hur inkomst påverkar allt från hälsa och välbefinnande till värderingar och politiskt deltagande, men det är svårt att isolera effekten av inkomst från andra faktorer som har med social status att göra. Och det vore absurt om vi började dela ut pengar till slumpmässigt valda personer, bara för att studera hur förmögenhet påverkar dessa utfall. Eller?

Lotterier är ett tydligt exempel på naturliga experiment med klassisk randomisering. Det är slumpmässigt vem som vinner i ett lotteri, men alla har inte samma chans att vinna. För att hitta en relevant referenspunkt jämför Doherty m. fl. (2006) personer som vunnit stora summor på lotteri med personer som spelat lika mycket men utan att vinna. De finner att lotterivinnarna är mer negativt inställda till fastighetsskatt och omfördelning, vilket visar att effekten av förmögenhet på politiska attityder åtminstone delvis är kausal. Andra studier har funnit att lotterivinster påverkar hälsa (Lindahl 2005) och välbefinnande (Gardner och Oswald 2007).

Vietnamlotteriet

Det mest välstuderade lotteriet är förmodligen det så kallade Vietnamlotteriet. År 1969 drogs i USA alla årets 366 dagar och rangordnades från 1 till 366. Personer som var födda på en dag med ett lågt nummer blev inkallade att tjänstgöra i Vietnamkriget, medan personer med höga lotterinummer slapp tjänstgöring i Vietnam.

Eftersom lotteriet gav upphov till en slumpmässig variation i vilka som deltog i kriget och vilka som inte gjorde det, har experimentet givit upphov till omfattande forskning. Ta en fråga som hur ens framtida inkomster påverkas av att delta i krig. Om vi jämför inkomsterna för personer som har deltagit i krig och de som inte har det, har vi i allmänhet svårt att separera effekten av kriget från effekten av alla de faktorer som gjorde att man tog värvning.

Trösklar

Så kallad Regression Discontinuity Design (RDD eller bara RD) utgår från förekomsten av en skarp gräns, sådan att analysenheter på den ena sidan gränsen utsetts för en behandling som inte analysenheterna på den andra sidan gränsen får. Om det inte är möjligt för analysenheterna att påverka huruvida de placerar sig precis ovanför eller nedanför tröskelvärde, kan allokeringen kring tröskeln – och därmed tilldelningen av behandlingen – betraktas som slumpmässig. Typiska trösklar är gränsen för att bli invald i en församling (jämför precis inkryssad med precis utanför), åldern för att få rösta (jämför personer födda dagen före med personer födda efter valdagen) och kommunstorlek som påverkar valsystemet (Sveriges införande av representativ demokrati).

Ett centralt antagande i varje RD är att fördelningen kring tröskelvärde kan betraktas som slumpmässig, så att observationerna precis ovanför tröskelvärde inte skiljer sig systematiskt från observationerna nedanför tröskelvärde, i något annat avseende än behandlingen. Om det är möjligt att påverka huruvida man kommer precis över eller under tröskeln, håller inte antagandet och vi kan inte längre behandla tilldelningen av behandlingen som slumpmässig.

Vinner regeringar val?

En intressant fråga för statsvetare är huruvida den sittande regeringen har en fördel i nästkommande val. Får dessa partier fler röster än vad de skulle få om de inte stött i regeringsställning? En metod för att besvara den frågan skulle vara att jämföra hur ofta en regering blir omvald med hur ofta oppositionen vinner valet. Den metoden skulle dock ha flera svårlösliga problem. Kanske blir regeringspartier omvalda för att de helt enkelt är populärare än oppositionspartierna? Om partier har olika popularitet skulle vi

ju förvänta oss att de mer populära partierna oftare sitter i regeringsställning.

Titunik (2011) studerar denna fråga genom att jämföra de partier som vann med de partier som förlorade i de brasilianska kommunval där utgången var jämn och ovisst. Hennes utgångspunkt är att det inte finns några systematiska skillnader mellan de partier som vann och de partier som förlorade, eftersom valen var så ojämna och det kan betraktas som slumpmässigt vilken sida som segrade. Titunik finner att partier som vann valet – och därmed styrde kommunen under efterföljande mandatperioden – presterade sämre än de partier som i stället utgjorde oppositionen. Effekten var stor och varierade mellan 4 och 19 procent av rösterna beroende på vilket parti som studerades.

Är kårpolitik meriterande?

