

Föreläsning 10: Regressionsdiagnostik och experimentell design

Pär Nyman

19 september 2014

Det här är kompletterande anteckningar till föreläsning 10. De är lite mer omfattande än övriga föreläsningsanteckningar och följer inte exakt samma disposition som föreläsningen. På flera områden går anteckningarna längre och djupare än vad som sades på föreläsningen. Innehållet består till största delen av sådant som inte tas upp i Teorell och Svensson. Därför är den här texten också en del av kurslitteraturen. Anteckningarna består liksom föreläsningen av två delar, vilka utan problem kan läsas var för sig.

Den första delen tar upp mer avancerade aspekter av regressionsanalys. Den riktar sig således främst till dem som tycker att kursen har varit för enkel och till dem som kommer att använda kvantitativa metoder i sitt uppsatsskrivande. De avsnitt som behandlades på föreläsningen – icke-linjära samband och outliers – kan komma att examineras på tentan. Övrigt innehåll är sådant som vi tror att ni kan ha nytta av i ert uppsatsskrivande och i grupparbetet inför det femte seminariet. Det kommer däremot inte att förekomma på tentan, om det inte också återfinns i övrigt kursmaterial.

Den andra delen handlar om experimentell design inom samhällsvetenskapen. Textens främsta syfte är att ge en förståelse för vilka problem vi kan lösa med hjälp av experiment, men också att peka på experimentens brister och begränsningar.

Gemensamt för båda delarna är att det viktiga innehållet ligger i helheten snarare än i detaljerna. En bra riktlinje för vad vi tycker är mest centralt återfinns i slutet under rubriken ”Att diskutera”.

Del 1

Fördjupning inom regressionsanalys

Den här texten tar upp mer avancerade aspekter av regressionsanalys än vad som behandlas i Teorell och Svensson. Det finns alltid en risk att ökad kunskap inom ett område gör att man lägger mindre vikt vid andra, minst lika viktiga saker. Jag vill därför påminna om att det viktigaste i en vetenskaplig undersökning sällan är att man använder en avancerad statistisk modell – tvärtom finns det många fördelar med enklare modeller, vilket diskuteras i avsnitt 1.5. De metoder som den här texten tar upp är verktyg vi kan använda för att svara på våra forskningsfrågor och för att skapa modeller som överensstämmer med våra teorier.

Texten kommer att ta upp följande problem och hur vi kan hantera dem:

- Multikollinearitet
- Icke-linjära samband
- Interaktionseffekter
- Outliers och viktiga observationer

Därefter avslutas kapitlet med ett avsnitt om fördelarna med att hålla en regressionsmodell simpel. Det finns nämligen många goda skäl för att göra det.

1.1 Multikollinearitet

Multikollinearitet uppstår när de oberoende variablerna är korrelerade med varandra. I detta avsnitt diskuteras varför det är ett problem, hur symptomen ser ut och hur vi kan lösa problemen.

När vi skattar en variabels regressionskoefficient med hjälp av OLS används endast variation som är unik för den aktuella variabeln. Variation

Tabell 1.1: Multikollinearitet

| | (1) | (2) | (3) |
|----------------|--------------------|--------------------|--------------------|
| Revenues | -0.33*** (0.10) | | -0.13 (0.26) |
| Expenditure | | -0.30*** (0.09) | -0.19 (0.23) |
| Constant | 45.12*** (4.39) | 44.58*** (4.13) | 45.33*** (4.43) |
| Observations | 26 | 26 | 26 |
| Adjusted R^2 | 0.282 | 0.294 | 0.271 |

Standardfel i parenteser. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

som delas med en annan variabel kan inte användas i detta syfte eftersom det är omöjligt att veta vilken av variablerna det är som påverkar den beroende variabeln. När den användbara variationen minskar blir våra skattningar mindre precisa, koefficienternas standardfel större och det blir svårare att hitta statistiskt signifikanta resultat.

