



Lunds universitet
Statistiska institutionen

Kostnadsutjämningsen inom Barnomsorgen i Sverige

- en diskriminantanalys

Rikke Berner
Pauline Persson

Uppsats i statistik
10 poäng
Nivå 41-60 poäng
September 2005

Handledare: Björn Holmquist

Abstract

Sweden's District councils and Country councils are in itself to a large extent responsible for the state financed welfare. Because of this, Sweden has a contribution- and adjustment system. The reason for this system is to create equal chance, for all District councils and Country councils, to be able to give the citizens service independently of their income and non-influential structural differences.

The adjustment is built upon the so-called standard cost method, which contains four different parts. These are called income-adjustment, cost-adjustment, general state contribution and rules of incorporation. The system is also designed to separate the standard cost of District councils and Country councils. The cost-adjustment for the District councils, in which we have chosen to look more closely into, is build upon a number of different sections. We have also restricted the investigation to the childcare section.

The structural cost differences for each section are calculated based upon factors that are able to reflect the structural needs- and cost-differences. Within the childcare section these factors are Tax levels, Levels of gainful employment amongst parents and Population density.

The purpose in this essay is to perform a discriminant analysis of the three factors in order to test if it is possible to, in advance, predict if a District council would win or lose within the cost-adjustment system. In other words, is it possible to predict if a District council's actual cost is bigger or smaller than its standard cost? We have also chosen to analyse four different years (2000-2003) in order to investigate if there exist similarities or differences in the results of the classification. The discriminant analysis is performed with three group variables (winners, neutrals and losers). After results from a Box test we adopt the method quadratic discriminant analyse.

Generally we got poor classifications. In spite of this, it was possible to show that most of the discriminations classified better than by chance. But still, they are not acceptable. There are no differences between the years. It is not possible to predict, in advance, if a District council is a winner or a loser.

Innehållsförteckning

ABSTRACT	2
INNEHÅLLSFÖRTECKNING	3
1 INLEDNING	4
1.1 BAKGRUND	5
1.2 SYFTE.....	5
1.3 DISPOSITION.....	6
1.4 AVGRÄNSNINGAR.....	6
2 DATA	7
2.1 DATA ÖVER SKATTEKRAFTSANDEL	7
2.2 DATA ÖVER BEFOLKNINGSTÄTHET	8
2.3 DATA ÖVER FÖRÄLDRARS FÖRVÄRVSFREKVENSS	8
3 METOD	9
3.1 GRUPPVARIABLER.....	9
3.2 MULTIVARIAT NORMALFÖRDELNING.....	9
3.3 KOVARIANSMATRISERNA.....	10
3.4 DISKRIMINANTANALYS.....	11
3.5 EVALUERING AV KLASSIFICERINGEN.....	11
3.6 CROSS-VALIDATION OCH RESUBSTITUTION.....	13
4 RESULTAT	14
4.1 INLEDANDE DISKUSSION.....	14
4.2 TOLKNING AV ANALYSEN.....	14
4.3 SAMMANFATTANDE RESULTAT	16
5 SAMMANFATTNING	18
APPENDIX 1	20
APPENDIX 2: ÅR 2000	21
APPENDIX 3: ÅR 2001	23
APPENDIX 4: ÅR 2002	25
APPENDIX 5: ÅR 2003	27
APPENDIX 6	29

1 Inledning

Idag tilldelar den svenska staten en avsevärd summa pengar till kommuner och landsting. Dessa resurser är kommunerna och landstingen själva ansvariga för att fördela mellan barnomsorg, äldreomsorg, grundskola, gymnasieskola och så vidare. Ju större del av den offentligt finansierade välfärden som kommuner och landsting själva ansvarar för, desto mer ökar behovet av utjämning av finansiella resurser. Detta för att förhindra att vissa kommuner tilldelas bidrag som inte är proportionella i förhållande till andra kommuner med samma behov. Med denna skevhet menas alltså, att om två kommuner har exakt samma behov av äldreomsorg tilldelas äldreomsorgen ändå inte samma kapital i de två kommunerna.

För att förhindra sådan skevhet har Sverige under många år haft ett statsbidrags- och utjämningsystem. Främsta syftet med systemet är att skapa likvärdiga förutsättningar för alla kommuner och landsting att kunna ge sina respektive invånare service oberoende av inkomstförhållande och skillnader som kommunerna inte själva kan råda över, så kallade strukturella skillnader.

Dagens statsbidrags- och utjämningsystem bygger på en metod som kallas standardkostnadsmetoden. Denna metod togs i bruk år 1996 och består av fyra delar; inkomstutjämning, kostnadsutjämning, generellt statsbidrag och införande regler. Systemet är även utformat som två separata delsystem, ett för kommuner och ett för landsting. Vi ska senare koncentrera oss på kostnadsutjämningsdelen inom kommunal barnomsorg, men först ska vi förklara mera ingående vad kostnadsutjämning innebär.

Syftet med kostnadsutjämningsdelen är att utjämna strukturella skillnader i behov och kostnader. Det är alltså inte möjligt för systemet att utjämna de faktiska kostnadsskillnaderna utan endast de skillnader som uppstår i behovet av kommunal service och de förutsättningar som finns för att producera sådan service. Med detta menas sådana skillnader som kommunerna inte själva kan råda över. Skillnader som beror på bland annat servicenivå och effektivitet kan alltså inte systemet utjämna, utan då en kommun har en högre servicenivå än andra kommuner måste kommunen själv finansiera detta med högre skatter eller avgifter.

Kostnadsutjämningsdelen är uppbyggd av ett antal delmodeller, som till exempel barnomsorg och äldreomsorg. De strukturella kostnadsskillnaderna för varje delmodell beräknas med hjälp av förklaringsvariabler som ska spegla de strukturella skillnaderna i behov och kostnader. Inom barnomsorgen antar man att skillnaderna beror på hur tätbefolkade kommunerna är, hur stor andel ensamstående eller deltidsarbetande föräldrar det finns samt hur skattekraftsandelen ser ut. Förklaringsvariablerna förklaras mera ingående i avsnitt 2.

För varje delmodell beräknas en kostnad, i kronor per invånare, per kommun respektive landsting. Denna kostnad kallas standardkostnad och beräknas utifrån de variabler som förklarar de skillnader vi beskrev i ovanstående stycke. I slutet av ett år är det sedan möjligt att beräkna den faktiska behovskostnaden, nettokostnaden. Nettokostnaden är

alltså det belopp en kommun eller ett landsting egentligen hade varit i behov av under året. Jämför man denna nettokostnad med standardkostnaden är det möjligt att se hur bra standardkostnaden beskrev behovet inom delmodellen.

Det har visat sig att Sverige i dagsläget har ett skevt statsbidrags- och utjämningsystem. Storstäder och förorter tenderar att tjäna på utjämnningen, då deras nettokostnader är mindre än deras respektive standardkostnad. Förlorarna är glesbygden och medelstora städer. Även om man på senare år har lyckats med en korrigering som minskar skevheten, då man har tagit fram en ny förbättrad modell (standardkostnadsmetoden) samt gjort förbättringar inom modellen, kvarstår fortfarande problem med skevheten.

Tidigare undersökningar visar på att skillnader i den sociala strukturen (till exempel andelen ensamstående och invånare med utländsk bakgrund) medför att storstäder oftast har ett större behov av offentlig finansiering. Kommuner med många enpersonshushåll och hög arbetslöshet skapar också ett tryck på den enskilda kommunens sociala budget. Skillnader i geografiska förutsättningar och klimatförhållanden kan innebära merkostnader, till exempel avseende väghållning och allmänt slitage.

För den läsare som söker mera kunskap om den svenska kostnadsutjämnings funktioner rekommenderar vi att läsa ”Kostnadsutjämnning för kommuner och landsting” (SOU 1998:151).

1.1 Bakgrund

Under ett antal år har regeringen, med hjälp av statistiska centralbyrån, försökt korrigera de beräkningsfel som finns inom den nuvarande standardkostnadsmetoden. När vi stötte på en problemformulering om att ta fram en ny förbättrad metod att beräkna standardkostnaden för delmodellen barnomsorg, på Statistiska Centralbyråns hemsida, ansåg vi att det vore mer intressant att analysera den nuvarande metoden och på så vis försöka lokalisera eventuella fel.

Då tidigare undersökningar pekar på att vinnarna är storstäderna och förlorarna är glesbygden frågade vi oss om det även var möjligt att definiera en vinnare eller en förlorare med avseende på de strukturella skillnader som finns.

1.2 Syfte

Syftet med uppsatsen är att med hjälp av diskriminantanalys, utförd med avseende på de faktorer som speglar kommunernas strukturella behovs- och kostnadsskillnaderna inom barnomsorgen, utvärdera om det är möjligt att förutspå om en kommun tenderar att vinna eller förlora på den kommunala kostnadsutjämnningen. Med detta menas alltså om det är möjligt att förutspå om en kommuns nettokostnader är större eller mindre än standardkostnaderna. Vi har även valt att analysera fyra olika år för att undersöka om det existerar eventuella likheter eller skillnader i beteendet hos klassificeringen.

1.3 Disposition

I avsnitt 2 behandlas och förklaras det insamlade datamaterial, varpå vi i avsnitt 3 går in på vilken metod vi har använt och bakomliggande teori. I det efterföljande avsnittet redovisar vi analysens resultat.

