

Lunds universitet
Statistiska institutionen

STAH11, Kandidatuppsats 15 hp
HT12
Handledare: Jakob Bergman

Brottsligheten i Sverige

En statistisk studie ur ett socioekonomiskt perspektiv

Jonas Stenberg
Christoffer Olsson

Abstract

Earlier studies in other countries have found that inequity and social economic factors influence crime. We examine if the same influence could be found in Sweden. Through a multiple regression we analyze material from Sweden's 290 municipalities.

The dependent variable is measured through reported acts of violent crime since this reflects the total crime rate better than other possible measurements.

Our results show that social economic factors indeed are an important indicator of crime in Sweden. Inequity, income and social benefits all affect the municipalities' crime rates in economic terms.

We also conclude that health and age affect the level of crime through that a high level of poor health and a low average aged population increase the crime rate. In contrast to what is expected municipalities' with a high proportion of females have a higher crime rate. We argue that the result can be derived to a higher proportion of females in metropolitan areas.

Further, immigration and separations shown influence in the model. Municipalities' with a high proportion of immigrants and a high proportion of children with separated parents have more crimes committed per capita. Based on these results we request measures against crime through social reforms.

Nyckelord: brott, våldsbrott, kommuner, socioekonomisk, ojämlikhet

Innehållsförteckning

1	Inledning	1
1.1	Syfte och frågeställning.....	1
2	Bakgrund	2
2.1	Forskningsbakgrund.....	2
3	Datamaterial	5
3.1	Inverkan på statistiken.....	5
3.1.1	Ändringar i statistikrutiner.....	5
3.1.2	Mörkertal.....	6
3.1.3	Anmälningbenägenhet.....	6
3.1.4	Regionala variationer.....	7
3.2	Variabler.....	7
3.2.1	Felaktigt utelämnade variabler.....	8
4	Metod	9
5	Resultat	14
5.1	Modellering.....	14
5.2	Modellvalidering.....	16
5.3	Resultatredovisning.....	17
6	Slutsats	20
6.1	Framtida forskning.....	20
7	Referenser	22
8	Bilagor	24
	Bilaga 1: Förklaring till variablerna.....	24
	Bilaga 2: Regression.....	25
	Bilaga 3: Korrelationsmatris.....	26
	Bilaga 4: Stegvis regression.....	27
	Bilaga 5: Bästa delmängdsregression.....	28
	Bilaga 6: Figurer och diagram.....	29

1 Inledning

Antalet anmälda brott i Sverige har ökat över tiden. Statistik från Brottsförebyggande rådet (Brå) visar att antalet anmälda brott har ökat kontinuerligt sedan 1970-talet (Brå 2011, s 13). Statens främsta uppgift enligt regeringsreformen är att trygga den enskildes personliga, ekonomiska och kulturella välfärd (Hofer 2011, s. 15). Genom att ställa brottsligheten i kontrast till denna trygghet tycks därmed uppgiften utförts sämre med åren.

Brottsligheten har ofta en stor inverkan på människor, inte minst blir offren till brotten påverkade på ett negativt sätt. År 2010 anmäldes drygt 1 370 000 brott i Sverige (Brå 2011, s. 13). Av dessa brott var 113 000 våldsbrott (Brå 2011, s. 26). Det är just våldsbrott som vi är intresserade av och väljer därför att använda oss av detta material i vår undersökning. Brottsligheten påverkar alltså en stor del av invånarna i Sverige samtidigt som det är troligt att siffran är ännu högre på grund av mörkertalet.

Det är därför intressant att veta vad som orsakar brottsligheten. Undersökningar i andra länder har visat att socioekonomiska faktorer har en viktig inverkan på brottsligheten (Blau & Blau 1982; Sachsida m.fl. 2011; Sampson m.fl. 1997). Sverige bör i detta avseende inte vara något undantag.

1.1 Syfte och frågeställning

Vi avser finna en förklaring till vad som påverkar brottsligheten i Sverige. Med hjälp av en multipel regression och med material från alla 290 svenska kommuner hoppas vi finna ett svar på vår frågeställning. Vi gör inga anspråk på att helt förklara uppkomsten till brottslighet utan inriktar oss på ett antal faktorer som utifrån tidigare forskning visat sig inverka. Genom att sedan kontrollera för dessa påverkansfaktorer hoppas vi se vad som i Sverige påverkar antalet brott på en samhällsnivå. Vår frågeställning är:

- Påverkar socioekonomiska förhållanden antalet våldsbrott i Sverige och i så fall på vilket sätt?

2 Bakgrund

Den svenska brottsstatistiken går långt bak i tiden. I ett brev från dåvarande svenska kungen 1830 uppmanades domstolarna att börja upprätta statistik över antalet brott i Sverige. Kungen uttryckte en tro om att detta skulle öka förståelsen till uppkomsten av brottslighet och även hur åtgärder mot den samma ska utformas. Detta var starten på en mångårig insamling av data som har förts i landet och Sverige har idag en rik brottsstatistik (Hofer 2011, s. 223). År 1974 startades Brottsförebyggande rådet på uppdrag av regeringen för att bland annat föra brottsstatistik. Myndigheten jobbar även för minskad brottslighet och ökad trygghet genom brottsförebyggande arbete (Brå 2).

Sedan brottsstatistik började föras har antalet anmälda brott varierat kraftigt. Beroende på vilken typ av brott det handlar om har kategoriseringen förändrats över tiden och även andra faktorer som lagförändringar påverkar statistiken. Det medför att det över tiden är svårt att jämföra olika brott (Hofer 2011, s. 183-192). Dock kommer detta inte utgöra något problem för oss då vi endast fokuserar på ett år.

Historiskt har skillnaden mellan olika grupper vad gäller brottslighet varit stor. Bland annat har skillnaden i brottslighet mellan män och kvinnor varit betydande. Det finns ett antal teorier kring detta, det är dock inget vi kommer gå in närmare på. Åldern på förövaren som begår grövre brott har på senare tid sjunkit. Också antalet brott som begås av utländska medborgare har ökat. Sedan 1950-talet har antalet brott från denna grupp ökat kraftigt. Detta hopp kan förklaras av den ökande arbetskraftsinvandringen under denna tid (Hofer 2011, s. 153-159).

2.1 Forskningsbakgrund

Tidigare studier har visat att brottslighet och ojämlikhet har ett samband. Socioekonomiska förhållanden i en region påverkar antalet brott som begås (Blau & Blau 1982; Sachsida m.fl. 2011; Sampson m.fl. 1997). Även etniska skillnader framhålls som en viktig indikator. Speciellt i USA har de etniska skillnaderna varit tongivande på grund av sin historia. Slumområden med en övervägande svart befolkning har troligen påverkat detta tidigare (Blau & Blau 1982, s. 115).

Fattigdom är en viktig del i de socioekonomiska förhållandena som visat sig påverka. Argumentet för detta är att fattigdom inverkar på de informella sociala normerna hos individen vilket senare resulterar i brottsliga handlingar (Sachsida m.fl. 2009, s. 95).

Vi har ovan nämnt att ojämlikhet kan ha en inverkan på kriminalitet. Men det finns även tecken på att det motsatta förhållandet kan råda, det vill säga att

kriminalitet inverkar på ojämlikhet. Det är något vi måste vara uppmärksamma på när resultatet sedan tolkas. Bland annat kan hög brottslighet i ett område betyda att investerare har en lägre benägenhet att investera där, och det kan medföra att dessa områden därmed blir fattigare och fattigare. Arbetsplatser i brottstäta områden kan också bli mindre attraktiva. Till exempel är en lärare mer benägen att jobba i ett område med låg brottslighet. Samtidigt är individer med ekonomiska resurser mindre benägna att bo i områden med hög brottslighet. Alla dessa faktorer kan påverka att brottsligheten istället skapar ojämlika förhållanden (Sachsida m.fl. 2009, s. 96).

Inkomst har tidigare visat sig vara en variabel att ha i åtanke. En relation mellan brott och inkomst kan förklaras av att alternativkostnaden för brott blir högre med en högre inkomst. Detta är förväntat utifrån ekonomiska modeller (Sachsida m.fl. 2009, s. 97). Fattigdom är också en viktig faktor som förväntas påverka brottsbenägenheten på samma sätt. Det eftersom även fattigdom minskar alternativkostnaden att begå brott. Väldigt fattiga människor ser därmed inte att de förlorar på att bli gripna för ett brott jämfört med förtjänsten brottet för med sig (Sachsida m.fl. 2009, s. 98).

Andra studier har visat att det finns en relation mellan brott och arbetslöshet. Det finns två effekter som arbetslöshet ger. Den ena är att arbetslösa har mindre pengar till förfogande vilket leder till att de har ett större incitament att begå brott. Den andra effekten handlar om att den arbetslöse känner en större motivation till att begå brott eftersom denne på grund av sin frånvaro från arbetsmarknaden känner sig utanför samhället (Sachsida m.fl. 2009, s. 97).

På individuell nivå har tidigare studier påvisat att ålder och kön påverkar brottsbenägenheten (Sampson m.fl. 1997). Statistiken påverkas till stor del av att yngre män är överrepresenterade i brottsstatistiken (Hofer 2011, s. 153-158). Även utbildning har visat sig vara en faktor som undersökningar tar hänsyn till. Utbildning har alltid nämnts ha en viktig effekt på kriminalitet (Sachsida m.fl. 2009, s. 97).