Låt oss anta att vi vill studera hypotesen att en stor offentlig sektor ökar graden av inkomstfördelning i ett land. Vi vill också veta om det är stora offentliga utgifter eller höga skatter som orsakar detta samband. Tabell 1.1 visar resultaten från tre regressioner av inkomstjämlighet, mätt med den så kallade ginikoefficienten. Den första kolumnen visar en bivariat regression med offentliga intäkter som oberoende variabel och den andra kolumnen en bivariat regression med offentliga utgifter som oberoende variabel. Den tredje kolumnen visar en trivariat regression där båda variablerna ingår i modellen. Båda variablerna har en statistiskt signifikant effekt i de bivariata regressionerna. Eftersom koefficienterna är negativa visar resultaten att länder med stora offentliga inkomster respektive stora offentliga utgifter har en jämnare inkomstfördelning. I den trivariata regressionen är ingen av dem signifikant. Standardfelen för båda koefficienterna har ökat kraftigt, vilket beror på att vi har problem med multikollinearitet. Det är inte heller överraskande. Länder med stora offentliga utgifter har naturligtvis även stora skatteintäkter, om de inte finansierar utgifterna med stora budgetunderskott. Eftersom korrelationen är så stark mellan de två variablerna kan vår modell inte avgöra vilken av dem det är som påverkar inkomstspridningen.

Eftersom multikollinearitet minskar variationen vi kan använda för våra skattningar, påminner problemet mycket om de svårigheter som kan uppstå när man har för lite data. Därför är det också svårt att säga hur mycket multikollinearitet vi kan hantera, på samma sätt som det inte finns någon tumregel för hur många observationer som krävs för att få statistiskt signifikanta resultat. Vid vilken grad av multikollinearitet som det uppstår

problem beror helt och hållet på vilken fråga vi ställer, storleken på de effekter vi skattar, hur många observationer vi har och hur mycket variation de innehåller samt hur avancerad regressionsmodell vi vill använda.

Ett tecken på multikollinearitet är att vi har god passning (högt R^2 och litet regressionsstandardfel) men inga statistiskt signifikanta effekter. Om vi för in en ny variabel och det ökar standardfelen på övriga variablers regressionskoefficienter är även det ett tecken på multikollinearitet. Vi kan också mäta graden av multikollinearitet genom att beräkna ett VIF-värde (VIF står för Variance Inflation Factor) för varje variabel. VIF-värdet för en variabel berättar hur väl den variabeln kan predikeras utifrån värdet på de övriga oberoende variablerna. Det förekommer många tumregler för vad som är oroväckande höga VIF-värden, men i enlighet med argumentationen i stycket ovanför är det i allmänhet bättre att göra en helhetsbedömning än att förlita sig på tumregler.

Det är viktigt att inse att multikollinearitet inte är något fel eller något vi till varje pris bör undvika. Tvärtom kan det ses som en naturlig konsekvens av att de fenomen vi observerar knappast är oberoende av varandra och att de därför ofta samvarierar. Om den samvariationen är stark är det – och bör så också vara – svårt att veta vilket av två eller flera tätt sammanknutna fenomen som påverkar vår beroende variabel. Så länge den inte utgör något avgörande problem skulle jag i de flesta fall rekommendera att man inte heller försöker åtgärda den.

Om den utgör ett problem finns det emellertid ett flertal möjliga utvägar. Det handlar dock inte om någon mirakelmedicin, utan lösningen går i allmänhet ut på att vi kringgår problemet på ett eller annat sätt. I grunden finns det tre olika metoder för att minska multikollineariteten. Både den andra och den tredje metoden innebär att vi frånsäger oss möjligheten att separera variabler från varandra. Vilket av dessa tre alternativ vi väljer beror därför i hög utsträckning på teoretiska resonemang som är specifika för det fenomen vi studerar. Mer konkret måste vi själva bestämma vilka åtskillnader mellan variabler som är teoretiskt relevanta.

1. Mer data.
2. Ta bort variabler.
3. Slå ihop variabler.

Om det är möjligt att använda mer data är det ofta den bästa lösningen, eftersom vi då ökar den variation vi kan använda i våra skattningar utan att vi behöver förändra vår modell. I regel är det dock kostsamt, problematiskt eller till och med omöjligt att öka antalet observationer.