I Appendix återfinns en stor del av resultaten från analysen. Vi har valt att bifoga dessa för att inte ha med för mycket tabeller och figurer i den löpande texten.

1.4 Avgränsningar

Vi har valt att studera den kommunala kostnadsutjämnningen för barnomsorg under perioden år 2000 till år 2003 och utifrån dessa fyra år analysera klassificeringen till de olika populationerna. Undersökningen görs med hjälp av diskriminantanalys.

2 Data

Undersökningens data består av de tre förklaringsvariablerna Skattekraftsandel, Befolkningstäthet och Föräldrarnas förvärvsfrekvens. Dessa variabler fungerar som underlag för den nuvarande barnomsorgsmodellen och ämnar förklara de strukturella skillnader som finns kommunerna sinsemellan. Tillsammans skattar de det så kallade barnomsorgsbehovet. Detta behov benämnes Kvolym och står för kostnadsjusterad servicevolym och mäter kommunernas produktion av barnomsorg per barn. I Appendix 1 finns en utförlig förklaring till hur man beräknar standardkostnaden för Kvolym.

Vissa remissinstanser är mycket kritiska till att använda Skattekraftsandel och Befolkningstäthets som förklaringsvariabler eller som underlag för beräkningarna. Men i tidigare undersökningar har förklaringsgraden blivit väldigt liten när man enbart har skattat behovet av barnomsorgen med förvärvsfrekvensvariabeln (SOU 1998:151). Även då endast Skattekraftsandelen, som är den mest omdiskuterade av variablerna, elimineras är förklaringsgraden liten. Tillsammans fungerar dock de tre variablerna som goda förklaringsvariabler till det så kallade barnomsorgsbehovet.

Samtliga data har vi fått från Statistiska Centralbyrån (SCB). Förklaringsvariablerna beskrivs mera ingående var för sig i nedanstående delavsnitt.

Inledningsvis valde vi att arbeta med samtliga data från landets alla kommuner. Det visade sig dock att två kommuner, Knivsta och Gullspång, hade otillräcklig data. Detta resulterade i ett bortfall på två kommuner och gav oss slutligen 288 observationer. Vår data sträcker sig från år 2000 till och med 2003.

2.1 Data över Skattekraftsandel

Skattekraftsandelen är ett relativt värde. Här är värdena beräknade som ett medelvärde över tre år. Detta medelvärde är begränsat nedåt till 82 procent och begränsat uppåt till 127 procent. Med detta menas alltså att om en kommun har en Skattekraftsandel som är 75 procent, av Sveriges genomsnitt, sätt detta ändå till 82 procent. På samma vis blir alltså kommuner vars skattekraftsandel är större än 127 procent, av Sveriges genomsnittliga Skattekraftsandel, registrerade som 127 procent. Orsaken till denna begränsning är att undersökningar visar på ett svagt samband mellan barnomsorgsbehovet och Skattekraftsandel vid låga och höga nivåer på skattekraft.

Själva Skattekraftsandelen beräknas genom att skatteunderlaget enligt taxering divideras med folkmängden i respektive kommun. Folkmängden beräknas som antalet invånare vid taxeringsårets ingång. Resultatet divideras med det genomsnittliga skatteunderlaget per invånare i riket.

2.2 Data över Befolkningstäthet

Måttet på Befolkningstäthet (invånare/km²) beräknas vart femte år. Senaste beräkningsåret var år 2000. Detta leder till att denna variabel är oförändrad över alla åren i undersökningen.

Täthetsmättet beräknas enligt Statistiska Centralbyrån (2004) på följande sätt: Befolkningstäthet = (Andel som bor i tätort > 199 invånare)*(befolkning i tätort > 199 invånare) / (yta för tätorter i hektar > 199 invånare) / 100. Även detta mått är begränsat uppåt och neråt. Den övre gränsen är satt till 0,23 och den undre gränsen är satt till 0,05. Begränsningarna är gjorda så att man maximerar förklaringsgraden mellan barnomsorgsbehovet och de förklarande variablerna.

Eftersom faktorn Befolkningstäthet är beräknad från förhållandet år 2000 kan det leda till eventuella fel i analysen. Detta då denna förklaringsvariabel möjligtvis har ändrats under de senaste åren.

2.3 Data över Föräldrars förvärvsfrekvens

Föräldrars förvärvsfrekvens är den andel mammor eller ensamstående pappor med barn i åldern 1-9 år som arbetar eller studerar minst 28 timmar per vecka. Andelen beräknas som ett snitt över två år.

Anledningen till att det enbart är mammans förvärvsarbete som mäts när det kommer till samboförhållande är att förklaringsgraden blir högre mellan barnomsorgsbehovet och de oberoende förklaringsvariablerna.

3 Metod

För att utvärdera om en kommun tenderar att "vinna" eller "förlora" på den kommunala kostnadsutjämnningen inom barnomsorgen har vi valt att använda oss av diskriminantanalys. Diskriminantanalysen görs med avseende på de tre förklaringsvariablerna Skattekraftsandel, Befolkningstäthet och Föräldrars förvärvsfrekvens, som alltså tillsammans utgör grunden för beräkningar av barnomsorgsbehovet Kvolym.

De statistiska beräkningarna har vi valt att göra i SAS, SPSS och MINITAB.

3.1 Gruppvariabler

För att med hjälp av diskriminantanalys utvärdera om det är möjligt att förutspå om en kommun tenderar att vinna eller förlora på den kommunala kostnadsutjämnningen krävs att man har olika gruppvariabler som definierar dessa "vinnare" och "förlorare". Vi har valt att ha tre gruppvariabler. De kommuner vars standardkostnader per person är 500 kronor eller mer, än deras nettokostnader sätts som "vinnare". De kommuner vars standardkostnader är mindre än deras nettokostnader sätts som "förlorare". De kommuner som ligger i gråzonen mellan vinnare och förlorare sätts som "neutrala". De "neutrala" kommunerna har alltså standardkostnader som är 0-499 kronor högre än deras nettokostnader. Det ska betonas att detta är helt subjektiva bedömningar.

3.2 Multivariat normalfördelning

Enligt Marcoulides och Hershberger (1997) antar man vid de flesta statistiska teknikerna att den variabel man studerar, då det är viktigt för hypotestester, följer en multivariat normalfördelning. Trots att man vet att observerad data inte är exakt normalfördelad, är antagandet av normalitet ytterst viktigt.

För att diskriminantanalys ska ge så liten felklassificering som möjligt är det nödvändigt att på bästa sätt transfor mera de variabler man diskriminerar efter för att på så vis erhålla så bra normalfördelning som möjligt. För att testa vilken typ av transformation som man bäst kan transformera ickenormala serier efter föreslår Ramanathan (2002) ett så kallat Box-Cox test. Detta test ger oss ett värde, lambda, vilket transformationerna hädanefter bygger på.

$$y^* = y^\lambda$$

När man plottar två eller flera variabler mot varandra är det möjligt att få en bild av hur klassificeringen går till. Då varje gruppvariabel i sig bildar ett område av observationer, är det lämpligt att dessa områden inte till för stor del täcker varandra. Två överlappande områden är nämligen svåra att klassificera efter, då man inte kan dra en bra "skiljelinje" mellan dessa två. "Skiljelinjen" fungerar alltså som en separator mellan populationerna. Om variablerna är normalfördelade bildar dessa områden av observationer ellipser.

Ellipserna gör det lättare att hitta ett linjärt samband mellan variablerna och ger en mer korrekt "skiljelinje". En ny observation tilldelas den gruppvariabel som ligger närmast observationen.

3.3 Kovariansmatriserna

Enligt Manly (1994) föreslår författarna Srivastava och Carter att man använder ett modifierat Bartlett's test, ett Box test, när man ska utvärdera om det finns likheter mellan två eller flera multivariata kovariansmatriser. Detta test är dock tämligen känsligt när det kommer till antaganden om multivariat normalitet. Det finns därför alltid en möjlighet att det signifikanta resultatet beror på att stickproven inte var multivariat normalfördelade, istället för att populationerna har skillnader i sina kovariansmatriser.

När man testar sambandet mellan flera kovariansmatriser utgår man ifrån hypotesen

$H_0 : \Sigma_1 = \dots = \Sigma_k$ och testar denna mot H_1 : minst ett par av kovariansmatriserna är signifikant olika.

Enligt Morrison (1976) antas i Bartlett's test att multivariat normalitet med p -dimensioner finns. Den väntevärdesriktiga skattningen till S_i är S_i och baseras på n_i frihetsgrader, där $n_i = N_i - 1$. N_i betecknar antalet slumpmässigt valda observationsvektorer i den i :te populationen. Om H_0 är sann är S , den poolade skattningen till den gemensamma kovariansmatrisen, väntevärdesriktig. S beräknas enligt följande:

$$S = \frac{1}{\sum n_i} \sum_{i=1}^k n_i S_i$$

Teststatistiken för att testa nollhypotesen är

$$M = \sum n_i \ln |S| - \sum_{i=1}^k n_i \ln |S_i|$$

Detta är en generalisering av Bartlett's test av homogenitet hos varianserna. Det är denna storhet som är starkt beroende av att observationerna är multivariat normalfördelade.