I amerikanska studier har etnicitet varit en flitigt använd variabel i sammanhanget (Blau & Blau 1982; Sampson m.fl. 1997). Det beror på USA:s historia och är troligen inte lika relevant för den svenska kontexten. I Sverige visar de flesta studier på området att invandrare i större utsträckning än den övriga befolkningen begår brott (Martens & Holmberg 2005; Kardell 2011, s. 11). Anledningen till dessa resultat är en känslig fråga och diskuteras flitigt (Kardell 2011, s. 13-34). Vi väljer att inte gå djupare in i ämnet men konstaterar att anledningar som nämns av Brå är sociala faktorer, brister i mottagandet av invandrare och svårigheter att bosätta sig i ett nytt land (Martens & Holmberg 2005, s. 12-13).

Det har visat sig att andra studier antyder att befolkningstäthet har en påverkan på antalet brott. De menar att detta påverkar eftersom interaktionen mellan kriminella och potentiellt kriminella ökar vilket därmed ökar den senare gruppens risk att bli kriminell (Sachsida 2009, s. 97). Tidigare studier antyder också att hushåll med separerade föräldrar påverkar kriminaliteten i ett område (Sampson m.fl. 1997).

Vi tror att vi kan förvänta oss liknande resultat som tidigare studier har fått fram eftersom Sverige inte bör skilja sig nämnvärt från andra länder.

3 Datamaterial

Materialet vi har använt oss av kommer till största del från Statistiska centralbyråns (SCB) statistikdatabas. De gånger materialet inte finns att hitta på SCB går vi till andra källor. Data för försörjningsstöd finns att tillgå på Socialstyrelsens hemsida (Socialstyrelsen). Den beroende variabeln anmälda våldsbrott i Sveriges alla kommuner har vi hämtat från Brå:s statistikdatabas på deras hemsida (Brå 1).

Definitionen av anmälda brott är ”alla de händelser som anmälts och registrerats som brott hos polis, åklagare eller tull. De anmälda brotten inkluderar även händelser som efter utredning visar sig inte vara brott” (Brå 2011, s. 23). Till våldsbrott räknas bland annat mord, misshandel och rån.

I statistiken kan man även utläsa var i Sverige brottet begåtts (Sporre & Standar 2006, s. 51). Statistiken över anmälda brott registreras vid den tidpunkt som anmälan har kommit till myndigheternas kännedom, vilket innebär att brottet inte behöver ha utförts under samma år (Brå 2011, s. 36).

Vi har valt att inrikta oss på året 2010 eftersom det är det tidsmässigt mest aktuella året med fullständig statistik. För 2011 finns en stor del av statistiken, men vissa data saknas.

3.1 Inverkan på statistiken

Under denna rubrik vill vi visa att anmäld brottslighet kan ge en snedvriden bild av verkligheten eftersom vi stöter på problem i form av bland annat mörkertal och anmälningsbenägenhet.

3.1.1 Ändringar i statistikrutiner

Något man ska ta i beaktning som kan påverka statistikens utfall är ändringar i statistikrutiner. Plötsliga och oväntade nivåförändringar i brottsligheten beror oftast på ändrad lagstiftning eller statistikomläggningar och inte på att brottsligheten har ändrats (Sporre & Standar 2006, s. 19).

För vår del har detta inte inneburit några problem eftersom redovisningen av brottsstatistik i Sverige i princip gjorts på samma sätt sedan 1975 (Sporre & Standar 2006, s. 30).

3.1.2 Mörkertal

Mörkertalet är andelen inträffade brott som inte har kommit till rättsväsendets kännedom och därför inte är inräknad i statistiken. Med andra ord kvoten mellan den faktiska och den anmälda brottsligheten.

Den officiella statistiken redovisar antalet anmälda brott. Därför är det viktigt att mörkertalet är så litet som möjligt för att få en korrekt uppfattning av verkligheten. Om statistiken helt saknar mörkertal är den anmälda brottsligheten lika med den faktiska brottsligheten (Sporre & Standar 2006, s. 66-67).

Inom samma brottskategori, i vårt fall våldsbrott, kan mörkertalet variera mellan olika brottstyper. Detta fenomen kallas för selektivt mörkertal. Anledningen till att detta uppstår kan bland annat vara relationen mellan gärningsman och offer samt brottets art. I denna situation ger den anmälda brottsligheten delvis en felaktig bild av den faktiska brottsligheten (Sporre & Standar 2006, s. 66-67).

För vissa brottstyper kan andelen anmälda brott som inte betraktas som brott vara hög (Sporre & Standar 2006, s. 31). Enligt tidigare studier av mord, dråp och misshandel med dödlig utgång är andelen anmälda brott, som i realiteten visat sig inte vara ett brott, hela 50 procent (Sporre & Standar 2006, s. 31). Detta medför att man får med en hög andel brott i statistiken som egentligen inte bör räknas in i datamaterialet vilket bidrar till en risk för svedvriden data (Sporre & Standar 2006, s. 97).

3.1.3 Anmälningsbenägenhet

Detta leder oss in på vårt andra problem med datamaterialet, nämligen anmälningsbenägenheten. De två viktigaste anledningarna till att anmälningsbenägenheten är stor eller liten när det gäller brott mot person är brottets grovhet och det sociala bandet mellan gärningsman och offer (Sporre & Standar 2006, s. 71).

Våldsbrott anmäls i större utsträckning om offer och gärningsman är obekanta med varandra. Detta medför att bortfallet av våldsbrott är större i situationer där förövare och offer har en nära relation, vilket i sin tur leder till att statistiken representeras i högre utsträckning av brott där ingen relation föreligger. Benägenheten att anmäla brottet ökar också om handlingen är av den grövre sorten. Då har relationen mellan de inblandade mindre betydelse (Sporre & Standar 2006, s. 71-72).

Ett annat skäl till att anmälningsbenägenheten är låg och därmed leder till bortfall i statistiken kan bero på integritetsrelaterade orsaker så som vissa våldtäkter och våldsbrott med lite våld samt möjligheten att upptäcka brottet (Sporre & Standar 2006, s. 71-72). Anmälningsbenägenheten för brott kan påverkas av fler faktorer. Ytterligare exempel på anledningar till att avstå från att anmäla ett brott kan vara att offret inte orkar genomgå de processer som det innebär att utreda brottet (Sporre & Standar 2006, s. 73).

I vårt datamaterial tar vi inte tagit hänsyn till hur anmälningsbenägenheten skiljer sig åt mellan olika kommuner.

3.1.4 Regionala variationer

I områden där befolkningstätheten är stor anmäls det fler brott i relation till befolkningen än i mer glesbefolkade områden. I Malmö, Göteborgs och Stockholms kommun anmäls det fler brott per hundrausen invånare än i de flesta övriga kommuner i landet. Undantagen är kommuner där genomströmningen av människor är stor såsom Sigtuna som har en internationell flygplats och Helsingborg med färjetrafik (Brå 2011, s. 35)

Något som bör nämnas om vårt material är att det kan finnas ett rumsligt beroende mellan våra observationer. Detta innebär att värdet på en variabel för en kommun kan påverkas av kommuner som angränsar eller ligger nära kommunen i fråga (Svensson Henning 2009, s. 249). Med andra ord kan det finnas ett samband mellan brottsligheten i två närliggande kommuner eftersom de som begår brotten inte tar hänsyn till kommungränser. Om det till exempel bor många brottsbenägna personer i en kommun kan brottsligheten i närliggande kommuner påverkas. Det rumsliga beroendet är dock inget vi kommer gå in närmare på.

3.2 Variabler

Mycket av forskningen kring ojämlikhet och brott fokuserar på våldsbrott som beroende variabel. Ett stort problem är när man mäter våldsbrott med endast variabeln mord. Det menar Jankowski (1995, s. 88) kan vara behändigt, men att det finns en stor risk att skillnadernas inverkan på brottsligheten underskattas. Problemet när brott likställs med mord är att man blandar ihop en akt med ett resultat. Mord är alltså endast mord då resultatet blir att en person dödas. Samma akt mot två olika personer kan resultera i två olika utfall.

På grund av detta väljer vi att använda variabeln våldsbrott som beroende variabel. Våldsbrott mäts som antal anmälda brott per hundrausen invånare och år. Det finns olika definitioner om vad som inkluderas som våldsbrott (Brå 2011, s. 307). Brå:s definition av våldsbrott är dödligt våld, försök till mord eller dråp, misshandel, våldtäkt, grov kvinnofridskränkning, grov fridskränkning, olaga förföljelse, våld mot tjänsteman och rån (Brå 1). Det är även denna definition vi kommer använda oss av. En stor fördel med att använda våldsbrott istället för totalt antal brott är att våldsbrott i högre utstäckning anmäls (Sporre & Standar 2006, s. 69). Det medför att mörkertalet blir mindre med våldsbrott som beroende variabel. Samtidigt reflekterar anmälda våldsbrott anmäld total brottslighet väl vilket medför att detta blir ett bra mått på brottslighet.