Ett vanligare alternativ är att vi tar bort någon variabel från modellen. Då har vi kapitulerat inför problemet och det faktum att vi inte kan isolera variablerna från varandra. Om vi tar bort en variabel innebär det att vi

inte kan utesluta att det faktiskt är den variabeln som påverkar vår beroende variabel. Ibland är det ett avgörande problem, särskilt om variablerna mäter helt olika saker och att den teoretiska relevansen av undersökningen förutsätter att vi vet vilken variabel som är viktig. Andra gånger är det ett mindre problem och vår teoretiska poäng är kanske giltig oavsett om det är den första, den andra eller båda variablerna som påverkar den beroende variabeln. Kanske finns det tidigare forskning som visat att variabeln vi tar bort inte är så viktig.

Om variablerna mäter liknande saker kan vi i stället för att ta bort en variabel välja att slå samman variablerna till en. Kanske har vi surveydata som innehåller tre frågor om respondentens tillit: till sina grannar, till politiker och till myndigheterna. Vi kan då reducera dessa frågor till en enda variabel, vilken mäter respondentens generella tillit. Det finns många metoder för att göra det. Enklast är att vi skapar något slags index.¹

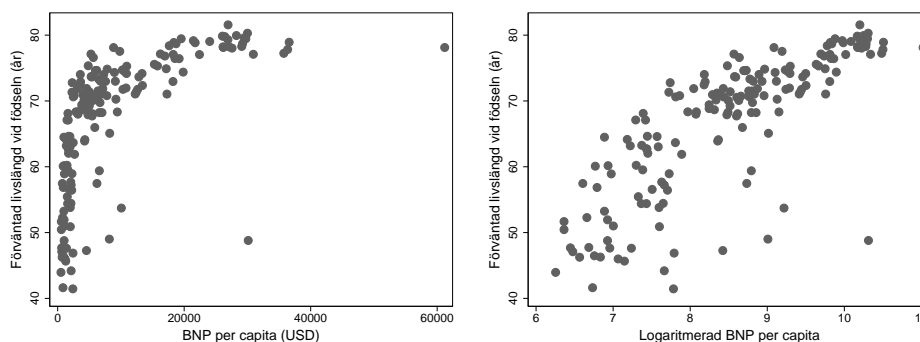
1.2 Icke-linjära samband

Den linjära regressions ekvationen antar alltid ett linjärt samband mellan å ena sidan den beroende variabeln och å andra sidan de oberoende variablerna – alltså att en förändring i en oberoende variabel orsakar samma förväntade förändring i den beroende variabeln oavsett var på x-axeln vi befinner oss. Många samband inom samhällsvetenskapen är emellertid avtagande (effekten av x på y är mindre vid höga värden på x), tilltagande (effekten av x på y är större vid höga värden på x), U-formade (effekten av x på y har olika tecken vid höga respektive låga värden på x) eller S-formade (effekten av x på y är störst vid medelhöga värden på x). Det här avsnittet handlar om hur vi upptäcker och hanterar sådana icke-linjära samband.

1.2.1 Avtagande och tilltagande samband

Det är välbelagt att ekonomiskt välstånd har en positiv men avtagande effekt på människors hälsa. Exempelvis skulle en ökning av medelinkomsten med 100 kr om dagen innebära mycket för medellivslängden i ett fattigt land som Zambia men kanske inte ha någon effekt alls i Sverige. Teoretiskt kan vi motivera det med att en inkomstökning i Zambia i högre grad används till att täcka upp grundläggande behov, som t.ex. livsmedel och medicin. Grafiskt kan vi upptäcka och illustrera icke-linjära samband genom att illustrera sambandet i ett punktdiagram. Sambandet mellan BNP och medellivslängd

¹Den enklaste formen av index är att bara beräkna medelvärdet av de olika variablerna. Om variablerna är mätta på olika skalor vill vi kanske standardisera variablerna innan vi slår ihop dem (standardisering innebär att vi transformerar variablerna så att de får medelvärde 0 och standardavvikelse 1). En mer avancerad metod är att vi använder faktoranalys för att skapa en variabel som fångar den gemensamma variationen i de olika delkomponenterna, vilket vi i detta fall antar är den generella tilliten.