Enligt Morrison (1976) har Box visat att skalfaktorn

$$C^{-1} = 1 - \frac{2p^2 + 3p - 1}{6(p+1)(k-1)} \left(\sum_{i=1}^k \frac{1}{n_i} - \frac{1}{\sum n_i} \right)$$

kan multipliceras med M vilket ger oss en storheten MC^{-1} som är approximativt χ^2 -fördelad med $\frac{1}{2}(k-1)p(p+1)$ frihetsgrader om n_i är större än tjugo och p och k inte överstiger fyra eller fem, vilket är fallet i vårt test.

Med andra ord ska vi alltså förkasta H_0 om $MC^{-1} > c_a^2(1/2(k-1)p(p+1))$, där p är antalet förklaringsvariabler och k är antalet gruppvariabler.

3.4 Diskriminantanalys

Diskriminantanalys är en multivariat teknik som syftar till att separera observationer till olika populationer med avseende på specifika beteenden i observationernas variabler. Efter att ha tagit fram en procedur som behandlar denna separation är det möjligt att allokera en helt ny observation till någon av populationerna.

Diskriminantanalys introducerades av R.A. Fisher (Johnson & Wichern, 2002, Kleinbaum et al, 1988).

Våra gruppvariabler får nu följande beteckningar; Vinnare (p_1), Neutrala (p_2) och Förlorare (p_3), och våra förklaringsvariabler; Skattekraftsandel (X_1), Befolkningstäthet (X_2) och Föräldrars förvärvsfrekvens (X_3).

Då vi söker efter en klassificeringsprocedur som minimerar risken för felbedömningar, tar vi med de så kallade apriorisannolikheterna i beräkningarna. Beteckningen för dessa är p_1 , p_2 respektive p_3 .

Beroende på om de olika populationerna har samma kovariansmatris eller inte använder man sig av olika klassificeringsprocedurer.

Om $\mathbf{X}_i \in N(\boldsymbol{\mu}_i, \mathbf{S})$, där $i=1, 2, 3$, så tilldelas \mathbf{x}_0 till p_i om

$$d_i = \max_{i=1,2,3} (d_i(\mathbf{x}_0)) \text{ där } d_i(\mathbf{x}_0) = \boldsymbol{\mu}_i' \mathbf{S}^{-1} \mathbf{x}_0 - \frac{1}{2} \boldsymbol{\mu}_i' \mathbf{S}^{-1} \boldsymbol{\mu}_i + \ln p_i$$

Denna metod kallas för linjär diskriminantanalys.

Om $\mathbf{X}_i \in N(\boldsymbol{\mu}_i, \mathbf{S}_i)$, där $i=1, 2, 3$, används kvadratisk diskriminantanalys, vilken klassificerar observationer efter;

$$d_i^Q(\mathbf{x}_0) = -\frac{1}{2} \ln |\mathbf{S}_i| - \frac{1}{2} (\mathbf{x}_0 - \boldsymbol{\mu}_i)' \mathbf{S}_i^{-1} (\mathbf{x}_0 - \boldsymbol{\mu}_i) + \ln p_i$$

Där \mathbf{x}_0 allokeras till p_i om

$$d_i^Q(\mathbf{x}_0) = \text{störst av } d_1^Q(\mathbf{x}_0), d_2^Q(\mathbf{x}_0) \text{ och } d_3^Q(\mathbf{x}_0)$$

3.5 Evaluering av klassificeringen

Ett sätt att evaluera en klassificering är enligt Johnson och Wichern (2002) att beräkna *the expected actual error rate* som betecknas $E(AER)$. Skattningen av denna kvot beräknas enligt;

$$\hat{E}(\text{AER}) = \frac{\sum_{i=1}^k n_{iM}^{(H)}}{\sum_{i=1}^k n_i}$$

där $n_{iM}^{(H)}$ är antalet missklassificerade observationer i den i :te populationen. I vårt fall är $k = 3$, vilket medför att $i = 1, 2, 3$.

För att klassificeringen ska vara godtagbar på 5 procentsnivån krävs alltså att $\hat{E}(\text{AER})$ är mindre eller lika med 0,05.

Ett annat sätt att evaluera om klassificeringen är lyckad är enligt Marcoulides och Hershberger (1997) att undersöka om den är bättre eller sämre än slumpen. Detta görs genom att man beräknar klassificeringens observerade "hit rate", H_o , och jämföra denna med ett "hit rate frequency expected by chance", H_c .

H_o = summan av diagonalelementen i klassificeringsmatrisen.

Med andra ord är alltså H_o antalet observationer som klassificerades rätt.

"The hit rate frequency expected by chance", H_c , beräknas enligt;

$$H_c = \frac{1}{k} \left(\sum_{i=1}^k N_i \right), \text{ där } k = \text{antalet gruppvaria bler.}$$

H_c är med andra ord antalet observationer som man sannolikt skulle ha klassificerat rätt om man slumpmässigt tilldelade alla observationerna en gruppvariabel. Formeln ovan används när man har lika apriorisannolikheter och kallas "the proportional chance criteria". Då man använder sig av olika apriorisannolikheter, exempelvis proportionellt viktade efter hur många observationer som tillhör respektive gruppvariabel, skattas efter H_c "the maximum chance criteria". Detta innebär att om man har tre olika apriorisannolikheter, till exempel 0,2, 0,3 och 0,5, används den högsta av dessa och multipliceras i sin tur med det totala antalet observationer.

Man kan pröva om klassificeringen är signifikant med;

$$z = \frac{H_o - H_c}{\sqrt{\frac{H_c(N - H_c)}{N}}} \text{ där } N \text{ är det totala antal observationer.}$$

För att klassificeringen ska anses signifikant ska z vara större än det kritiska värdet från en standardiserad normalfördelning.

Man kan också beräkna *IOCC* som står för "the improvement over chance criterium" och betecknar hur mycket bättre klassificeringen är i relation till slumpen. Förbättringen beräknas enligt följande;

$$IOCC = \frac{H_o/N - H_c/N}{1 - H_c/N}$$

För att klassificeringen ska anses godtagbart bör *IOCC* enligt Marcoulides och Hershberger (1997) vara större än 0,90.

3.6 Cross-validation och resubstitution

En deskriptiv diskriminantanalys kan genomföras på olika sätt. Vi har valt att jämföra två olika klassificeringsmetoder, cross-validation och resubstitution, enligt Afifi och Clark (1990).

Cross-validation, som även kallas för holdout eller jackknifing, genomförs genom att en observation utelämnas innan klassificeringsberäkningen. Denna observation klassificeras därpå efter de specifika beteenden som klassificeringen tog fram.

Klassificeringstekniken resubstitution använder sig, till skillnad från cross-validation, av alla observationerna när klassificeringen genomförs.

4 Resultat

Detta avsnitt inleds med en diskussion om vad vi initialt förväntar oss av analysen. Därefter beskrivs analysens resultat. I detta avsnitt hänvisar vi till appendix. Dessa appendix är i sin tur uppdelade efter år. Således utgör appendix 2 resultaten för år 2000, appendix 3 resultaten för år 2001 och så vidare. I appendix 6 kan läsare se vilka procar vi har använts av i SAS. Sist i avsnittet för vi ett mer allmänt resonemang om analysens resultat.

4.1 Inledande diskussion

Då det i tidigare undersökningar (SOU 1998:151) har visat sig att kommuner och landsting med många ensamstående och många arbetslösa tenderar att tjäna på utjämnningen torde vår diskriminantanalys med avseende på Föräldrarnas förvärvsfrekvens vara fördelaktig. Även Befolkningstätheten torde vara en bra klassificeringsvariabel då man har sett att storstäder och förorter, som en följd av många ensamstående och hög arbetslöshet, tjänar på kostnadsutjämnningen. Om skevheten har att göra med att man i de olika kommunerna betalar olika mycket skatt vore klassificeringsvariabeln Skattekraftsandel fördelaktig.

4.2 Tolkning av analysen

Då vi söker en så liten felklassificering som möjligt använder vi oss av Box-Cox test för att på så vis se vilka transformationer vi ska göra, på våra tre förklarande variabler, för att på bästa sätt omvandla våra ickenormala variabler. Variablerna transformeras således efter de resultat som går att utläsa i Box-Cox testen i appendix 2-5. Det finns inte några större skillnader i transformationerna mellan åren.

När vi plottar variablerna, både två- och tredimensionellt, mot varandra ser vi att det är mycket stor spridning bland observationerna. Detta betyder att det blir svårt att diskriminera efter våra tre förklarande variabler, både för våra transformerade och våra icke transformerade data. Vi kan även se, från scatterplottarna i appendix, att våra populationer till stor del överlappar varandra. Detta gör att det blir svårt att finna en bra skiljelinje som på bästa sätt separerar de olika populationerna. Allt detta pekar på att det blir problematiskt att diskriminera och klassificera med avseende på Skattekraftsandel, Befolkningstäthet och Föräldrars förvärvsfrekvens. Eftersom vi har svårt att utifrån plottarna avgöra om våra transformerade eller våra icke transformerade data bäst går att klassificera efter väljer vi att utföra analysen på båda dessa. Vi anser dock, med hänvisning till avsnittet om multivariat normalfördelning, att våra transformerade data borde ge bättre klassificering.

Vi har i undersökningen valt att utföra analyserna dels med lika apriorisannolikheter och dels med viktade sådana. Med viktade ariorisannolikheter menas i vårt fall att p_1 , p_2 och p_3 är proportionellt viktade efter hur många observationer som tillhör respektive gruppvariabel. Detta görs, som vi beskrev i metodkapitlet, för att minimera felklassificeringen.