De oberoende variabler vi väljer att titta närmare på utifrån vår teori och tidigare forskning är redovisade i tabell 1. Dessa är: *kön, ålder, inkomst, separationer, hälsa, försörjningsstöd, invandring, ojämlikhet, befolkningstäthet,*

utbildning och *arbetslöshet*. Det är dock inte säkert att alla dessa senare kommer ingå i modellen. En mer utförlig förklaring av dessa variabler går att finna i bilaga 1.

Tabell 1: Förklaring av variablerna

Variabel	Förklaring
<i>Beroende</i>	
Brott	Anmälda våldsbrott per 100 000 invånare
<i>Oberoende</i>	
Kön	Andel kvinnor
Ålder	Medelålder
Inkomst	Medelinkomst
Separationer	Antal separationer per 100 barn
Hälsa	Beräknat ohälsotal
Försörjningsstöd	Andel personer med försörjningsstöd
Invandring	Andel utrikesfödda
Ojämlighet	Gini-index
Befolkningstäthet	Antal invånare per kvadratkilometer
Utbildning	Andel med eftergymnasial utbildning
Arbetslöshet	Andel arbetslösa

Not: För en mer utförlig förklaring till variablerna se bilaga 1.

3.2.1 Felaktigt utelämnade variabler

För att vi ska få bästa möjliga förklarande modell vill vi undvika att utesluta variabler som tillför något till vår modell, det vill säga variabler som är signifikanta. Felaktigt utelämnade variabler leder till att vår regressionsfunktion inte är korrekt specificerad. För att undvika att detta oönskade fenomen uppstår kan man ta med alla variabler i modellen. Detta är å andra sidan en ganska dålig lösning eftersom det medför att även irrelevanta variabler inkluderas, vilket vi vill undvika (Westerlund, s. 157-158).

För att försäkra oss om att vi inte missar några relevanta variabler använder vi oss av metoderna Stegvis regression och Bästa delmängdsregression. Dessa metoder återkommer vi till i kapitel 4.

4 Metod

Vi vill undersöka vad som påverkar antalet anmälda brott i Sverige genom att jämföra alla 290 kommuner. Den metod vi använder för detta är multipel regression. Fördelen med metoden är att den kan anpassas till i stort sett varje situation där man vill förklara vad som påverkar en beroende variabel. Vid multipel regression används flera oberoende variabler, till skillnad från enkel regression där endast en oberoende variabel används (Djurfeldt 2009a, s.53; Djurfeldt 2009b s. 105). Vi tar hjälp av statistikprogrammet Minitab vid de flesta uträkningar. Då metoderna eller testen inte finns i programvaran genomför vi uträkningarna för hand.

Den linjära relationen mellan beroende och oberoende variabler fås genom (1). β -koefficienten anger effekten som den oberoende variabeln x har på modellen. Den tolkas som genomsnittlig förändring i den beroende variabeln då den oberoende variabeln ökar med ett, givet allt annat konstant. Residualen e anger den slumpmässiga delen som inte kan förklaras av modellen och α är interceptet (Westerlund 2005, s. 137-139). Man skattar α och β -koefficienterna med hjälp av minsta kvadratmetoden och väljer dessa så att summan av de kvadrerade residualerna blir minimal (Westerlund 2005, s. 75).

Vi kommer senare att bygga en modell som på bästa sätt kan förklara den beroende variabeln. Detta sker genom att förklara så stor del som möjligt av variationen i den beroende variabeln med våra oberoende variabler (Westerlund 2005, s. 67).

För multipel regression finns det ett antal antaganden som vi måste ta hänsyn till. Dessa sex antaganden är följande (Westerlund 2005, s. 72 & s. 139-149):

- 1) Modellen är linjär

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni} + e_i \quad (1)$$

- 2) Det förväntade värdet av residualerna är 0

$$E(e_i) = 0 \quad (2)$$

- 3) Residualerna är homoskedastiska, dvs har samma varians

$$\sigma^2 = Var(e_i) \quad (3)$$

- 4) Residualerna är inte autokorrelerade

$$Cov(e_i, e_j) = 0 \text{ om } i \neq j \quad (4)$$

5) De oberoende variablerna korrelerar ej

6) Residualerna är normalfördelade

$$e_i \sim N(0, \sigma^2) \quad (5)$$

De två första antagandena är till största del formella. Antagande ett säger egentligen endast att vår beroende variabel kan förklaras linjärt av våra oberoende variabler och feltermen. Samtidigt som det andra antagandet betyder att vårt stickprov är slumpmässigt utvalt (Westerlund 2005, s. 73). Dock har vi inget stickprov utan hela populationen av svenska kommuner.

Antagande tre och fyra är kanske tillsammans med antagandena sex de viktigaste. Det tredje antagandet säger att residualerna bör ha samma varians. Med andra ord betyder detta att osäkerheten kring observationernas medelvärde är lika stor för alla x-värden. Variansen påverkas alltså inte av vilket x-värde som antas. Om villkoret ej är uppfyllt råder heteroskedasticitet och det vill vi undvika (Westerlund 2005, s. 73). Vi testar om detta antagande är uppfyllt genom att grafiskt kontrollera efter ett beroende mellan residualerna i ett spridningsdiagram.

För att testa om heteroskedasticitet råder använder vi oss även av Whites test. Detta test är väldigt generellt vilket kan vara en nackdel. Dock kan det samtidigt vara en stor fördel då testet kan kontrollera för ett stort antal alternativa påverkansfaktorer (Verbeek 2012, s. 106-107). Hypoteserna lyder:

H_0 : Homoskedasticitet

H_1 : Heteroskedasticitet

Det finns olika varianter av detta test men vi väljer det som Westerlund framhåller (2005, s. 180-182). Denna variant utförs likt (6) genom en regression av modellens kvadrerade residualer mot modellens oberoende och kvadrerade oberoende variabler.

$$\hat{e}_i^2 = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_8 x_8 + \beta_9 x_1^2 + \beta_{10} x_2^2 + \dots + \beta_{16} x_8^2 + u_i \quad (6)$$

R^2 -värdet från denna regression multipliceras sedan med antalet observationer vilket därmed skapar testets observerade värde. Under nollhypotesen är värdet χ^2 -fördelat (Verbeek 2012, s. 106-109).

Det fjärde antagandet säger att varje observation kan ses som oberoende av de andra observationerna. Detta innebär också att residualerna inte ska vara beroende av varandra. Den ena residualen ska alltså inte påverka den andra. Om villkoret ej är uppfyllt är residualerna autokorrelerade (Westerlund 2005, s. 73). Autokorrelation är något som är vanligt i tidsseriedata och är således inget problem i vårt material (Djurfeldt 2009b, s. 115).

Det femte antagandet anger att de oberoende variablerna ej bör bero av varandra. Finns det ett systematiskt beroende mellan de oberoende variablerna råder multikolinjäritet. Det är svårt att helt undvika multikolinjäritet men den bör hållas på en sådan låg nivå som möjligt. För att testa detta kan man göra en enkel korrelationsmatris som anger Pearsons korrelation. Är detta värdet över 80 procent bör detta uppmärksammas (Westerlund 2005, s. 160).

Det är även möjligt att det råder multikolinjäritet mellan en grupp av oberoende variabler. Sådant upptäcks inte alltid i en korrelationsmatris. För att pröva för detta kan man använda sig av VIF-värdet (Variance Inflation Factors). Detta räknas ut genom att skatta en hjälpregression för varje oberoende variabel där denna agerar beroende variabel. I denna nya regression ställs den nya beroende variabeln mot övriga oberoende variabler. R^2 -värdet som utkommer av varje regression används sedan till att räkna ut VIF för varje oberoende variabel likt (7).

$$VIF = \frac{1}{1 - R^2} \quad (7)$$

En regression med VIF-värden på 1 saknar helt multikolinjäritet vilket såklart är eftersträfvansvärt (Westerlund 2005, s. 160). Rekommendationerna för vilket kritiskt värde som ska användas varierar kraftigt och vi har sett förslag mellan 2,5 och 10 (Djurfeldt 2009b, s. 114; Wooldridge 2009, s. 99). Vi väljer att inte vara extrem i någon riktning och sätter därför vårt kritiska värde till 5, vilket får räknas som en medelväg.

Det sjätte och sista antagandet är att residualerna ska vara normalfördelade. Det följer även att den beroende variabeln är normalfördelad (Wooldridge 2009, s. 172). För att testa om residualerna i modellen är normalfördelade kan vi således kontrollera om den beroende variabeln är normalfördelad. Det gör vi först genom att grafiskt kontrollera normalitetsantagandet i ett normalfördelningsdiagram.

Vi beräknar Anderson-Darlings test som kontroll för att normalitet råder i den beroende variabeln. Hypoteserna ser ut som följer:

$$\begin{aligned} H_0: y &\text{ är normalfördelad} \\ H_1: y &\text{ är ej normalfördelad} \end{aligned}$$

Ligger vårt p-värde under en signifikansnivå på 5 procent förkastas nollhypotesen och normalitet föreligger därmed inte i den beroende variabeln.

Råder inte normalitet i den beroende variabeln använder vi oss av Box-Cox-transformation. Metoden kan användas för att se vilken transformation som är att rekommendera, om det visar sig att den beroende variabeln inte är normalfördelad. Det bygger på att värdet λ räknas fram utifrån (8).

$$B(x, \lambda) = \begin{cases} \frac{x^\lambda - 1}{\lambda}, & \text{då } \lambda \neq 0 \\ \log x, & \text{då } \lambda = 0 \end{cases} \quad (8)$$

Utifrån det λ som framkommer kan vi välja vilken transformation som rekommenderas. Vid $\lambda=0,5$ ska den beroende variabeln transformeras med kvadratroten (Davidson & Mackinnon 2004, s. 440-443).