Figur 1.1: Sambandet mellan BNP och förväntad livslängd

illustreras i Figur 1.2.1. Det vänstra diagrammet visar det ursprungliga sambandet. Om vi skulle försöka dra en rät regressionslinje genom datapunkterna skulle den bli uppenbart missvisande. I det högre diagrammet har vi transformerat BNP-variabeln genom logaritmering och sambandet är nu mer linjärt. Det här avsnittet handlar om hur vi kan hantera icke-linjära samband inom ramen för vår regressionsanalys.

Den vanligaste metoden för att hantera icke-linjära samband är att transformera den oberoende variabeln.² Med en transformering menas att alla variabelvärden (x_i) ersätts med transformerade värden (z_i) vilka är en funktion av de ursprungliga värdena ($z_i = f(x_i)$).

Den vanligaste transformeringen för avtagande samband kallas för logaritmering. Det innebär att vi transformerar variabeln så att den i stället antar logaritmen av sina värden:

$$z = \log x + e \tag{1.1}$$

Logaritmerade variabler har den attraktiva egenskapen att en ökning av z med 1 alltid motsvarar en viss relativ förändring i x ³ När vi logaritmerar den oberoende variabeln analyserar vi därför effekten av relativa förändringar ("om x ökar med 100 procent") i stället för absoluta förändringar ("om x ökar med 10"). Valet att logaritmera kan motiveras både empiriskt ("passningen blir bättre") och teoretiskt ("det är rimligt att en fördubbling av BNP orsakar en lika stor förändring i förväntad livslängd oavsett BNP-nivå").

Det vanligaste exemplet på logaritmering är kanske BNP och inkomster, eftersom det är en vanlig hypotes att effekten av en inkomstförändring står i proportion till hur stor inkomstförändringen var i relativa termer (en

²Det förekommer att man i stället transformerar den beroende variabeln, men det är inte lika vanligt.

³Hur stor relativ förändring det motsvarar beror på vilken bas som har använts vid logaritmeringen. En ökning av det logaritmerade värdet med 1 motsvarar en fördubbling av x för \log_2 , en tioudubbling för \log_{10} , och så vidare.

fördubbling av inkomsten antas ge samma effekt, oavsett hur hög inkomsten är).

Sambandet i Figur 1.2.1 blev inte perfekt linjärt efter logaritmeringen, men å andra sidan är det möjligt att med denna transformation tolka sambandet som att en fördubbling av BNP per capita innebär en viss ökning av medellivslängden. Enkelhet och tydliga teoretiska tolkningar kan alltid påverka vilken transformering vi väljer.

1.2.2 U-samband och polynom

En annan typ av icke-linjära samband är så kallade U-samband, vars namn kommer från formen på sambandet. U-samband innebär att både låga och höga värden på den oberoende variabeln sammanfaller med höga värden på den beroende variabeln. Motsatsen är minst lika vanlig och kallas ibland för upp-och-ner-vända U-samband. Exempelvis är sambandet mellan ålder och inkomst sådant att inkomsterna är höga för medelålders personer och lägre för både unga och gamla. Ett annat exempel på ett upp-och-ner-vänt U-samband är den så kallade Kuznetskurvan, vilken visar sambandet mellan den genomsnittliga inkomsten och ojämlikheten i ett land. Både fattiga och rika länder tenderar att ha en liten inkomstspridning. Ojämlikheten är som störst i länder som är genomsnittligt rika.

För att modellera ett U-samband i vår regressionsmodell brukar vi använda det otransformerade värdet av en variabel tillsammans med samma variabel i kvadrat. Så om vi skapar en variabel z enligt ($z = x^2$), så skulle vår regressionsmodell för ett U-samband skrivas

$$y = a + b_1x + b_2z + e \quad (1.2)$$

Ett U-samband känns igen på att b_1 är negativ och b_2 positiv. På samma sätt har vi ett upp-och-ner-vänt U-samband om b_1 är positiv och b_2 negativ.