Tittar man på kovariansmatriserna i tabell 1 för respektive år kan vi se att de ser olika ut. För att avgöra om olikheterna är signifikanta testar vi detta med Box test. Eftersom det i Box test väljs ett $\alpha = 0,10$ blir det kritiska området för $MC^{-1} > c_{10}^2[\frac{1}{2}(k-1)p(p+1)] = c_{10}^2(12) = 6,3035$. I tabell 1 ser vi att vi kan förkasta samtliga nollhypoteser om lika kovariansmatriser då alla våra MC^{-1} är betydligt större. Då kovariansmatriserna är olika utför vi därför kvadratisk diskriminantanalys.

Från tabell 1 kan vi se att klassifikationen med icke transformerade variabler ger bäst allokering vid användning av lika apriorisannolikheter. Resubstitution är överlag bättre än Korsvalidering. För transformerade värden är det svårt att utskilja något mönster i vad som anses som bästa allokering av Resubstitution och Korsvalidering respektive lika eller viktade apriorisannolikheter. Överlag visar det sig dock att de icke transformerade variablerna ger bättre klassificeringsresultat. Detta finner vi märkligt då variablernas normalfördelning är sämre än för de transformerade variablerna. Vi ser även att det inte finns några större skillnader mellan åren. De små skillnader som finns är att åren 2000 och 2003 klassificerar aningen bättre än åren 2001 och 2002 om vi ser på observerad "hit rate".

Eftersom vi har ett stort antal observationer använder vi oss av $\alpha = 0,01$, vilket medför att det kritiska området för $z > 2,3263$ när man testar om klassificeringen är bättre än slumpen. Detta z -värde finns bland annat i Körner (2000) tabell 3b. I tabell 1 nedan finns det ett fåtal z -värden, 8 stycken, som är mindre än vad z -värdet är på 1 procentsnivån. Dessa z -värden återfinns oftast för år 2003. Ett för litet observerat z -värde tyder på att klassifikationen inte är signifikant bättre än slumpen. Även om vi oftast har bättre klassificering än slumpen ser vi att alla värdena på IOCC är mycket låga, vilket tyder på att klassificeringen ändå inte är godtagbar.

Ett annat sätt att evaluera om en klassificering är bra är att beräkna *the expected actual error rate*. För att allokeringen ska vara godtagbar på 5 procentsnivån krävs alltså att $\hat{E}(AER)$ är mindre eller lika med 0,05. Värden på $\hat{E}(AER)$ från tabell 1, vilka ligger runt 0,5, är betydligt större än 0,05. Detta betyder att vi har så hög felklassificering att vi inte kan acceptera allokeringen.

Tabell 1: Sammanfattning av resultaten.

År	X-värde	MC ⁻¹	P-värde	Apriori	Resub/Cross	H ₀	H _c	IOCC	z	$\hat{E}(AER)$
2000	vanligt	84,94013	<0,0001	lika	Resub	149	96	0,276042	6,625	0,5158
	vanligt			lika	Cross	141	96	0,234375	5,625	0,5412
	transf	40,26653	<0,0001	lika	Resub	145	96	0,255208	6,125	0,5047
	transf			lika	Cross	131	96	0,182292	4,375	0,5541
	vanligt	84,94013	<0,0001	viktad	Resub	149	122	0,162651	3,22	0,4826
	vanligt			viktad	Cross	145	122	0,138554	2,743	0,4965
	transf	40,26653	<0,0001	viktad	Resub	144	122	0,13253	2,624	0,5
	transf			viktad	Cross	135	122	0,078313	1,55	0,4761
2001	vanligt	54,32043	<0,0001	lika	Resub	138	96	0,21875	5,25	0,534
	vanligt			lika	Cross	131	96	0,182292	4,375	0,5571
	transf	24,18707	0,0192	lika	Resub	135	96	0,203125	4,875	0,5349
	transf			lika	Cross	126	96	0,15625	3,75	0,5666
	vanligt	54,32043	<0,0001	viktad	Resub	137	105	0,163934	3,918	0,5243
	vanligt			viktad	Cross	132	105	0,147541	3,306	0,5417
	transf	24,18707	0,0192	viktad	Resub	135	105	0,163934	3,673	0,5312
	transf			viktad	Cross	125	105	0,10929	2,449	0,566
2002	vanligt	58,79618	<0,0001	lika	Resub	126	96	0,15625	3,75	0,5326
	vanligt			lika	Cross	123	96	0,140625	3,375	0,5411
	transf	26,73656	0,0084	lika	Resub	120	96	0,125	3	0,5718
	transf			lika	Cross	110	96	0,072917	1,75	0,6061
	vanligt	58,79618	<0,0001	viktad	Resub	132	117	0,087719	1,8	0,5417
	vanligt			viktad	Cross	122	117	0,02924	0,6	0,5764
	transf	26,73656	0,0084	viktad	Resub	136	117	0,111111	2,28	0,5278
	transf			viktad	Cross	121	117	0,023392	0,48	0,5149
2003	vanligt	67,75418	<0,0001	lika	Resub	156	96	0,3125	7,5	0,482
	vanligt			lika	Cross	148	96	0,270833	6,5	0,511
	transf	33,97943	0,0007	lika	Resub	138	96	0,21875	5,25	0,5285
	transf			lika	Cross	128	96	0,166667	4	0,5592
	vanligt	67,75418	<0,0001	viktad	Resub	145	126	0,117284	2,257	0,4965
	vanligt			viktad	Cross	143	126	0,104938	2,019	0,5035
	transf	33,97943	0,0007	viktad	Resub	158	126	0,197531	3,801	0,4514
	transf			viktad	Cross	147	126	0,12963	2,494	0,4896

4.3 Sammanfattande resultat

I avsnittets inledande diskussion antog vi att undersökningen skulle visa på en godtagbar klassifikation med avseende på våra tre förklaringsvariabler. Vi var lite skeptiska till faktorn Skattekraftsandel, men vi förutsatte ändå att de två andra variablerna skulle medföra en bra klassificering. Detta resonemang visade sig dock vara felaktigt.

Vi fick överlag dåliga klassificeringar. Visserligen kunde vi visa att de flesta diskrimineringarna klassificerade bättre än slumpen. Men när det kom till de övriga evalueringssätten "the expected actual error rate" och "the improvement over chance criteria" var klassificeringarna inte godtagbara. Det fanns inte heller några större skillnader mellan åren. Möjligtvis att åren 2000 och 2003 allokerade lite bättre än åren

2001 och 2002. Det är alltså inte möjligt att på förhand avgöra om en kommun är vinnare eller förlorare på kostnadsutjämnningen inom barnomsorgen i Sverige med avseende på de tre förklaringsvariablerna Skattekraftsandel, Befolkningstäthet och Föräldrars förvärvsfrekvens. Hade vi däremot funnit att en eller flera av våra förklaringsvariabler gav goda klassificeringar kan detta tyda på att det finns allvarliga fel i beräkningssättet av Kvolym.

Vilka förklaringsvariabler som är bättre att klassificera efter är svårt att avgöra eftersom inte statsbidrags- och utjämningsystemet kan jämföra ut de faktiska kostnaderna. Man tvingas därför hitta variabler som speglar de strukturella skillnaderna i behovet. Sådana skillnader är svåra att beräkna och vi anser ändå att de tre variablerna som finns speglar behovet relativt bra. Vad man kan göra, och även gör, är att utvärdera de befintliga variablerna och hitta kriterier som förbättrar förklaringsvariabeln. Detta kan till exempel göras genom att ändra begränsningarna uppåt och neråt samt beräkna variablerna efter andra kriterier. Även om förklaringsgraden kan öka till följd av ändrade kriterier är det ändå tveksamt huruvida sådana ändringar skulle förändra resultaten i vår diskriminantanalys.

5 Sammanfattning

Sveriges kommuner och landsting ansvarar till stor del själva för den offentligt finansierade välfärden. Därför finns det svenska statsbidrags- och utjämningsystemet. Syftet med detta system är att skapa likvärdiga förutsättningar, för alla kommuner och landsting i landet, att kunna ge sina respektive invånare service oberoende av inkomstförhållande och opåverkbara strukturella skillnader.

Standardkostnadsmetoden, som är det statsbidrags- och utjämningsystem som finns idag, består av fyra delar. Dessa är inkomstutjämning, kostnadsutjämning, generell statsbidrag och införande regler. Systemet är även utformat som två separata delsystem, ett för kommuner och ett för landsting. Den kommunala kostnadsutjämnningen, vilken vi har valt att se närmare på, bygger på ett antal delmodeller. Även här har vi inskränkt oss till att enbart se till delmodellen barnomsorg.

De strukturella kostnadsskillnaderna för varje delmodell beräknas med hjälp av faktorer som ska spegla de strukturella skillnaderna i behov och kostnader mellan kommunerna. Inom delmodellen barnomsorg är dessa faktorer Föräldrars förvärvsfrekvens, Skattekraftsandel och Befolkningstäthet. Med hjälp av dessa faktorer kan man beräkna en kommuns standardkostnad för delmodellen barnomsorg.