Vi väljer även att testa normalitetsantagandet direkt för residualerna genom att använda Jarque-Beras test. Detta test prövar om toppigheten och skevheten av residualernas fördelning liknar de för en normalfördelning (Westerlund 2005, s. 134). Jarque-Beras test skrivs som

$$JB = \frac{N}{6} \left(S^2 + \frac{(k-3)^2}{4} \right) \quad (9)$$

JB är testets observerade värde, S är ett mått på skevheten och k är ett mått på toppigheten av feltermernas sannolikhetsfördelning, där

$$S = \frac{1}{N} \sum \frac{\hat{e}_i^3}{\hat{\sigma}^3} \quad (10)$$

och

$$k = \frac{1}{N} \sum \frac{\hat{e}_i^4}{\hat{\sigma}^4} \quad (11)$$

Följande hypoteser gäller för testet:

$$\begin{aligned} H_0: e_i &\text{ är normalfördelad} \\ H_1: e_i &\text{ är ej normalfördelad} \end{aligned}$$

Under nollhypotesen är JB χ^2 -fördelat. Därmed bör det observerade JB-värdet vara nära noll för att residualerna ska vara normalfördelade. Det medför att S och k är noll respektive tre vid perfekt normalfördelning (Westerlund 2005, s. 135).

För att avgöra vilka variabler som ska ingå i vår modell använder vi oss av Stegvis regression och Bästa delmängdsregression.

Inom Stegvis regression finns det tre olika varianter av metoden. Vi väljer att använda standardversionen som tar bort och lägger till variabler utifrån signifikans i modellen. Resultatet ges av gränsen 0,15 som anger om variabeln ska inkluderas eller uteslutas ur modellen.

En nackdel med metoden är att den inte tar hänsyn till förklaringsgraden. Vi tar därför även hjälp av Bästa delmängdsregression för att få ytterligare en indikator på vilken modell som är bäst. Bästa delmängdsregressionen identifierar den bästa modellen genom att hitta de kombinationer med högst förklaringsgrad (R^2). Vi tittar även på R^2 -adjusted. Denna typ av förklaringsgrad justeras efterhand som fler oberoende variabler tillförs till modellen. Det justerade R^2 -värdet kan vara negativt om en modell har en väldigt liten förklaringsgrad. R^2 -adjusted är alltid mindre eller lika med R^2 (Davidson & MacKinnon 2004, s.117).

Värdet för Mallows CP är också något man bör ta i beaktning vid Bästa delmängdsregression. Det bästa möjliga värdet som kan antas är samma som antal

variabler inklusive interceptet i modellen. Om vi har en modell med tre variabler är det mest optimala värdet fyra för Mallows CP.

För att avgöra vilken modell vi ska välja jämför vi modellen från Stegvis regression och förslagen som Bästa delmängdsregression ger oss. När förklaringsgraden inte ökar tillräckligt mycket för varje ny variabel som tillförs till modellen finns det ingen anledning att ta med fler variabler. Vi tar även hänsyn till Mallows CP i vår urvalsprocess.

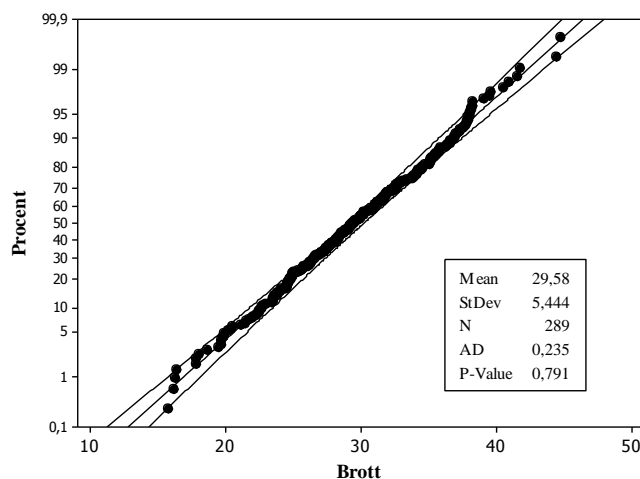
5 Resultat

Det resultat vi får fram av regressionen presenteras nedan. Vi försöker i detta kapitel finna svar på om socioekonomiska förhållanden påverkar brottsligheten i de svenska kommunerna och i så fall vilka. Till att börja med redovisar vi hur modellen tagits fram och om modellen lever upp till de antaganden som presenteras i avsnitt 4.

5.1 Modellering

Det finns en observation som kraftigt sticker ut från de andra och denna visar sig vara Dorotea kommun. Brottsligheten per hundra tusen invånare är mycket högre i Dorotea jämfört med övriga kommuner. År 2010 sticker också ut för Dorotea jämfört med statistik från kommunens övriga år. Vi hittar inget som förklarar denna kraftiga ökning utifrån den medierapportering vi tagit del av. Med stöd av detta och den residualanalys som vi genomfört finner vi bäst i att ta bort observationen. Detta borttagande av uteliggaren visar sig medföra stora fördelar till modellen.

Vi kontrollerar om den beroende variabeln är normalfördelad genom att ta fram ett normalfördelningsdiagram. Variabeln Antalet anmälda våldsbrott är inte normalfördelad. Vi använder oss därmed av Box-Cox metod för att avgöra vilken transformation av variabeln som är att rekommendera. Det intressanta är vilket värde på λ som metoden ger. Värdet vi får är 0,5 vilket gör att vi följer rekommendationen om att transformera den beroende variabeln med kvadratroten. Vi väljer därmed att åter kontrollera för normalitet.



Figur 1: Normalfördelningsdiagram för den beroende variabeln

Genom att grafiskt avgöra om den beroende variabeln är normalfördelad kontrollerar vi detta i ett normalfördelningsdiagram. Figur 1 tyder på att det råder normalfördelning. Detta styrks även av Anderson-Darlings test. Vi får ett p-värde på 0,791 vilket innebär att det är större än signifikansnivån på fem procent. Därmed förkastar vi inte nollhypotesen och testet visar på normalfördelning.

Vi kontrollerar för multikolinjaritet genom att studera korrelationen mellan de förklarande variablerna. I korrelationsmatrisen i bilaga 3 ser vi att det inte finns några variabler som korrelerar mer än 80 procent. Däremot korrelerar variablerna *utbildning* och *ojämlikhet* med nästan 80 procent (79,4 procent). På grund av detta misstänker vi att dessa variabler förklarar samma sak. Korrelationen med den beroende variabeln är större för *ojämlikhet* än för *utbildning*. Dessutom är den senare inte signifikant. Av de båda anledningarna misstänker vi starkt att *utbildning* bör tas bort. Vi väljer dock att inkludera båda variablerna tillsvidare för att titta på om VIF-värdet förstärker vår misstanke.

Som vi kan se av VIF-värdena i tabell 2 så ligger värdet över gränsen fem enbart för *utbildning*. Av korrelationsmatrisen att döma är det inget överraskande resultat att VIF-värdet är högt för just denna variabel. Enligt vår beslutsregel är det alltså endast en variabel som exkluderas på grund av för stark korrelation.

Tabell 2: Variance Inflation Factor samtliga variabler

Variabel	VIF
Kön	1,598
Ålder	2,567
Inkomst	2,970
Separationer	1,099
Hälsa	3,321
Försörjningsstöd	1,947
Invandring	2,398
Ojämlikhet	4,815
Befolkningstäthet	1,705
Utbildning	5,175
Arbetslöshet	2,305

Vi granskar vilka variabler som bör ingå i modellen. För detta använder vi oss av de båda metoderna Bästa delmängdsregression och Stegvis regression.

Vid stegvis regression i bilaga 4 föreslår metoden att vi tar med åtta variabler i vår modell. Dessa är: *invandring*, *kön*, *försörjningsstöd*, *separationer*, *hälsa*, *ålder*, *ojämlikhet* och *inkomst*. Därmed föreslås att *arbetslöshet* och *befolkningstäthet* utesluts ur modellen.

Vi tittar även på Bästa delmängdsregressionen. I bilaga 5 kan vi utläsa att det finns modeller med fler variabler som har högre förklaringsgrad. Detta är inte oväntat eftersom förklaringsgraden inte kan sjunka när vi använder fler variabler i modellen. När den bästa modellen tas fram finns det fler aspekter att ta hänsyn till än endast förklaringsgraden. Till exempel vill vi ha en så enkel modell som möjligt och ett lågt Mallows CP.

I Bästa delmängdsregressionen är värdet på R^2 -adjusted för åtta variabler det mest tillfredsställande med tanke på den minimala ökningen när ytterligare variabler tillkommer. Vi anser att förklaringsgraden inte ökar tillräckligt mycket när vi utökar modellen från åtta variabler och uppåt. Värdet på Mallows CP är nära det önskade när vi har med åtta variabler och förbättras endast marginellt med fler variabler. Däremot förbättras värdet betydligt när vi utökar modellen från sju till åtta variabler.