När vi undersöker den statistiska signifikansen i den här typen av samband blir det lite mer komplicerat än vad vi är vana vid. Att b_1 är signifikant säger inte så mycket alls. Om b_2 är signifikant betyder det att vi kan vara säkra på att sambandet är icke-linjärt. För att testa om variabeln är signifikant vill vi testa nollhypotesen att varken den ursprungliga eller den transformerade variabeln har någon effekt:

$$\begin{aligned} \beta_1 &= 0 \\ \beta_2 &= 0 \end{aligned} \quad (1.3)$$

1.2.3 Dummyvariabler

Ett annat tillvägagångssätt är att skapa en uppsättning dikotoma variabler där varje variabel motsvarar ett intervall av värden på den ursprungliga

värden. Därefter kan man i en regression skatta separata effekter för varje intervall. Det är en väldigt flexibel metod i betydelsen att effekterna kan anta vilka värden som helst – vi sätter inga restriktioner på vilken form som sambandet kan anta. Å andra sidan låser vi effekten till att vara konstant inom varje intervall.

Exempelvis kan vi i stället för att modellera inkomst som ett omvänt U-samband av ålder, skapa tre ålderskategorier för unga, medelålders och gamla. Vi tvingar då modellen att predikera samma inkomst för olika åldrar inom en åldersgrupp, men vi tillåter effekterna för de olika åldersgrupperna att anta vilka värden som helst. Det kan vara en bra metod om vi inte har någon aning om formen på sambandet eller om den ursprungliga variabeln lämpar sig för att skapa kategorier. Nominalskalor och ordinalskalor med få skalsteg är ofta bra kandidater för att skapa dummyvariabler.

1.3 Interaktion

Hittills har vi antagit en helt additiv modell där effekterna av olika variabler är oberoende av varandra, men så ser inte alltid verkligheten ut. Ofta tänker vi oss att effekten av en variabel är beroende av värdet på en annan variabel. Det har till exempel visats att små undervisningsgrupper i skolan förbättrar skolresultaten mer för personer med utländsk eller arbetarbakgrund än för personer med inrikes födda och högutbildade föräldrar.

För att testa den hypotesen skulle vi kunna ställa upp en regressionsanalys där en elevs prestationer (y) utgör en funktion av storleken på undervisningsgrupperna (x), huruvida eleven har utländsk bakgrund (z) samt en interaktionseffekt av de oberoende variablerna ($x \times z$).

$$y = a + b_1x + b_2z + b_3xz + e \quad (1.4)$$

Om $z = 1$ för personer med utländsk bakgrund och $z = 0$ för övriga, kan vi tolka b_1 som effekten av undervisningsgruppens storlek för personer utan utländsk bakgrund (eftersom $z = 0$ så är $xz = 0$) och $b_1 + b_3$ som effekten av gruppstorlek för personer med utländsk bakgrund (eftersom $z = 1$ så är $xz = z$).

Dikotoma variabler är populära vid utformandet av interaktionsvariabler, eftersom det blir så enkelt att tolka effekterna. De interagerande variablerna behöver emellertid inte vara dikotoma variabler.

Precis som med U-samband är det lite mer komplicerat att studera statistisk signifikans när vi har interaktionsvariabler. Huruvida regressionskoefficienterna för de ursprungliga variablerna är signifikanta säger inte så mycket. Om interaktionstermen har en signifikant effekt kan vi tolka det som att effekten av den ena variabeln beror på värdet på den andra variabeln och att den skillnaden i effekt är statistiskt signifikant. Vi kan även välja att testa nollhypotesen att flera av populationens regressionskoefficienter är noll,

för att testa om den gemensamma effekten av de ursprungliga variablerna och interaktionsvariabeln är statistiskt signifikant.

1.4 Outliers och viktiga observationer

Enstaka observationer som skiljer sig kraftigt från de flesta övriga observationer i ett datamaterial kallas för outliers. Sådana observationer är ofta avgörande för de resultat vi får när vi genomför regressionsanalyser. Om vi väljer att exkludera dem från våra modeller, får vi ofta helt andra resultat än när vi har dem med. Det här avsnittet handlar om att identifiera outliers, analysera vikten av dem och hur vi kan åtgärda eventuella problem. Låt oss börja med att skilja på tre typer av outliers. Kategorierna är inte ömsesidigt exkluderande, utan en observation kan samtidigt ha flera typer av "outlierness".