Syftet med uppsatsen är att med hjälp av diskriminantanalys, utförd med avseende på de förklaringsvariabler som speglar kommunernas strukturella behovs- och kostnadsskillnaderna inom barnomsorgen, utvärdera om det är möjligt att förutspå om en kommun tenderar att vinna eller förlora på den kommunala kostnadsutjämnningen. Med detta menas alltså om det är möjligt att förutspå om en kommuns nettokostnader (de faktiska kostnaderna) är större eller mindre än standardkostnaderna. Vi har även valt att analysera fyra olika år (2000-2003) för att undersöka om det existerar eventuella likheter eller skillnader i beteendet hos klassificeringen. Diskriminantanalysen har gjorts med tre gruppvariabler (vinnare, neutrala och förlorare). Utefter resultat från Box test använder vi kvadratisk diskriminantanalys.

Överlag fick vi dåliga klassificeringar. Vi kunde visserligen visa att de flesta diskrimineringar klassificerade bättre än slumpen, men de är ändå inte godtagbara. Det fanns inte heller några större skillnader mellan åren.

Det är alltså inte möjligt att i förväg avgöra om en kommun är vinnare eller förlorare avseende Skattekraftsandel, Befolkningstäthet och Föräldrars förvärvsfrekvens.

Litteraturförteckning

Affi, A.A. och Clark, V. (1990) *Computed-Aided Multivariate analysis*, Van Nostrand Reinhold, New York.

Johnson, R.A. och Wichern, D.W. (2002) *Applied Multivariate Statistical Analysis*, Prentice-Hall, New Jersey.

Kleinbaum, D.G., Lawrence, L.K. och Keith, E.M. (1988) *Applied Regression Analysis and Other Multivariable Methods*, PWS-KENT Publishing Company, Boston.

Körner, S. (2000) *Tabeller och formler för statistiska beräkningar*, Studentlitteratur, Lund.

Manly, B.F.J. (1994) *Multivariate Statistical Methods*. Chapman & Hall, London.

Marcoulides, G.A. och Hershberger S.L. (1997) *Multivariate Statistical Methods, a first course*, Lawrens Erlbaum Associates, New Jersey.

Morrison, D.F. (1976) *Multivariate Statistical Methods*. McGraw-Hill, Tokyo.

Ramanathan, R. (2002) *Introductory Econometrics with Applications*, South-Western, Australia.

SCB, (2004) *Kommunalekonomisk utjämning för kommuner*, SCB, Örebro (bilaga 5a)

(SOU 1998:151) *Kostnadsutjämning för kommuner och landsting*
(<http://www.regeringen.se/sb/d/126/a/136/action/search/type/simple>) senast uppdaterad 2004-05-05

Appendix 1

Standardkostnaden för förskoleverksamhet och skolbarnomsorg, som tillsammans bildar delmodellen barnomsorg, beräknas med avseende på kommunens åldersstruktur och andra faktorer som förklarar skillnader av barnomsorg mellan kommuner.

$$\text{Standardkostnad} = \text{Korrigerad åldersersättning} * \text{Kvolym-index}$$

$$\text{Korrigerad Åldersersättning} = (42\,827 * \text{andel barn 1-3 år} + 41\,863 * \text{andelen barn 4-5 år} + 26\,547 * \text{andelen 6 åringar} + 15\,273 * \text{andelen barn 7-9 år}) * \text{Justeringsfaktorn åldersersättning}$$

Med korrigerad åldersersättning menas att åldersersättningen är åldersjusterad efter den genomsnittliga vägda åldersersättningen i riket. Denna överrensstämmer med den vägda genomsnittliga nettokostnaden för barnomsorgen.

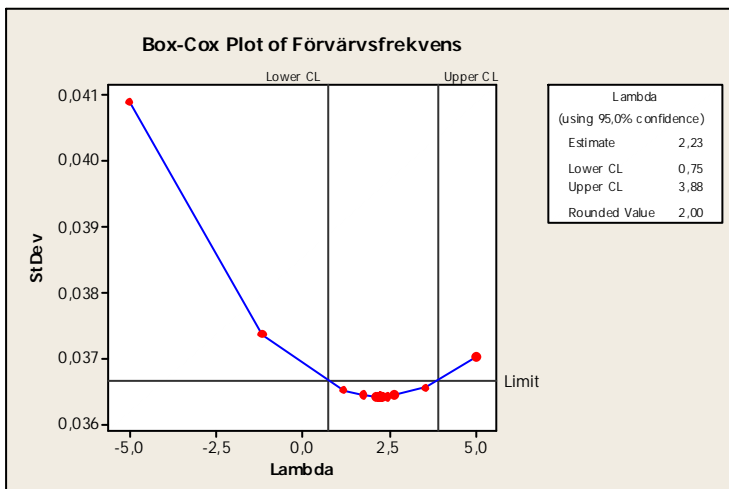
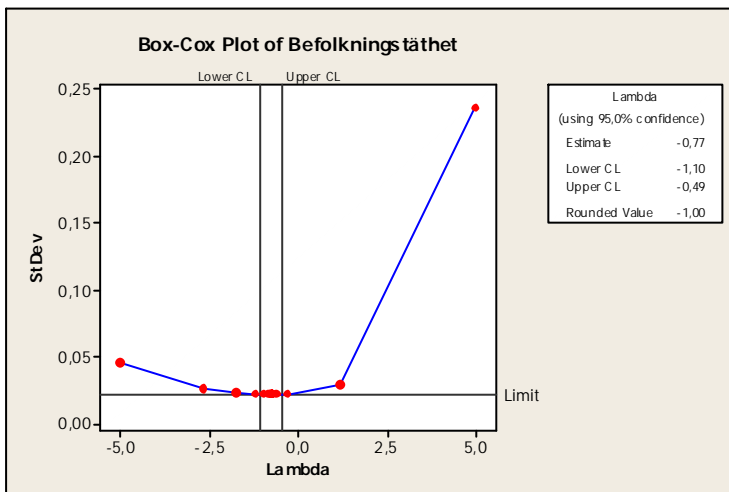
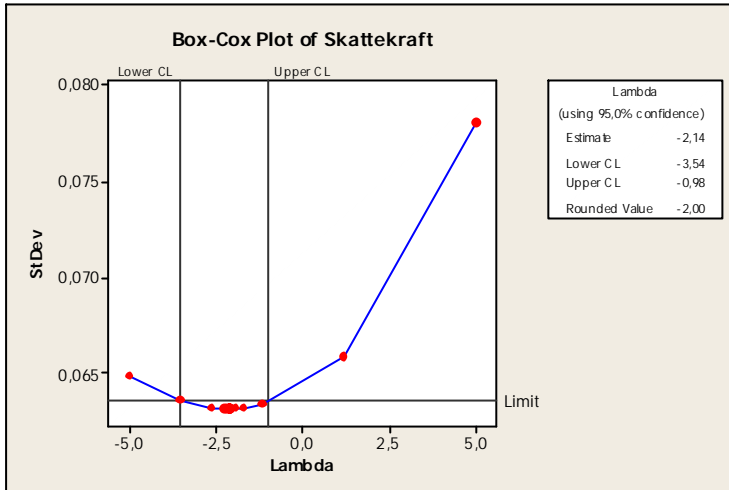
Behovet av de andra faktorerna som förklarar skillnader i barnomsorg skattas med en regressionskoefficient, det så kallade Kvolym. SCB (2004) har tagit fram en modell som beräknar Kvolym enligt;