Pär Zetterberg, verksam vid statsvetenskapliga institutionen i Uppsala, har tillsammans med två andra forskare studerat hur sannolikheten att kandidera till kommunfullmäktige påverkas av att under tiden som student ha suttit i kårfullmäktige (Lundin m. fl. 2013). För att lösa problemet med att personer som är aktiva i kårpolitiken redan innan är mer politiskt engagerade än personer som inte är det, jämför de personer som precis blev invalda med personer som var precis under strecket för att bli invalda. Dessa två grupper borde vara närmast identiska, med den skillnaden att den ena gruppen blev invald i fullmäktige medan den andra gruppen inte blev det. Författarna finner att personer som blev invalda har en större chans att få ett arbete inom tre år och även högre inkomster i genomsnitt. Effekten är dock övergående och inte permanent. Det skulle kunna tolkas som att kårengagemanget fungerar som ett substitut för arbetslivserfarenhet.

Väder och andra instrument

Den typ av experiment som är svårast att förstå är förmodligen instrumentvariabler. Vi vill ofta studera ett samband mellan variablerna x och y , men där vi misstänker att det kan finnas andra variabler som påverkar både x och y och därför snedvrider det samband vi observerar. En instrumentvariabel är då en variabel z som har en stor effekt på x och inget samband med y utöver den effekt som går via x . Vi kan då använda z för att göra prediktioner av x . Eftersom variationen i z inte påverkas av några problematiska bakomliggande variabler kommer inte heller våra prediktioner av x att göra det. Vi kan därför använda prediktionerna i våra regressioner av y utan risk för spuriösa samband eller samband med omvänd orsaksriktning. Metoden kan beskrivas i följande två ekvationer.

$$\begin{aligned}\hat{y} &= a_1 + b_1\hat{x} \\ \hat{x} &= a_2 + b_2z\end{aligned}\tag{2}$$

Väder som tillväxt

Låt oss anta att vi vill studera huruvida ekonomisk tillväxt påverkar risken för inbördeskrig. Ett uppenbart problem är att sambandet förmodligen går åt båda hållen, eftersom inbördeskrig förmodligen påverkar den ekonomiska tillväxten. Om vi skulle genomföra en regressionsanalys med inbördeskrig som beroende variabel och tillväxt som oberoende skulle b-värdet fånga båda effekterna och vi skulle därför inte kunna säga något om den effekt vi var intresserade av. Dessutom skulle det föreligga en stor risk för spuriösa samband, där andra variabler påverkar både tillväxt och sannolikheten för inbördeskrig.

Miguel m. fl. (2004) använder variationen i nederbörd som ett instrument för att predikera tillväxt. De studerar 41 afrikanska länder där regn är avgörande för jordbruket. Under de år som nederbörden är låg blir även tillväxten låg eftersom skördarna blir sämre när det råder torra.

Det fina med väder är att variationen i nederbörd kan betraktas som slumpmässig. Eftersom förekomsten av inbördeskrig inte kan påverka mängden nederbörd kan kriget inte heller påverka den del av variationen i BNP-tillväxt som orsakats av variation i nederbörd. Vi kan därför använda den utifrån nederbörd förväntade tillväxten för att undersöka effekten av tillväxt på inbördeskrig, utan att riskera ett spuriöst samband eller effekter som går i motsatt orsaksriktning.

Denna studie har ett problem att brottas med. Det finns en viss risk att nederbörd påverkar sannolikheten för inbördeskrig på andra sätt än via tillväxten. Kanske förhindrar nederbörd krig genom att regnet förstör vägarna som militärfordonen färdas på? Eller genom att det inte är lika varmt de år som har mycket nederbörd och soldaterna därför är mindre aggressiva?

Författarna tvingas därför lägga mycket utrymme på att utreda alternativa mekanismer för hur vädret kan påverka sannolikheten för inbördeskrig. Efter att ha studerat data på vägsador och temperatur drar de emellertid slutsatsen att deras ursprungliga hypotes håller och att en minskning av den ekonomiska tillväxten (orsakad av minskat regn) med 5 procentenheter ökar sannolikheten för inbördeskonflikt med 50 procent.

Nazister och sverigedemokrater

Kåre Vernby, statsvetare i Uppsala, har tillsammans med två andra forskare studerat sambandet mellan var nazisterna var starka på 1930-talet och var högerextrema partier (framför allt Sverigedemokraterna) är starka idag (Dahlberg m.fl 2012). Korrelationen är ganska stark och där nazisterna en gång i tiden var starka hittar vi många högerextrema idag.