- Observationer med *leverage* är observationer som har ovanliga värden på de oberoende variablerna. Om värdet på den beroende variabeln är i linje med vad vi skulle förvänta oss utifrån regressionslinjen (helst en linje skattad utan extremvärdet) har sådana observationer ingen påverkan på regressionskoefficienterna, annat än att estimaten blir mer precisa.
- Observationer som har ovanliga variabelvärden på den beroende variabeln givet värdena på de oberoende variablerna kallas för *vertikala outliers* eller *regressionsoutliers*. De har ofta stora residualer, men variabelvärdena behöver inte vara ovanliga i sig. Det är nästan samma sak som Teorell och Svensson kallar för avvikande fall.
- Vertikala outliers som dessutom har högt leverage kommer att ha stor påverkan på våra regressionsresultat. Den engelska termen för sådana observationer är *influential observations*. Vi säger ofta att observationerna drar regressionslinjen till sig.⁴

Outliers kan uppstå av många olika skäl och vi bör alltid fråga oss om de extrema värdena speglar verkligheten eller någon sorts mätfel. Om vi har extrema värden för att någon har kodat fel är det självklart att vi vill exkludera dem från våra analyser. Men om de extrema värdena återspeglar verkligheten har vi ett mer svårlöst problem. Å ena sidan är det ju verkligheten vi vill beskriva, men å andra sidan tenderar den typen av observationer att få en orimligt stor betydelse för de effekter vi skattar i våra regressionsanalyser. Det finns i grunden tre metoder för att hantera outliers:

⁴Det vanligaste måttet på leverage är så kallade hat values. För att identifiera vertikala outliers kan man använda studentized residuals med Bonferronijusterade p-värden. Det vanligaste måttet på influence är Cook's Distance eller Cook's D.

- Behålla dem.
- Ta bort dem.
- Behålla dem men ge dem mindre betydelse.

Om vi vet att våra outliers representerar mätfel är valet i allmänhet enkelt⁵, men annars finns det inga självklara svar. Jag brukar rekommendera att utföra regressionerna både med och utan outliers för att se om resultaten påverkas.⁶

Hur vi behandlar extremvärden handlar också om vilka generaliseringssamband vi har. Tar vi bort observationer med extrema värden på den oberoende variabeln begränsar vi förstås våra möjligheter att extrapolera resultaten till att gälla observationer utanför vårt sample, men beroende på våra syften behöver det inte vara något problem.

1.5 Flexibla eller enkla modeller?

Den här texten har i hög utsträckning handlat om varför vi ibland måste skapa mer komplicerade regressionsmodeller och olika metoder för att göra detta. Mycket lite har sagts om varför det kan vara en dålig idé att använda mer avancerade modellspecifikationer.

Varför antar vi att ett samband är linjärt, när vi kan lämna det öppet och låta en mer flexibel modell avgöra formen på sambandet? Varför inkluderar vi inte bara alla variabler för vilka vi har data, med alla tänkbara transformationer och interaktioner, och testar vad som är signifikant eller erbjuder bäst passning? Det finns i alla fall tre anledningar till varför vi ofta föredrar enklare modeller.

För det första har enkelhet ett värde i sig. Enkla modeller är lättare att beskriva och ta till sig. Det handlar inte bara om bekvämlighet, utan gör dem även jämförbara med modeller som andra forskare använder. Om två forskare har använt liknande modeller är det enklare att jämföra resultaten, än om forskarna har använt mer komplicerade modeller där varje effekt beror på värdet på flera andra variabler.

För det andra svarar enkla modeller i regel bättre mot de teorier vi vill testa. Vi kanske har en idé om att storleken på en undervisningsgrupp påverkar elevernas lärande – kanske till och med att effekten beror på om eleven har utländsk bakgrund – men sällan säger vår teori att effekten av gruppstorlek följer ett sjundegradspolynom. Eftersom våra teorier ofta

⁵Kom dock ihåg att vi kan skapa en bias i våra resultat om vi exempelvis bara tar bort slumpmässiga mätfel som ger för höga värden, och inte de mätvärden som är för låga.

⁶Ibland är det emellertid bättre att sikta på ett mellanläge, så att man inkluderar alla observationer men justerar deras relativa betydelse så att man reducerar det inflytande som outliers har. De flesta sådana metoder går under namnet *robust regression*.

motsvarar ganska enkla modeller, är det inte ovanligt att vi använder mer flexibla specifikationer för våra kontrollvariabler än för de variabler som ingår i vår forskningsfråga.