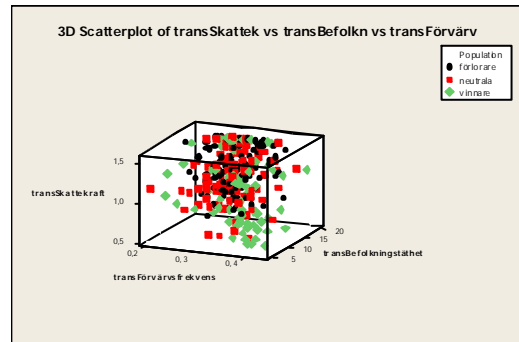
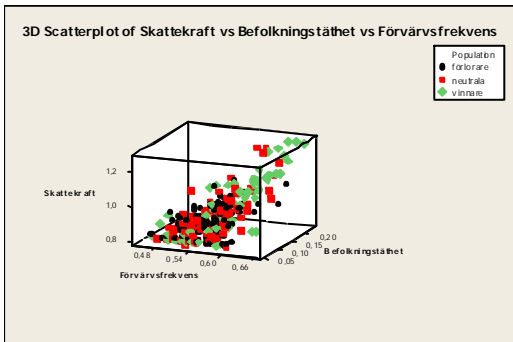
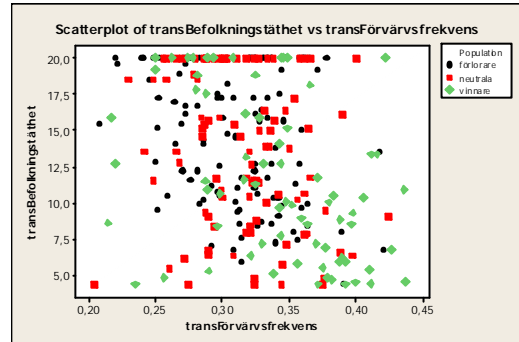
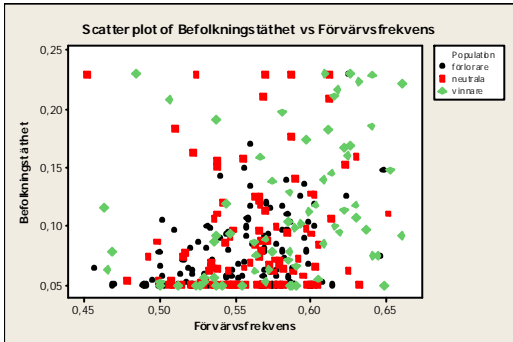
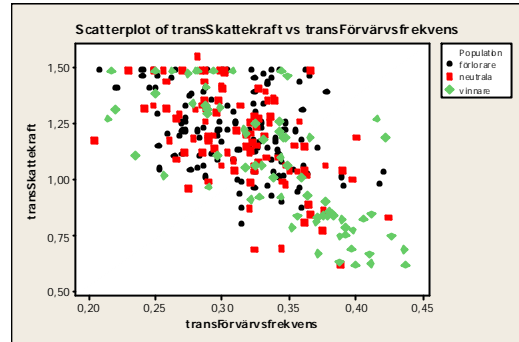
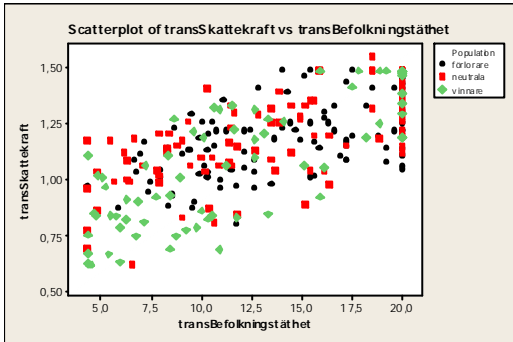
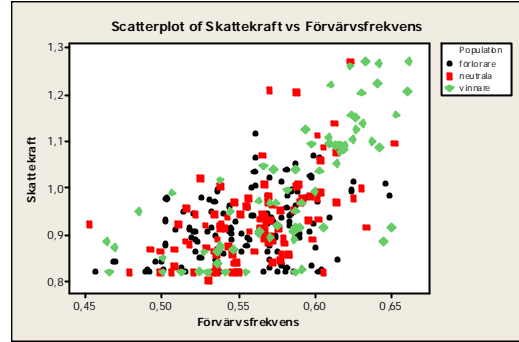
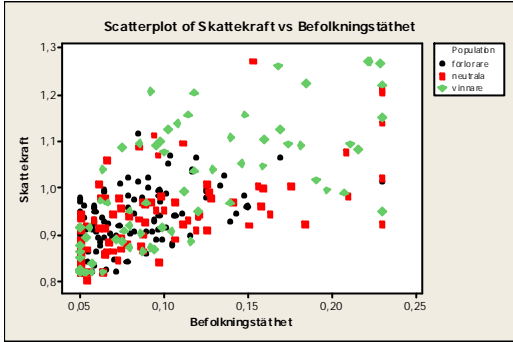
$$\text{Kvolym} = 0,0226 + 0,1944 \cdot \text{Skattekraftsandel} + 0,4025 \cdot \text{Befolkningstäthet} + 0,1158 \cdot \text{Föräldrars förvärvsfrekvens}$$

$$\text{Kvolym-index} = \frac{\text{Kvolym}}{\text{Justeringsfaktorn för Kvolym}}$$

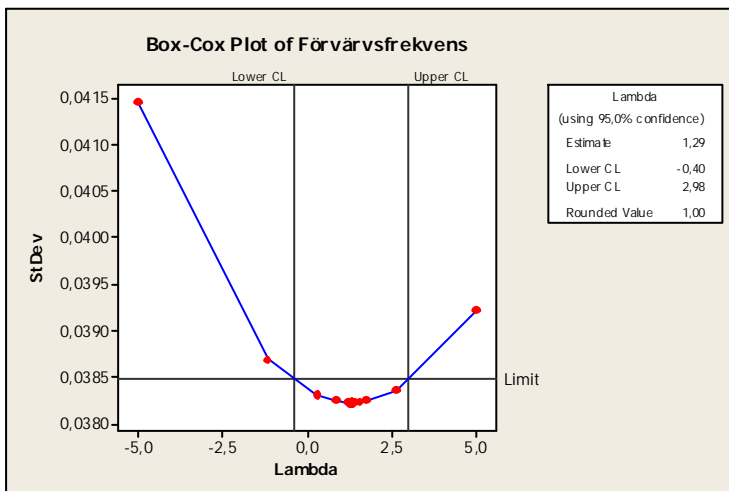
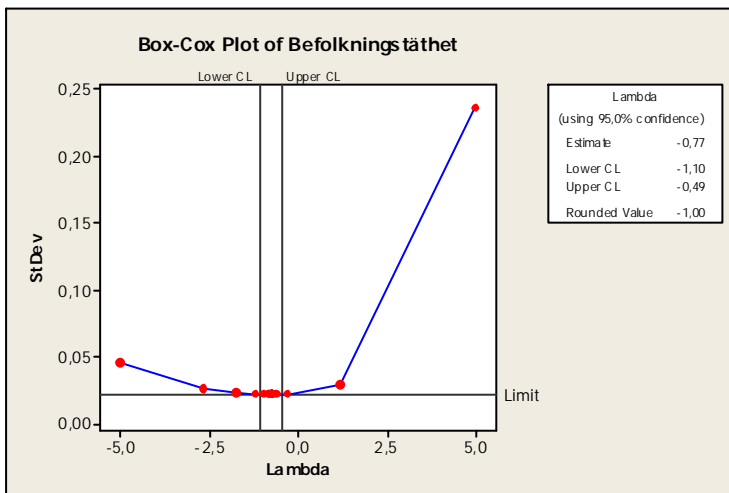
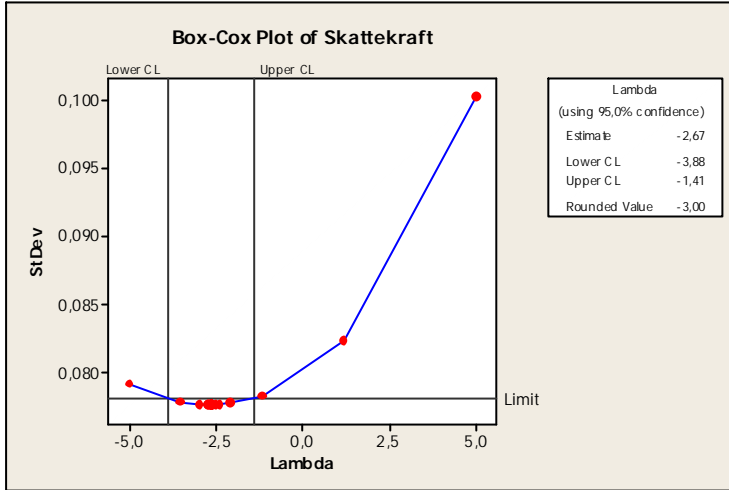
Justeringsfaktorn för Kvolym är det vägda medelvärdet, av Sveriges alla kommuner, av Kvolym.

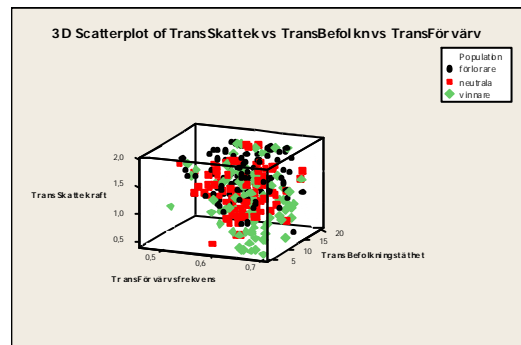
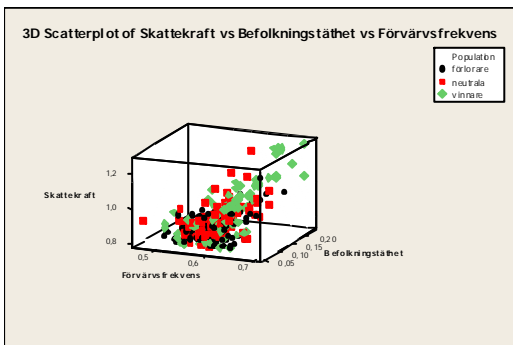
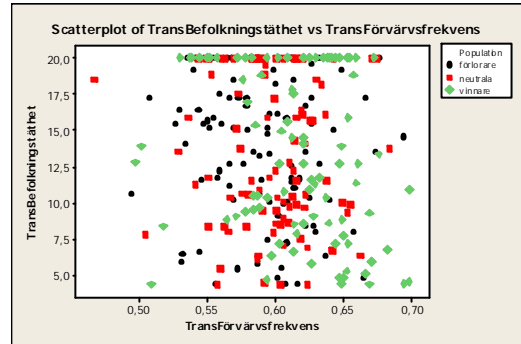
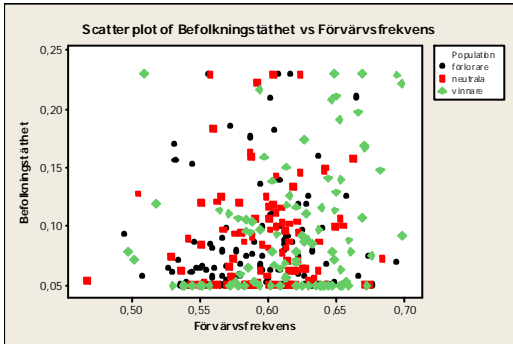
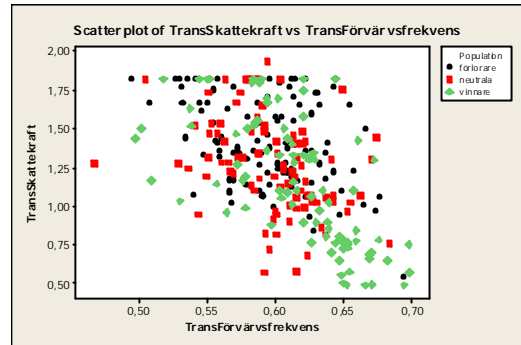
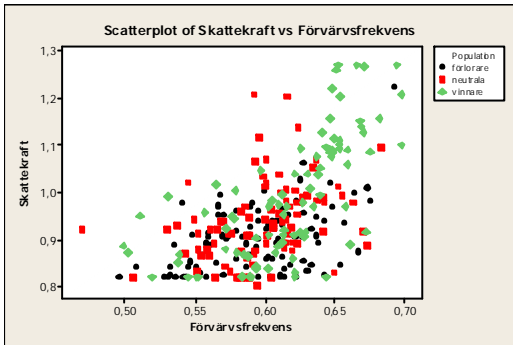
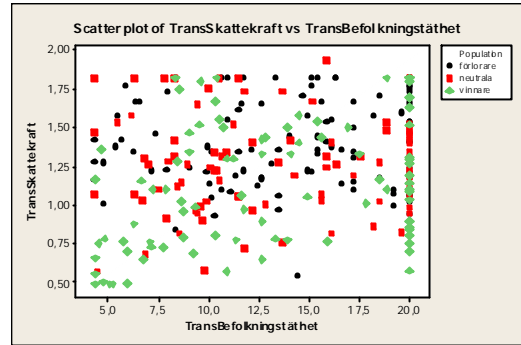
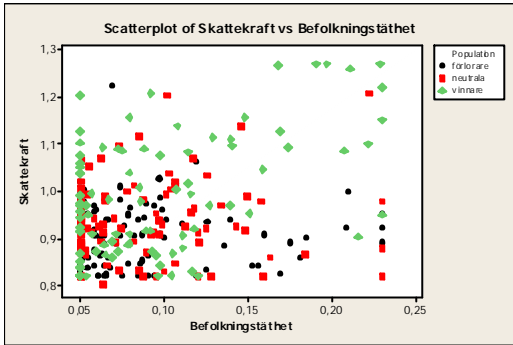
Appendix 2: År 2000