Författarna vill tolka det som en kausal effekt, sådan att förekomsten av nazister har påverkat andelen sverigedemokrater genom att organisationer och personliga band har levt kvar över decennierna. Men hur vet vi att

det inte handlar om att vissa regioner helt enkelt är mer mottagliga för högerextremism och rasism, vilket också skulle kunna förklara att SD är stora i samma områden som nazisterna hade sina starka fästen.

Dahlberg m.fl. besvarar den frågan genom att använda försvarsnedläggningarna 1925 som ett instrument för var stödet senare var starkt för nazisterna. De visar att nazisterna blev starka där man lade ner regementen, vilket kan tolkas som att arbetslösa militärer lockades av nazismen. Fördelen med detta instrument är att sådana faktorer som att vissa regioner är mer mottagliga för högerextremism inte kan ha påverkat vilka regementen som lades ner. Att även Sverigedemokraterna idag är starka där man lade ner regementen kan därför tolkas som en kausal effekt av var nazisterna var stora, och inte av huruvida regionen i övrigt har gynnsamma förutsättningar för högerextremism.

Kritik och begränsningar

Kritiken mot experiment brukar komma i två former. Det vanligaste är att ifrågasätta av experimentens relevans. Det kan ske genom att man kritiserar den externa relevansen eller argumenterar för att experimentet som ideal leder till snäva och irrelevanta frågeställningar. En annan baksida av experimenten är att de kan dras med etiska problem samt att de ofta begränsas av just etiska hänsyn.

Relevans

Den viktigaste kritiken mot experiment går ut på att experimentet inte säger någonting som är relevant för samhället eller vetenskapen – att de saknar teoretisk relevans. Det kan uttryckas som att frågeställningarna blir för begränsade när vi måste kunna besvara frågan med ett experiment eller som att experimentet har dålig extern validitet – att det inte går att generalisera resultaten till verkliga situationer. Det betyder naturligtvis inte att detta gäller alla experiment som utförs, men det är inget kontroversiellt påstående att de flesta experiment dras med den typen av begränsningar.

Mest tydligt är detta problem i laboratorieexperiment och surveyexperiment, där skillnaderna i regel är stora mellan experimentsituationen och det experimentet vill förklara. Säger verkligen ett kontrollerat och övervakat spel någonting om hur en verklig löneförhandling går till? Och hur vet vi att människor beter sig på samma sätt som de uppger i en surveyundersökning?

En annan sida av samma problem är att forskaren anpassar frågorna som ställs efter vad som kan besvaras med hjälp av en experimentell design. Då är risken att frågorna reduceras till väldigt smala fenomen och frågor som är mindre relevanta än de frågor vi egentligen vill ställa.

En ytterligare risk är att experimentet endast studerar *partiella* effekter, eftersom experimentsituationerna ofta är mer begränsade än verkligheten.

Kanske är den BNP-tillväxt som orsakats av mycket nederbörd någonting annat än den BNP-tillväxt som beror på att fler kvinnor kommer ut på arbetsmarknaden och kanske har dessa typer av tillväxt olika effekt på sannolikheten för inbördeskrig?

Till sist bör noteras att experiment är bra på att identifiera storleken på en effekt, men de lämnar oss ofta frågande om varför effekten såg ut som den gjorde. Att spåra mekanismer är i regel inte en del av den experimentella studien.

Etiska problem

När forskaren inte bara observerar utan även försöker manipulera verkligheten, ökar genast risken att vi åsamkar våra analysobjekt skada. Det kan handla om att vi skapar kostnader för tredje part, som när vi skickar ut jobbansökningar på låtsas för att testa teorier om diskriminering vid rekrytering eller när vi medvetet sprider falsk information i ett surveyexperiment. Men det kan också handla om etiskt tveksam särbehandling, som när vi låter en grupp människor få en fördel jämfört med andra grupper.

En viktig omständighet är att många fältexperiment utförs i länder där det är relativt billigt för forskaren att manipulera förutsättningarna för analysobjekten. I ett fattigt land kanske vi kan erbjuda vinstsummor som motsvarar en månadslön för lokalbefolkningen. Då ökar också risken att vi bryter mot etiska regler för vad som är ok, jämfört med om den vinnande parten i ett spel kan vandra hemåt med en trisslott eller biobiljett i handen.