Utifrån vår jämförelse av förklaringsgrad och Mallows CP för de olika modellerna tillsammans med vår strävan efter en enkel modell, beslutar vi oss för att välja samma modell som den vi fick från stegvis regression. Variablerna i modellen blir därmed enligt tabell 3.

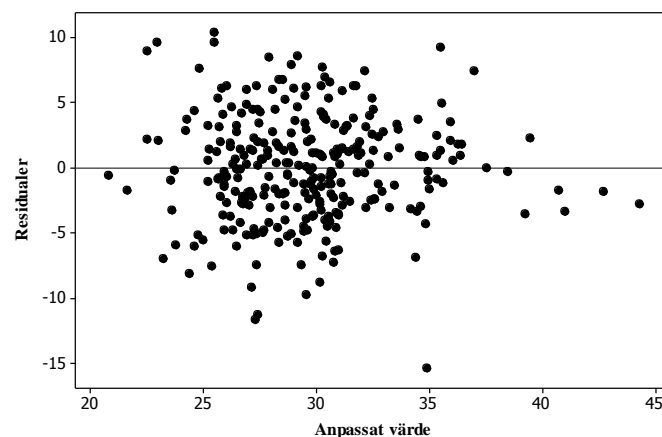
Tabell 3: Variabler i modellen

Brott (y)	Ålder (x_3)	Försörjningsstöd (x_6)
Inkomst (x_1)	Kön (x_4)	Invandrare (x_7)
Hälsa (x_2)	Separationer (x_5)	Ojämlighet (x_8)

5.2 Modellvalidering

Under denna rubrik testar vi för de tre viktigaste av de sex antagandena i avsnitt 4. Detta görs för att kontrollera att det råder homoskedasticitet, normalitet och att det inte råder en för hög multikolinjäritet.

Vi undersöker om residualerna är homoskedastiska. Detta gör vi genom att grafiskt kontrollera om det finns ett beroende mellan residualerna. Som vi ser av figur 2 finns inget beroende. Det finns inget som tyder på att residualernas varians inte är konstant.



Figur 2: Residualdiagram

För att även testa för homoskedasticitet använder vi oss av Whites test. Av testet i tabell 4 får vi fram det observerade värdet 27,166. Vårt kritiska värde är 27,59 i och med 17 frihetsgrader och en signifikans på fem procent. Det medför

att vi inte förkastar nollhypotesen som innebär att det råder homoskedasticitet i residualerna. Följaktligen är antagande tre uppfyllt.

Tabell 4: Whites test för homoskedasticitet

R ²	9,4%
White-statistika	27,166

Kontroll av multikolinjäritet görs genom att se till VIF-värdena för varje variabel. Vi har även tidigare visat att korrelationen mellan variablerna i bilaga 3 tyder på låg multikolinjäritet. Av tabell 5 kan vi utläsa att inget av VIF-värdena ligger över det kritiska värdet på fem. Därmed kan vi fastslå att multikolinjäriteten ligger på en nivå som vi accepterar.

Tabell 5: Variance Inflation Factor för modellen

Variabel	VIF
Kön	1,427
Ålder	2,450
Inkomst	2,734
Separationer	1,071
Hälsa	3,071
Försörjningsstöd	1,498
Invandring	1,958
Ojämlighet	3,487

Slutligen testar vi normalitetsantagandet för residualerna genom att använda Jarque-Beras test. Som vi ser av tabell 6 får vi ett lågt JB-värde. Vi kan därmed inte förkasta hypotesen om att normalfördelning råder i residualerna. Resultatet är inte speciellt förvånande utifrån normalitetstestet av den beroende variabeln.

Tabell 6: Jarque-Beras test av normalfördelning

k	3,162
S	-0,114
JB-statistika	0,948

5.3 Resultatredovisning

Det resultat vi får fram från regressionen gällande om socioekonomiska förhållanden påverkar brottslighet presenteras nedan. Vi analyserar även resultaten men understryker att vi inte gör anspråk på att helt förklara fenomenet utan vi resonerar och diskuterar endast kring dessa.

Som vi ser av tabell 6 har alla åtta variablerna en signifikant påverkan på den beroende variabeln. Alla variablerna är signifikanta på enprocentnivån. Det vill säga de oberoende variablerna påverkar den beroende variabeln med 99 procents säkerhet, vilket får anses bra.

Tabell 6: Regression

Variabel	$\hat{\beta}$	Medelfel
Konstant	- 67,20***	(19,90)
Kön	195,51***	(38,40)
Ålder	- 0,543***	(0,149)
Inkomst	- 1,631**	(0,560)
Separationer	0,802**	(0,275)
Hälsa	0,284***	(0,062)
Försörjningsstöd	58,61***	(17,14)
Invandring	24,19***	(6,230)
Ojämlighet	42,97**	(14,55)
R ²	43,9%	
N	289	

Not: Signifikant på 5 % (*), 1 % (**), 0,1 % (***).

Följande ekvation visar hur den beroende variabeln anmälda våldsbrott påverkas av de oberoende variablerna:

$$y_i = -67,2 + 196x_1 - 0,543x_2 - 1,631x_3 + 0,802x_4 + 0,284x_5 + 58,6x_6 + 24,2x_7 + 43,0x_8 \quad (12)$$

Vi ser nu till $\hat{\beta}$ -värdenas påverkan på modellen. I vårt fall är det svårt att analysera och tolka värdena på $\hat{\beta}$ eftersom den beroende variabeln är transformerad. De oberoende variablerna har dessutom olika enheter och kan inte jämföras sinsemellan. Vad vi däremot kan titta på och kommentera är vilket tecken $\hat{\beta}$ -värdet antar, alltså om variabeln har en positiv eller negativ påverkan på våldsbrott.

En variabel som påverkar andelen brott negativt är ålder. Det vill säga att kommuner med en hög medelålder har en lägre brottslighet. Det är också något som vi förväntar oss och beror möjligen på att personer väljer att flytta till mer lugna områden utanför storstäderna med åldern. Arbetsföra äldre personers starkare ekonomiska situation kan medföra att denna grupp har större möjligheter till ett tryggare boende.

Kommuner med en hög ohälsa har i större utsträckningen en högre brottslighet. Detta resultat kan möjligen sammankopplas med att personer med låginkomstyrken oftare har ett högre ohälsotal. Eventuellt medför de mer slitsamma låginkomstyrkena en högre ohälsa och att dessa personer på grund av sina ekonomiska förutsättningar bosätter sig i kommuner med hög brottslighet.

Samma gäller för en hög andel separationer. Det ligger i nivå med vad vi förväntar oss och beror troligen på att det kan vara svårare för barnet att växa upp

med skilda föräldrar. Ekonomiska förhållanden kan möjligtvis påverka denna variabel. Ensamstående föräldrar har ofta en sämre ekonomi vilket kan tvinga dessa att bostätta sig i mer brottstäta områden på grund av de lägre bostadskostnaderna.

Ojämlikhet påverkar brottsligheten genom att kommuner med stora inkomstskillnader i större utsträckning har en högre andel våldsbrott. Kommuner med en låg medelinkomst har samtidigt högre brottslighet. Vilken av dessa variabler som är viktigast i sammanhanget är svårt att säga. Man kan tänka sig att det är en kombination av de båda variablerna som skapar högst brottslighet. Möjligen gör de låga inkomsterna att incitamenten att begå brott ökar på grund av inkomstsituationen. Detta samtidigt som det finns mer välbärgade människor i närheten som gör dessa personer påmind om sin situation vilket kan leda till en utanförskapskänsla och frustration, tillsammans med att det finns ett utbud av brottsoffer. Det är dock svårt att säga vad som påverkar vad. Den höga brottsligheten i en kommun kan medföra att det blir mindre attraktivt att bosätta sig och driva sitt företag där. Dessa faktorer kan påverka att brottsligheten istället skapar ojämlika förhållanden och låg inkomst.

Att också kommuner med en hög andelen försörjningsstödstagare har högre brottslighet är inget som är förvånande. Det hänger troligen ihop med de ekonomiska incitamenten som tidigare nämnts.

I enlighet med vad vi förväntat oss visar resultatet även att kommuner med större andel invandrare har en högre brottslighet. Varför denna grupp är överrepresenterade är omtvistat och tidigare studier har försökt finna svar på detta. Trots att dessa kontrollerar för sociala påverkansfaktorer kvarstår detta resultat (Kardell 2011). Troligen hänger invandrades överrepresentation samman med personernas anledning till immigration samtidigt som att många invandrare bosätter sig i storstadskommuner med hög brottslighet.

Det finns dock en variabel som inte betar sig som vi förväntar oss. Resultatet antyder att kommuner med en högre andel kvinnor har en högre brottslighet. Vi förväntade oss dock motsatsen eftersom män i högre utsträckning än kvinnor begår brott. Resultatet kan möjligen vara ett skensamband och istället mäta det faktum att en högre andel kvinnor bor i storstäderna där fler brott begås. Samtidigt som andelen män är större i landsbygdskommunerna där mindre brott begås.

Hur mycket av variansen i den beroende variabeln kan förklaras av de oberoende variablerna? Utifrån vår undersökning kan vi inte uttala oss om hur stor förklaringsgrad varje enskild variabel bidrar med till modellen. Istället tittar vi på hela modellens förklaringsgrad. I jämförelse med tidigare studier står sig förklaringsgraden i vår undersökning väl och det visar sig även att vår modell i vissa fall har en högre förklaringsgrad än tidigare studier (Blau & Blau 1982; Sachsida m.fl.1997).