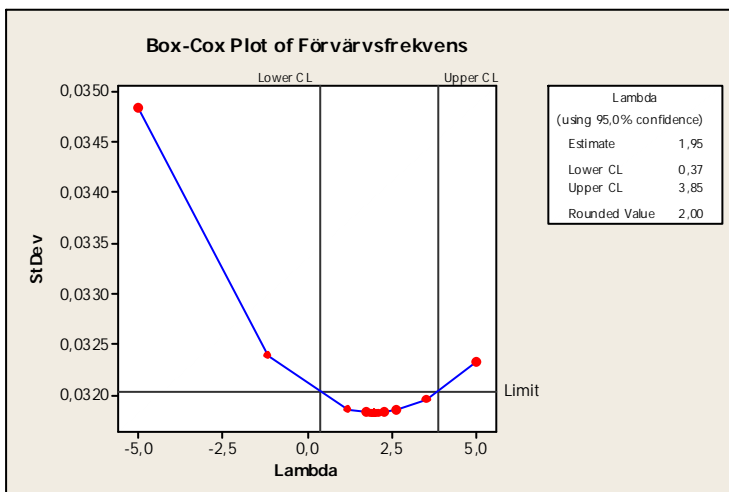
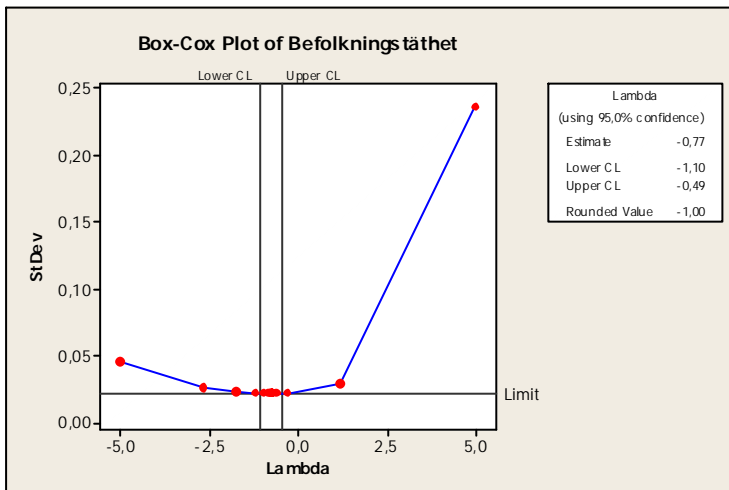
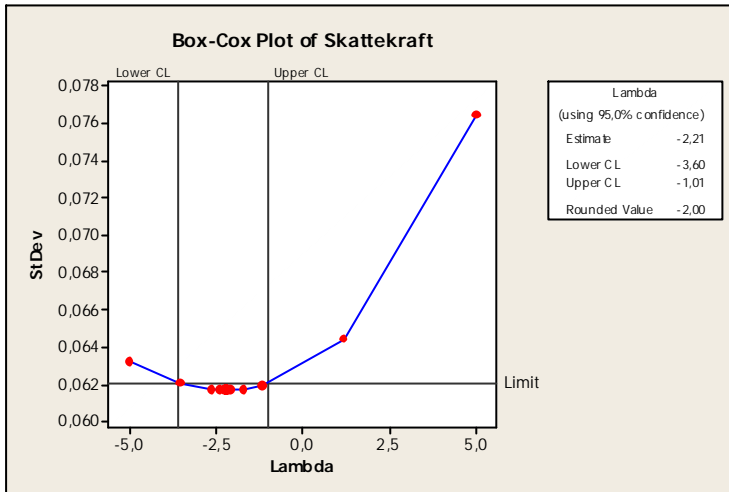


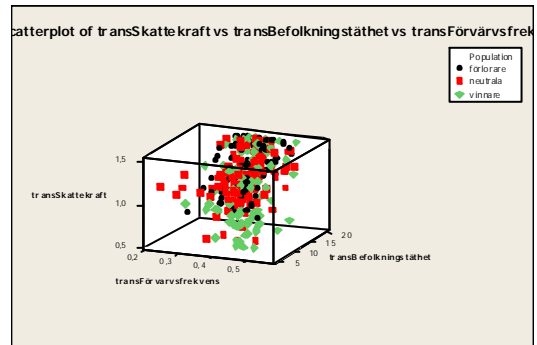
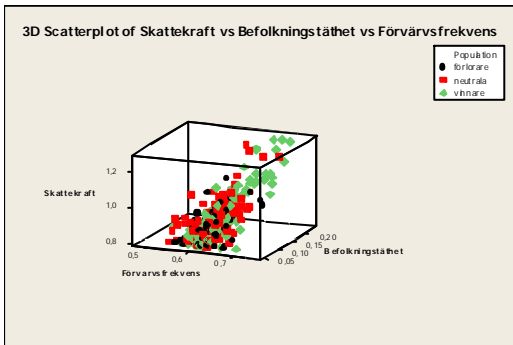
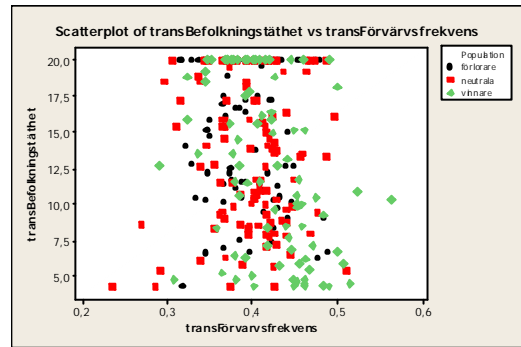
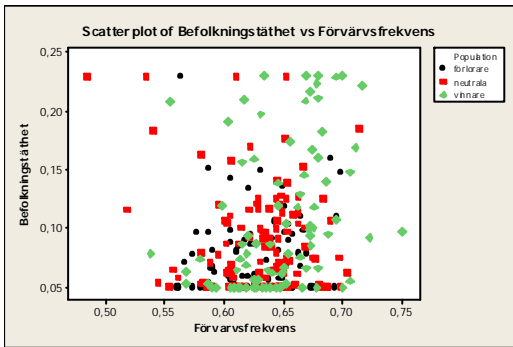
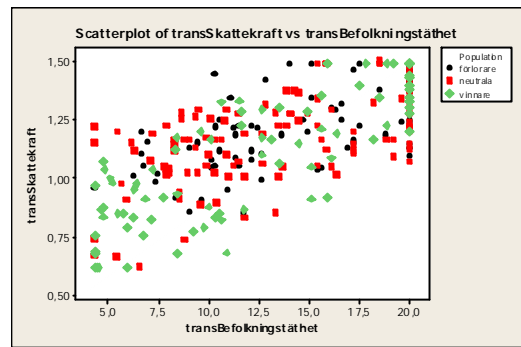
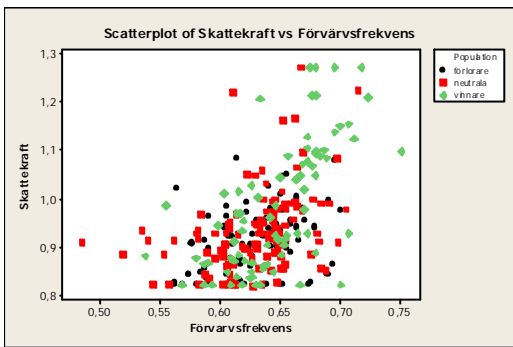
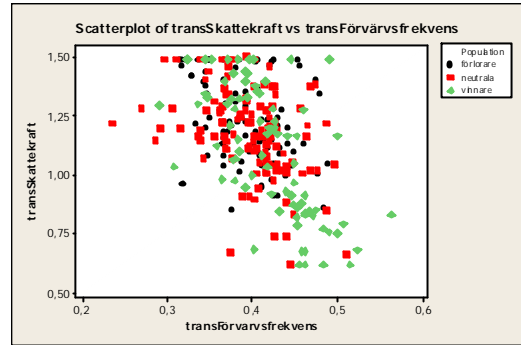
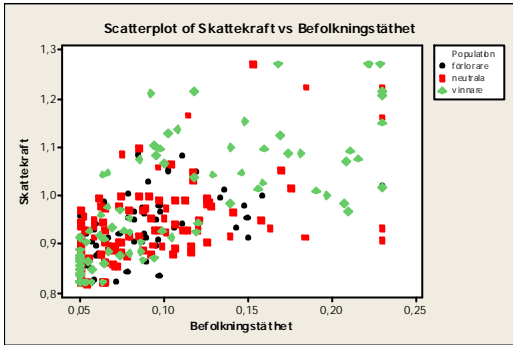
Appendix 3: År 2001