För det tredje finns det statistiska skäl för att hålla regressionsmodellerna relativt enkla. Ju fler parametrar som ska estimeras, desto mindre precisa blir dessa uppskattningar. Ju mer flexibel en modell är, desto mer anpassar den sig efter de mest närbelägna observationerna. Därför kommer den också att fånga upp mer slumpmässiga variationer än vad en mer begränsad modell gör. Således kan en enklare modell med sämre passning i ett urval beskriva populationen eller den datagenererande processen bättre än vad en mer komplex modell gör. Anledningen är att den flexibla modellen kan överreagera på osystematisk variation i urvalet.⁷

Sist men inte minst är det viktigt huruvida vi teoretiskt kan motivera mer avancerade modeller. Om de tendenser vi kan skönja i våra observationer motsvarar våra teoretiska förväntningar, minskar risken att de beror på slumpmässig variation i data och vi blir mer benägna att använda en mer flexibel regressionsmodell.

⁷I sammanhanget bör det kanske nämnas att det existerar metoder för att inom modellen avgöra hur känslig den ska vara för lokala variationer i data.

Del 2

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 avsnitt 2.1 logiken bakom experiment och vilka problem de löser. Det är den viktigaste delen av texten. De mest omfattande delarna av texten är avsnitt 2.2 och 2.3, vilka tar upp de viktigaste typerna av kontrollerade och naturliga experiment samt lämnar ett stort antal exempel från tidigare forskning. Avsnitt 2.4 och 2.5 avslutar med att diskutera den viktigaste kritiken som brukar riktas mot experiment samt avrundar med några allmänna lärdomar om design.

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

2.1 Experimentets logik

2.1.1 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

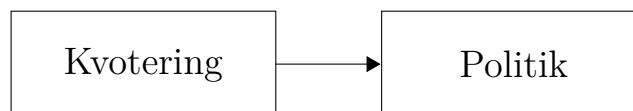
merely lightened, not removed: men knew how to raise water in pumps long before they adverted to what was really the operating circumstance in the means they employed, namely, the pressure of the atmosphere on the open surface of the water.

- J.S. Mill (1843), p. 481.

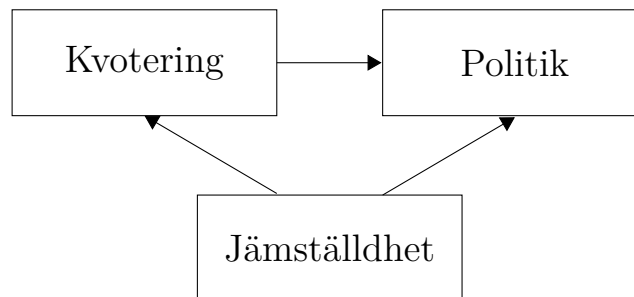
Många av de problem vi stöter på när vi använder icke-experimentell data handlar om att vi har svårt att belägga ett av de fyra orsakskriterierna – isolering. I en multivariat regression försöker vi åstadkomma isolering genom att inkludera relevanta kontrollvariabler, men det finns flera skäl för varför 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. 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. Som bonus skulle vi lösa eventuella problem med omvänd kausalitet (att y påverkar x)!

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).

2.2 Kontrollerade 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.

Det finns många sätt att kategorisera experiment. Här har de kontrollerade experimenten delats upp i tre kategorier: laboratorieexperiment, fältexperiment och surveyexperiment.

2.2.1 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.

Ultimatums spel

Ett klassiskt experiment är det så kallade ultimatumspellet, 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 psykologistuderenter 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.

2.2.2 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. Tanken är att experimentmiljön ska vara identisk med en verklig situation.

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 träffa fler väljare, 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 skillnaden mellan grupperna endast kunde bero på dörrknackandet.

2.2.3 Enkätexperiment

Att alternera frågeformuleringarna eller bakgrundsinformationerna 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 respondenden 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, Sagan och Valentino (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. Det finns många metoder för att hantera det problemet. 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 in på honom. 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 skatta hur stor andel som anser att påståendet är korrekt. 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.