6 Slutsats

Brottsligheten i de svenska kommunerna påverkas av ett antal faktorer. Vi kan visa att socioekonomiska faktorer är en stor anledning till nivån av brottslighet i kommuner. De faktorer som påverkar kan härledas till ekonomiska anledningar i form av inkomst, skillnader i inkomst och andel av befolkning med försörjningsstöd.

Dessutom finner vi att påverkansfaktorer som ålder, invandring och kön inverkar. Dock förväntade vi oss motsatt effekt av den senare. Samtidigt finner vi stöd för att andelen separationer och hälsa påverkar nivån av brottslighet.

Detta får oss att efterfråga sociala insatser för att minska brottsligheten i Sverige. Genom att minska de negativa sociala faktorerna bör det enligt vårt resultat minska incitamenten till brott och därmed minska antalet brottslingar.

6.1 Framtida forskning

Det skulle vara intressant att forska vidare kring om samma signifikanta variabler kommer med i modellen, om vi istället för *Våldsbrott per hundra tusen invånare* använde *Brott per hundra tusen invånare* som beroende variabel. Vi kan därmed se om det finns en skillnad mellan typen av brottslighet.

Vidare skulle det vara spännande att undersöka om vi hade fått en annan modell om vi hade använt oss av data från tidigare år. Man skulle till exempel kunna jämföra om modellerna skiljer sig åt beroende på om vi använder data från 2010, 2000 eller 1990. På så sätt kan vi se om våldsbrott har påverkats av andra faktorer tidigare år än de vi har kommit fram till.

Ytterligare en möjlighet är att studera om våldsbrott per hundra tusen invånare påverkas av olika faktorer i olika stadsdelar inom storstäder i Sverige.

För att göra en närmare analys för enskilda typer av våldsbrott skulle det vara intressant att titta på om modellerna skiljer sig åt beroende på vilken typ av våldsbrott som används som beroende variabel. För vissa våldsbrott finns det dock för få observationer för att kunna göra en analys.

Man skulle även kunna tänka sig att göra en prediktion för utvecklingen av våldsbrott. Till exempel huruvida antalet våldsbrott har ökat eller minskat om tio år utifrån materialet vi använder oss av.

En bristfällighet med vårt datamaterial är att det inte tar hänsyn till mörkertalet. Detta kan ha påverkat vårt resultat genom att det blir snedvridet. Men troligen finns det inga större regionala skillnader gällande mörkertalet. Däremot kan man tänka sig att det rumsliga beroendet påverkar resultatet eftersom

brottsligheten inte tar hänsyn till kommungränser. Det är något som vore intressant att kontrollera för i framtida studier.

7 Referenser

- Blau, Judith R. & Peter M. Blau (1982). "The Cost of Inequality", *American Sociological Review* vol. 47, s. 114-129.
- Brå 1 = Brottsförebyggande rådets hemsida, Brott och statistik. [Elektronisk] Tillgänglig: <http://bra.se/bra/brott--statistik.html#&panel1-1>. Hämtdatum: 2013-01-07.
- Brå 2 = Brottsförebyggande rådets hemsida, Om Brå. [Elektronisk] Tillgänglig: <http://www.bra.se/bra/om-bra.html>. Hämtdatum: 2013-01-07.
- Brå (2011). *Kriminalstatistik 2010*. Stockholm: Brottsförebyggande rådet (Brå).
- Davidson, Russell & James G. MacKinnon (2004). *Econometric Theory and Methods*. New York: Oxford University Press.
- Djurfeldt, Göran (2009a) "Enkel regressionsanalys" i Djurfeldt, Göran & Mimmi Barmark (red.) *Statistisk verktyglåda 2: Multivariat analys*. Stockholm: Studentlitteratur, s 53-67.
- Djurfeldt, Göran (2009b) "Multipel regressionsanalys (MRA)" i Djurfeldt, Göran & Mimmi Barmark (red.) *Statistisk verktyglåda 2: Multivariat analys*. Stockholm: Studentlitteratur, s. 105-124.
- Hofer, Hanns von (2011). *Brott och straff i Sverige: Historisk kriminalstatistik 1750-2010: Diagram, tabeller och kommentarer*. 4 uppl. Stockholm: Kriminologiska institutionen, Stockholms universitet.
- Jankowski, Martin Sánchez (1995). "Ethnography, Inequality and Crime in the Low-Income Community", i Hagan, John & Ruth Peterson (red.) *Crime and Inequality*. Stanford, California: Stanford University Press, s. 80-94.
- Kardell, Johan (2011). *Utländsk bakgrund och registrerad brottslighet: Överrepresentationen i den svenska kriminalstatistiken*. Licentiatavhandling Stockholm: Stockholms universitet.
- Martens, Peter & Stina Holmberg (2005). *Brottslighet bland personer födda i Sverige och i utlandet*. Stockholm: Brottsförebyggande rådet (Brå).
- Sachsida, Adolfo, Mario Jorge Cardoso de Mendonca, Paulo R.A. Loureiro & Maria Bernadete Sarmiento Gutierrez (2010). "Inequality and Criminality Revisited: Further Evidence from Brazil", *Empirical Economics*, vol. 39, s. 93-109.
- Sampson, Robert J., Stephen W. Raudenbush & Felton Earls (1997). "Neighborhoods and Violent Crime: A Multilevel Study of Collective Efficacy". *Science* vol. 277, s. 918-924.
- SCB = Statistiska centralbyråns hemsida, statistikdatabasen. [Elektronisk] Tillgänglig: http://www.scb.se/Pages/SSD/SSD_TreeView____340478.aspx. Hämtdatum: 2013-01-07.

- Socialstyrelsen = Socialstyrelsens hemsida, Statistikdatabas. [Elektronisk]
Tillgänglig: <http://www.socialstyrelsen.se/statistik/statistikdatabas>. Hämt-
datum: 2013-01-07.
- Sporre, Tove & Robert Standar (red.) (2006). *Konsten att läsa statistik om brottslighet*. 3 uppl. Stockholm: Brottsförebyggande rådet (Brå).
- Svensson Henning, Martin (2009). ”Tekniker för analys av rumsliga data: En översikt” i Djurfeldt, Göran & Mimmi Barmark (red.). *Statistisk verktyglåda 2: Multivariat analys*. Stockholm: Studentlitteratur, s. 229-267.
- Verbeek, Marno (2012). *A Guide to Modern Econometrics*. 4 uppl. Hoboken, New Jersey: Wiley.
- Westerlund, Joakim (2005). *Introduktion till ekonometri*. Lund: Studentlitteratur
- Wooldridge, Jeffrey M. (2008). *Introductory Econometrics: A Modern Approach*. 4 uppl. Mason, Ohio: South-Western Cengage Learning.

8 Bilagor

Bilaga 1: Förklaring till variablerna

Arbetslöshet: Arbetslöshet i procent. Täljaren består av personer som någon gång under året registrerats som öppet arbetslösa i sökandekategori för öppen arbetslöshet. Nämnaren består av befolkningen i åldern 20-64 år. (SCB).

Befolkningstäthet: Antal invånare per kvadratkilometer. Vid beräkning av befolkningstätheten sätts befolkningen den 31 december i relation till landarealen den 1 januari året därpå. (SCB)

Brott: Antal anmälda våldsbrott per 100 000 invånare. I denna brottskategori ingår mord, dråp, barnadråp och misshandel med och utan dödlig utgång, våldtäkt inklusive grov våldtäkt, grov fridskränkning, grov kvinnofridskränkning, våld mot tjänsteman samt rån inklusive grovt rån. (Brå 1).

Försörjningsstöd: Andel biståndsmottagare inklusive barn. Individuellt behovsprövad insats som beslutas med stöd av socialtjänstlagen. Biståndet kan gälla försörjning (försörjningsstöd) eller livsföring i övrigt. Ekonomiskt bistånd (socialbidrag) är det begrepp som täcker all ekonomisk hjälp enligt socialtjänstlagen. Försörjningsstöd är en del av det ekonomiska biståndet och avser de vanligaste och regelbundna återkommande levnadskostnaderna. (Socialstyrelsen).

Hälsa. Ohälsotal. Täljaren består av antalet summerade nettodagar med sjukpenning, arbetsskadesjukpenning, sjukersättning, aktivitetsersättning och rehabiliteringsersättning under året. Nämnaren består av befolkningen i åldern 20-64 år. (SCB).

Inkomst: Medelvärdet av individens disponibla inkomst omräknat till antal prisbasbelopp. Ålder 20-64 år. (SCB).

Invandring: Andel utrikesfödda. Täljaren består av antalet personer födda utomlands. Nämnaren består av befolkningen. (SCB).

Kön: Andelen kvinnor i kommunen. Uppgifterna avser förhållandena den 31 december för valt år enligt den regionala indelning som gäller den 1 januari året efter. (SCB).

Ojämlighet. Gini-index. Mäter inkomstfördelningen inom kommunerna. Siffran är mellan 0 och 1 och baseras på kommunens nettoinkomst. Beskriver gapet mellan rika och fattiga, där 0 representerar jämlikhet och 1 representerar ojämlikhet (SCB).