Dessa risker innebär att den som vill utföra experiment som berör människor ofta måste etikpröva sin ansökan. Eftersom sådana ansökningsprocesser kan ta lång tid, och kostar pengar, är det ingenting som rekommenderas för den som vill skriva en C-uppsats.

Några avslutande ord om design

Något ni bör ta med er, både från den här föreläsningen och från kursen i stort, är vikten av en bra design. Med en bra plan och strategi för en undersökning kan vi ge oss goda möjligheter att dra slutsatser om kausala samband eller generalisera våra slutsatser till en större population.

Den kvantitativa samhällsvetenskapliga forskningen utvecklas i hög grad mot allt mer avancerade statistiska metoder. Den utvecklingen är logisk, givet att vi måste modellera verkligheten för att kunna isolera vårt samband från andra förklaringar. Världen är komplex och även de mest sofistikerade modeller är grova förenklingar av den dynamik och de samband som återfinns i verkligheten.

En bra design minskar i regel behovet av avancerade statistiska modeller. Om experimentets ideal är uppfyllda, kommer forskaren i allmänhet långt med att bara jämföra medelvärden för den behandlade och den obehandlade

gruppen! Om designen brister är det däremot svårt att åtgärda detta. Även med en mycket avancerad statistisk modell, skulle många välja att lita mer på en enkel studie som använder en snillrik design.

Det finns en uppfattning om att experiment är någonting som ”kvantare” och inte ”kvallare” sysslar med. Jag tror att den uppfattningen är fel. Visserligen estimeras effekterna i ett experiment med kvantitativa metoder, men att hitta och utvärdera naturliga experiment kräver den typ av detaljkunskap vi ofta förknippar med – eller erhåller genom – kvalitativ forskning. Hur vet vi att systemet för kvotering i Indien var helt slumpmässigt? Genom att läsa lagtext och intervjuar beslutsfattare!

Kombinationsstudier

Precis som att kvalitativa och kvantitativa studier kan kombineras för att utnyttja de båda traditionernas relativa fördelar, bör experiment kombineras med andra metoder. Experimenten är ofta överlägsna andra metoder i att identifiera kausala effekter, men har klara nackdelar när det kommer till generaliseringsmöjligheter och att förstå varför en viss effekt uppkommer. Vi bör därför inte se experimenten som ett substitut till andra metoder, utan som ett välkommet komplement.

Litteratur

- Chattopadhyay, R och E Duflo (2004). "Women as policy makers: Evidence from a randomized policy experiment in India". *Econometrica* 72.5, s. 1409–1443.
- Coffman, Katherine B. m. fl. (2013). "The size of the LGBT population and the magnitude of anti-gay sentiment are substantially underestimated".
- Doherty, D m. fl. (2006). "Personal income and attitudes toward redistribution: A study of lottery winners". *Political Psychology* 27.3, s. 441–458.
- Falk, Armin och James J Heckman (2009). "Lab experiments are a major source of knowledge in the social sciences." *Science (New York, N.Y.)* 326.5952, s. 535–8.
- Gardner, Jonathan och Andrew J Oswald (2007). "Money and mental well-being: a longitudinal study of medium-sized lottery wins." *Journal of health economics* 26.1, s. 49–60.
- Gneezy, Uri och Aldo Rustichini (2000). "A fine is a price". *J. Legal Stud.* 29.1, s. 1–17.
- Levitt, SD och JA List (2007). "What do laboratory experiments measuring social preferences reveal about the real world?" *The journal of economic perspectives* 21.2, s. 153–174.
- Lindahl, M (2005). "Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income". *Journal of Human Resources* 40.1, s. 144–168.
- Lundin, Martin m. fl. (2013). "Political training as a pathway to power".
- Miguel, E m. fl. (2004). "Economic shocks and civil conflict: An instrumental variables approach". *Journal of political Economy* 112.4.
- Milgram, S (1963). "Behavioral study of obedience." *The Journal of Abnormal and Social Psychology* 67.4.
- Mill, J S (1843). *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation v. 1. John W. Parker.

- Press, Daryl G. m. fl. (2013). "Atomic Aversion: Experimental Evidence on Taboos, Traditions, and the Non-Use of Nuclear Weapons". *American Political Science Review* 107.01, s. 188–206.
- Titunik, Rocío (2011). "Incumbency Advantage in Brazil : Evidence from Municipal Mayor Elections".