Eftersom det kan finnas slumpmässiga skillnader i hur många som instämde i de fyra första påståendena, och eftersom vi vill generalisera våra resultat utanför det studerade urvalet, bör vi beräkna ett konfidensintervall för skillnaden. Då kan vi använda följande formel för konfidensintervall för differenser 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.

$$\bar{x}_1 - \bar{x}_2 \pm t_{kv} \sqrt{\frac{s_{x_1}^2}{n_{x_1}} + \frac{s_{x_2}^2}{n_{x_2}}} \quad (2.1)$$

Ekonomen Katherine Coffman (Coffman, Coffman och Ericson 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. Endast 11 procent svarade ja när de fick en direkt fråga, trots att enkäten var anonym.

2.3 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.

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 analysenheter på den andra sidan gränsen får. Om det inte är möjligt för analysenheter 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 påverka eller påverkas av någonting annat av relevans.

2.3.1 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, Gerber och Green (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 lotterinumner 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.

2.3.2 Trösklar

Så kallad Regression Discontinuity Design (RDD) 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 tröskelvärdet, kan allokeringen kring tröskeln \hat{U} - och därmed tilldelningen av behandlingen \hat{U} - 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 RDD är att fördelningen kring tröskelvärdet kan betraktas som slumpmässig, så att observationerna precis ovanför tröskelvärdet inte skiljer sig systematiskt från observationerna nedanför tröskelvärdet, 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 suttit 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 svårösliga 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 oviss. 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, Skans och Zetterberg 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.

2.3.3 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.

$$\begin{aligned}\hat{y} &= a_1 + b_1\hat{x} \\ \hat{x} &= a_2 + b_2z\end{aligned}\tag{2.2}$$

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, Satyanath och Sergenti (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 torka.

Det fina med väder är att variationen i nederbörd kan betraktas som slumpmässig. Eftersom förekomsten av inbördeskrig inte kan påverka mängden nederbörd kan krigen 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 vägskador 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 emottagliga för högerextremism och rasism, vilket då både skulle förklara variationen i nazister på 1930-talet och variationen i sverigedemokrater idag?

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.

2.4 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.

2.4.1 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.

2.4.2 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 jobbsökningar 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.

2.5 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 intervjua beslutsfattare!

2.6 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.

Att diskutera

Icke-linjära samband

- Vad menar vi med icke-linjära samband?
- Vad innebär det teoretiskt att ett samband är linjärt, avtagande eller U-format?

Outliers

- Vad är en outlier?
- Varför bör vi bry oss om dem?
- Vad bör vi göra om vi har outliers i vårt datamaterial?

Experiment

- Vilka problem löser vi genom att utföra experiment?
- Vad finns det för begränsningar med experiment?

Litteratur

- Chattopadhyay, R och E Duflo (2004). "Women as policy makers: Evidence from a randomized policy experiment in India". I: *Econometrica* 72.5, s. 1409–1443.
- Coffman, Katherine B., Lucas C. Coffman och Keith M. Marzilli Ericson (2013). "The size of the LGBT population and the magnitude of anti-gay sentiment are substantially underestimated".
- Doherty, D, AS Gerber och DP Green (2006). "Personal income and attitudes toward redistribution: A study of lottery winners". I: *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." I: *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." I: *Journal of health economics* 26.1, s. 49–60.
- Gneezy, Uri och Aldo Rustichini (2000). "Fine is a price, a". I: *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?" I: *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". I: *Journal of Human Resources* 40.1, s. 144–168.
- Lundin, Martin, Oskar Nordström Skans och Pär Zetterberg (2013). "Political training as a pathway to power".
- Miguel, E, S Satyanath och E Sergenti (2004). "Economic shocks and civil conflict: An instrumental variables approach". I: *Journal of political Economy* 112.4.
- Milgram, S (1963). "Behavioral study of obedience." I: *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., Scott D. Sagan och Benjamin a. Valentino (2013). "Atomic Aversion: Experimental Evidence on Taboos, Traditions, and the Non-Use of Nuclear Weapons". I: *American Political Science Review* 107.01, s. 188–206.

Titunik, Rocío (2011). "Incumbency Advantage in Brazil : Evidence from Municipal Mayor Elections".