Separationer: Antal separationer per 100 barn. Barn definieras som 0-17 år. (SCB).

Utbildning: Andel som har eftergymnasial utbildning, 3 år eller mer, åldersgrupp 16 år och uppåt. (SCB).

Ålder: Medelålder för befolkningen i kommunen. Uppgifterna avser förhållandena den 31 december för valt år enligt den regionala indelning som gäller den 1 januari året efter. (SCB).

Bilaga 2: Regression

Regression Analysis: Brott versus Inkomst; Hälsa; ...

The regression equation is

Brott = - 67,2 - 1,63 Inkomst + 0,284 Hälsa - 0,543 Ålder + 196 Kön
 + 0,802 Separationer + 58,6 Försörjningsstöd + 24,2 Invandrare
 + 43,0 Ojämlighet

Predictor	Coef	SE Coef	T	P	VIF
Constant	-67,20	19,90	-3,38	0,001	
Inkomst	-1,6313	0,5595	-2,92	0,004	2,734
Hälsa	0,28411	0,06274	4,53	0,000	3,071
Ålder	-0,5428	0,1492	-3,64	0,000	2,450
Kön	195,51	38,40	5,09	0,000	1,427
Separationer	0,8017	0,2750	2,91	0,004	1,071
Försörjningsstöd	58,61	17,14	3,42	0,001	1,498
Invandrare	24,186	6,230	3,88	0,000	1,958
Ojämlighet	42,97	14,55	2,95	0,003	3,487

S = 4,13522 R-Sq = 43,9% R-Sq(adj) = 42,3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	8	3747,53	468,44	27,39	0,000
Residual Error	280	4788,02	17,10		
Total	288	8535,55			

Bilaga 3: Korrelationsmatris

Korrelationsmatris

	Brott	Kön	Ålder	Inkomst	Hälsa	Försörjningsstöd	Separationer	Invandring	Ojämlighet	Utbildning	Arbetslöshet	Befolkningsstätt
Brott	1	,239	-,187	-,148	,088	,382	,183	,453	,230	,088	,319	,221
Pearson Correlation												
Sig. (2-tailed)		,000	,001	,012	,137	,000	,002	,000	,000	,136	,000	,000
N	289	289	289	289	289	289	289	289	289	289	289	289
Kön		1	-,366	,415	-,404	-,184	-,067	,098	,497	,599	-,181	,282
Pearson Correlation												
Sig. (2-tailed)			,000	,000	,000	,005	,257	,098	,000	,000	,002	,000
N	289	289	289	289	289	289	289	289	289	289	289	289
Ålder			1	-,493	,708	,146	,170	-,417	-,564	-,597	,353	-,345
Pearson Correlation												
Sig. (2-tailed)				,000	,000	,013	,004	,000	,000	,000	,000	,000
N	289	289	289	289	289	289	289	289	289	289	289	289
Inkomst				1	-,658	-,465	-,227	,079	,654	,709	-,570	,360
Pearson Correlation												
Sig. (2-tailed)					,000	,000	,000	,183	,000	,000	,000	,000
N	289	289	289	289	289	289	289	289	289	289	289	289
Hälsa					1	,357	,190	-,211	-,648	-,704	,542	-,344
Pearson Correlation												
Sig. (2-tailed)						,000	,001	,000	,000	,000	,000	,000
N	289	289	289	289	289	289	289	289	289	289	289	289
Försörjningsstöd						1	,179	,252	-,147	-,319	,639	-,081
Pearson Correlation												
Sig. (2-tailed)							,002	,000	,013	,000	,000	,123
N	289	289	289	289	289	289	289	289	289	289	289	289
Separationer							1	-,021	-,143	-,192	,265	-,047
Pearson Correlation												
Sig. (2-tailed)								,717	,015	,001	,000	,425
N	289	289	289	289	289	289	289	289	289	289	289	289
Invandring								1	,505	,215	,082	,429
Pearson Correlation												
Sig. (2-tailed)									,000	,000	,163	,000
N	289	289	289	289	289	289	289	289	289	289	289	289
Ojämlighet									1	,794	-,296	,498
Pearson Correlation												
Sig. (2-tailed)										,000	,000	,000
N	289	289	289	289	289	289	289	289	289	289	289	289
Utbildning										1	-,390	,536
Pearson Correlation												
Sig. (2-tailed)											,000	,000
N	289	289	289	289	289	289	289	289	289	289	289	289
Arbetslöshet											1	-,123
Pearson Correlation												
Sig. (2-tailed)												,037
N	289	289	289	289	289	289	289	289	289	289	289	289
Befolkningsstätt												1
Pearson Correlation												
Sig. (2-tailed)												
N	289	289	289	289	289	289	289	289	289	289	289	289

Bilaga 4: Stegvis regression

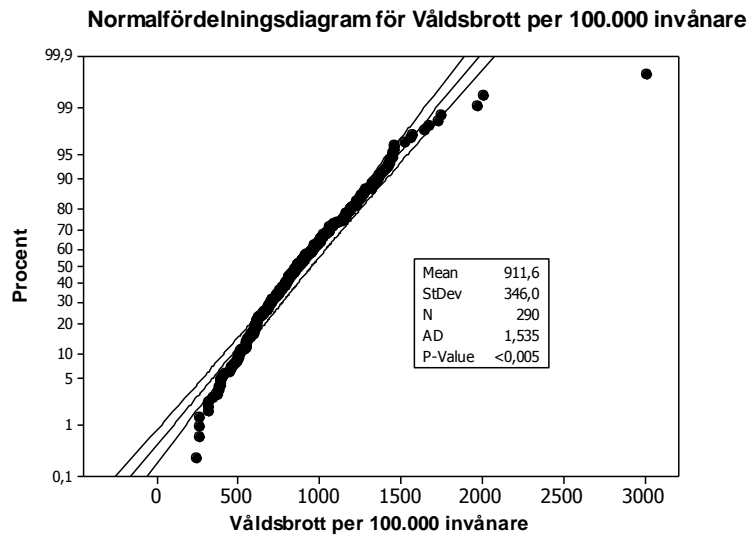
Stepwise Regression: Brott versus Inkomst; Hälsa; ...										
Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15										
Response is Brott on 11 predictors, with N = 289										
Step	1	2	3	4	5	6	7	8	9	10
Constant	24,66	18,02	-75,73	-80,48	-82,82	-100,91	-74,57	-69,35	-65,51	-67,20
Invandrare	45,1	42,7	39,8	35,6	36,2	39,8	33,4	27,9	24,0	24,2
T-Value	8,61	8,57	8,29	7,26	7,45	7,91	6,30	4,58	3,86	3,88
P-Value	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Arbetslöshet	0,90	0,90	1,05	0,66	0,56	0,34	0,40	0,41	0,29	
T-Value	5,65	5,65	6,82	3,40	2,87	1,61	1,87	1,93	1,35	
P-Value	0,000	0,000	0,000	0,001	0,004	0,108	0,062	0,054	0,177	
Kön			187	198	199	230	214	187	191	196
T-Value			5,32	5,70	5,79	6,33	5,93	4,80	4,95	5,09
P-Value			0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
Försörjningsstöd							54	55	46	59
T-Value							2,81	2,88	2,40	3,42
P-Value							0,005	0,004	0,017	0,001
Separationer							0,78	0,80	0,75	0,80
T-Value							2,77	2,85	2,71	2,91
P-Value							0,006	0,005	0,007	0,004
Hälsa							0,234	0,283	0,264	0,284
T-Value							3,96	4,38	4,11	4,53
P-Value							0,000	0,000	0,000	0,000
Ålder							-0,50	-0,52	-0,55	-0,54
T-Value							-3,35	-3,45	-3,71	-3,64
P-Value							0,001	0,001	0,000	0,000
Ojämlighet								23	41	43
T-Value								1,82	2,84	2,95
P-Value								0,070	0,005	0,003
Inkomst									-1,46	-1,63
T-Value									-2,56	-2,92
P-Value									0,011	0,004
S	4,86	4,62	4,41	4,34	4,30	4,26				
R-Sq	20,52	28,50	34,96	37,22	38,69	39,99	4,19	4,17	4,13	4,14
R-Sq (adj)	20,25	28,00	34,28	36,33	37,61	38,71	42,29	42,96	44,27	43,90
Mallows Cp	112,5	74,6	44,3	35,0	29,6	25,1	40,85	41,33	42,47	42,30
							15,6	14,3	9,7	9,5

Bilaga 5: Bästa delmängdsregression

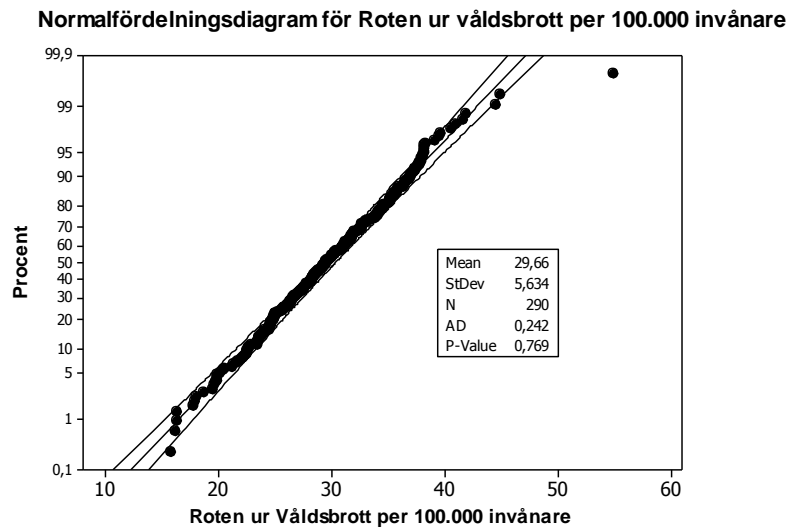
Response is Brott

Vars	R-Sq	R-Sq (adj)	Mallows Cp	S	t	a	t	r	n	r	t	d	e	t
1	20,5	20,2	113,9	4,8618										X
1	14,6	14,3	143,9	5,0410										X
2	28,5	28,0	75,9	4,6195									X	X
2	28,2	27,7	77,6	4,6303									X	X
3	35,0	34,3	45,4	4,4134						X		X	X	X
3	34,7	34,0	46,9	4,4236						X		X	X	X
4	37,7	36,9	33,5	4,3255	X					X		X	X	X
4	37,2	36,3	36,1	4,3439						X		X	X	X
5	39,6	38,5	26,1	4,2678	X	X				X		X	X	X
5	39,4	38,4	27,0	4,2739	X					X	X	X	X	X
6	41,6	40,3	18,3	4,2055	X	X	X			X	X	X	X	X
6	40,7	39,5	22,5	4,2359	X	X	X			X	X	X	X	X
7	42,3	40,9	16,7	4,1868	X	X	X	X		X	X	X	X	X
7	42,2	40,8	17,1	4,1900	X	X	X	X		X	X	X	X	X
8	43,9	42,3	10,6	4,1352	X	X	X	X	X		X	X	X	X
8	43,1	41,5	14,5	4,1642	X	X	X	X	X		X	X	X	X
9	44,3	42,5	10,6	4,1279	X	X	X	X	X	X		X	X	X
9	44,3	42,5	10,7	4,1291	X	X	X	X	X	X	X		X	X
10	44,6	42,6	11,0	4,1238	X	X	X	X	X	X	X	X		X

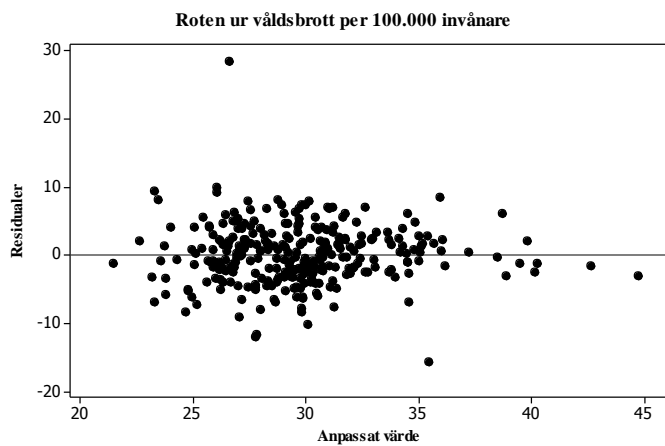
Bilaga 6: Figurer och diagram



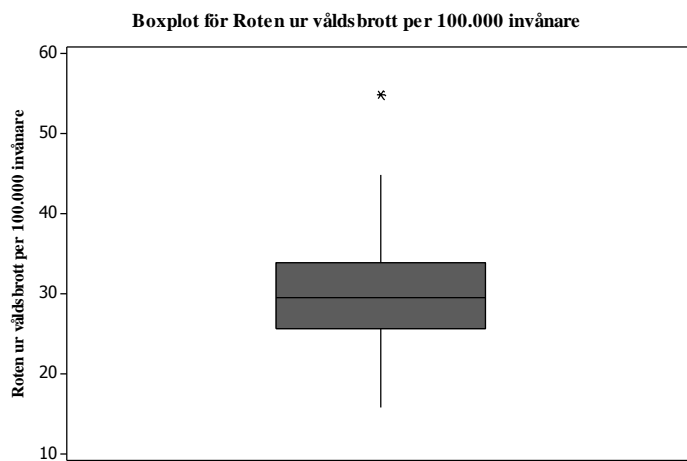
Normalfördelningsdiagram för våldsbrott per 100.000 invånare med Dorotea.



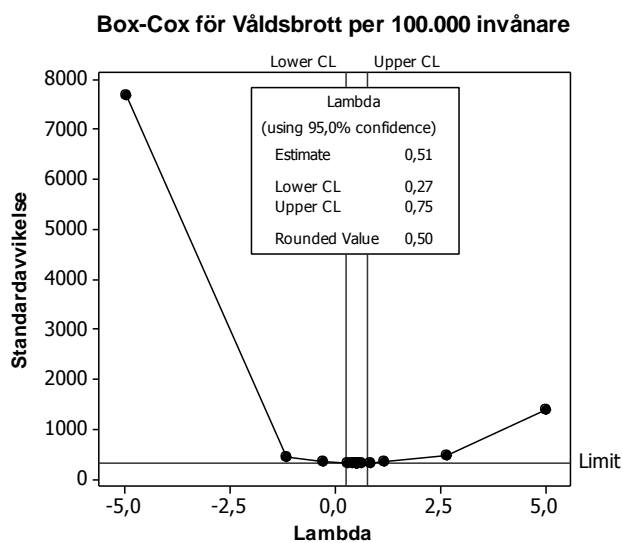
Normalfördelningsdiagram för roten ur våldsbrott per 100.000 invånare med Dorotea.



Residualplot med alla variabler med roten ur våldsbrott per 100.000 invånare som beroende variabel. Med Dorotea. Dorotea är pricken högst upp.

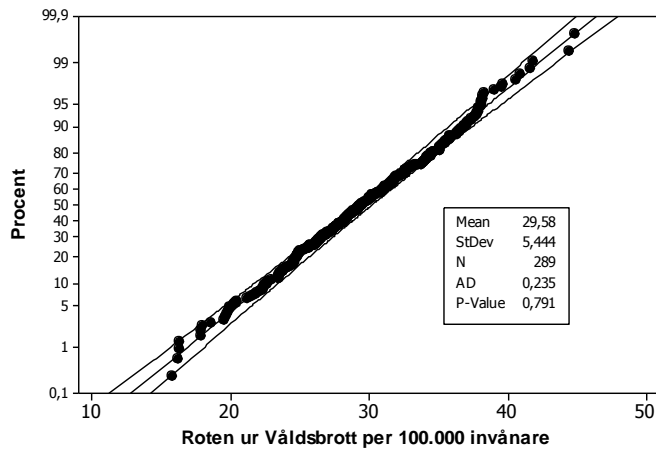


Boxplot för roten ur våldsbrott per 100.000 invånare med Dorotea. Pricken är Dorotea.



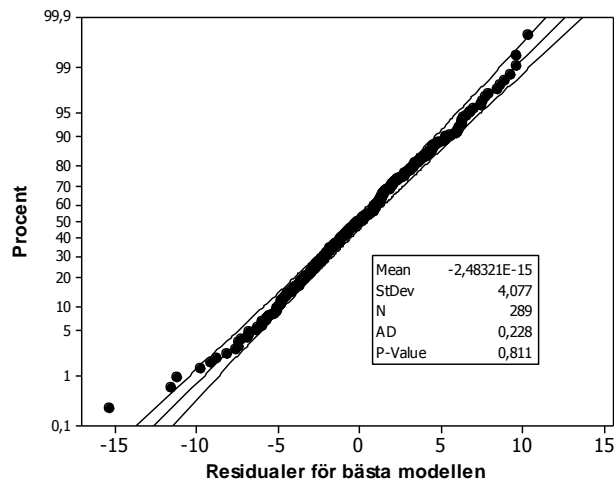
Resultat av Box-Cox. Utan Dorotea.

Normalfördelningsdiagram för Roten ur Våldsbrott per 100.000 invånare



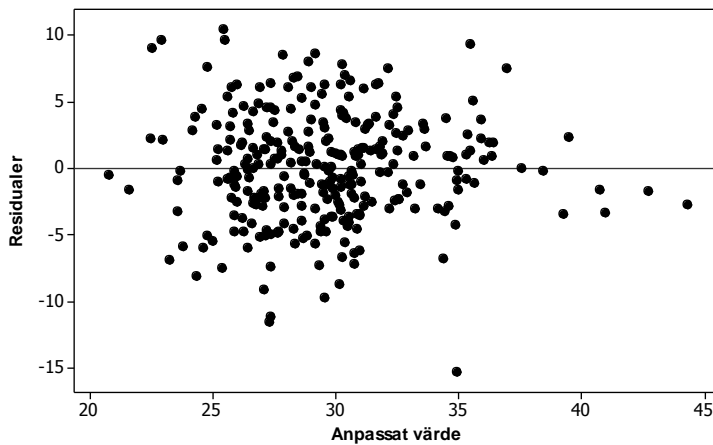
Normalfördelningsplot för variabeln Roten ur våldsbrott per 100.000 invånare. Utan Dorotea.

Normalfördelningsdiagram för residualer



Normalfördelningsdiagram av residualerna för vår bästa modell utan Dorotea.

Roten ur våldsbrott per 100.000 invånare



Residualplot för bästa modell utan Dorotea.