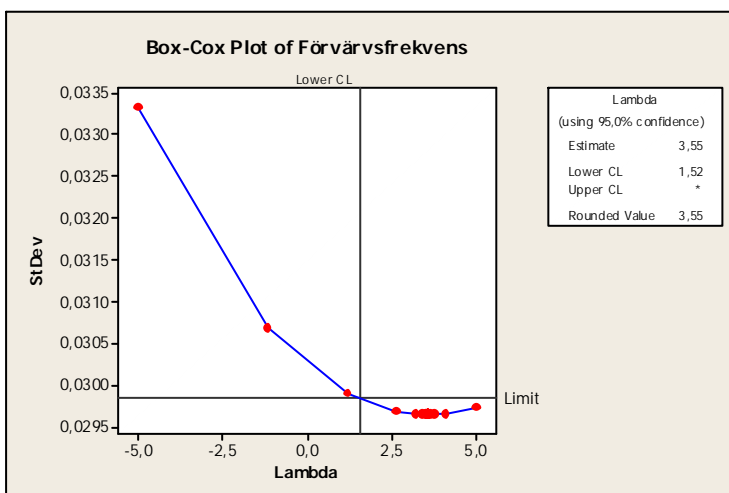
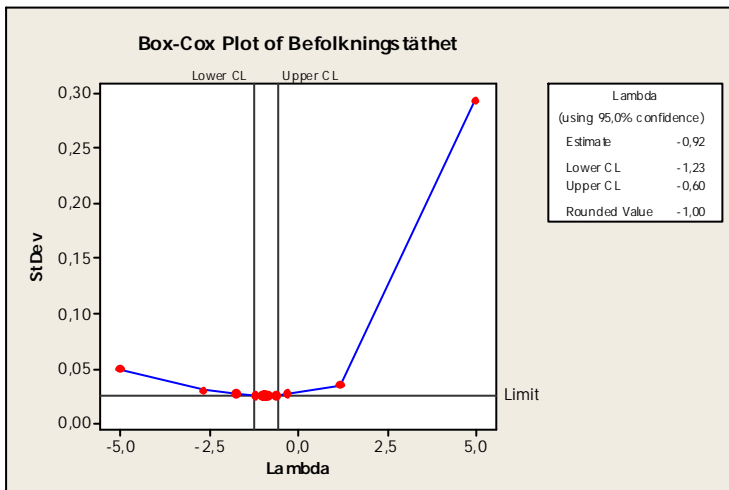
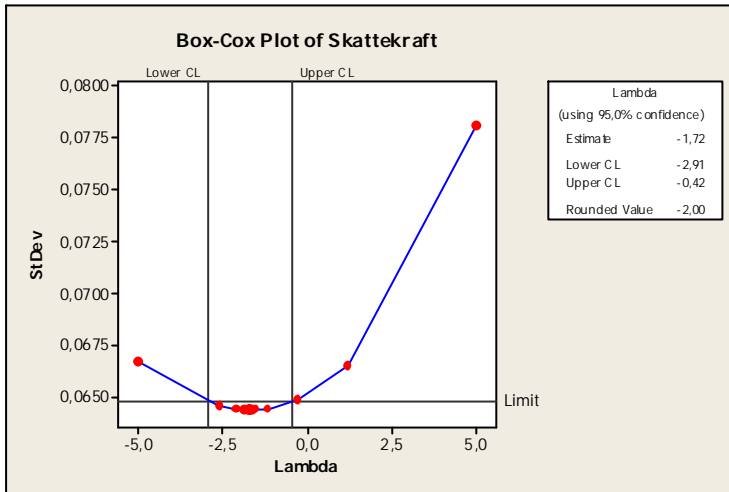


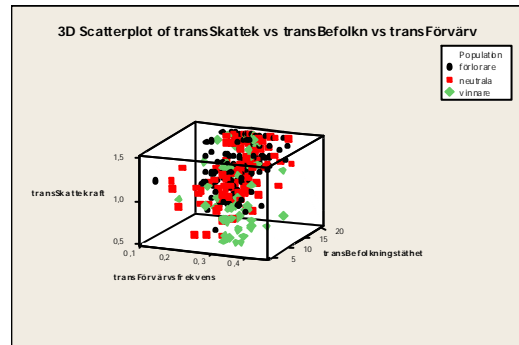
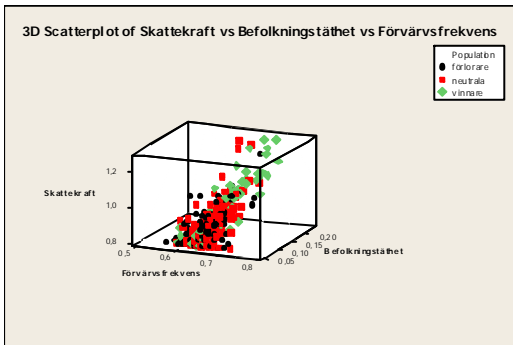
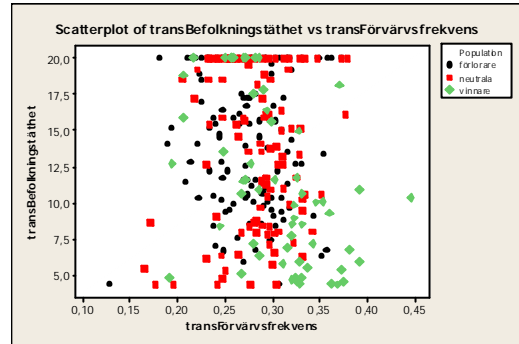
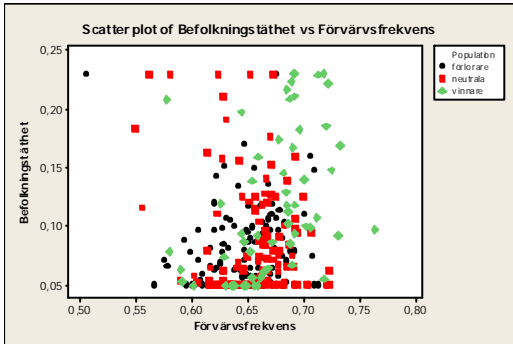
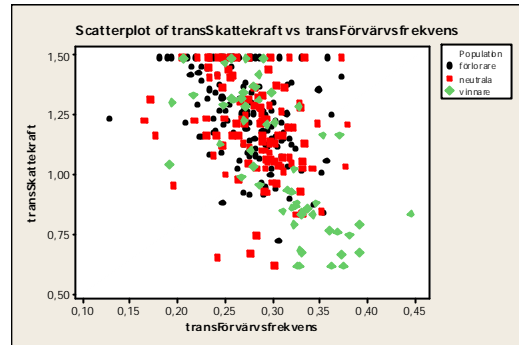
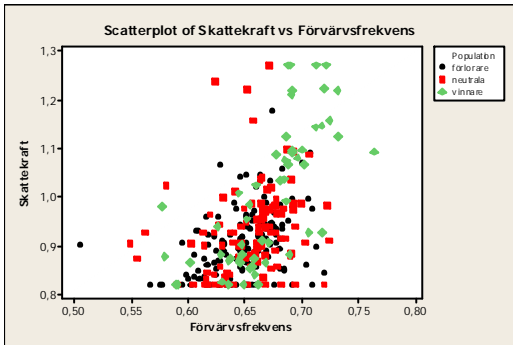
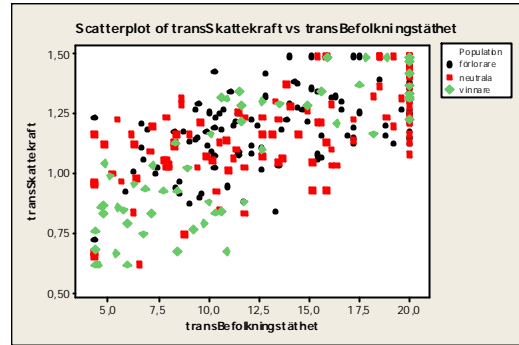
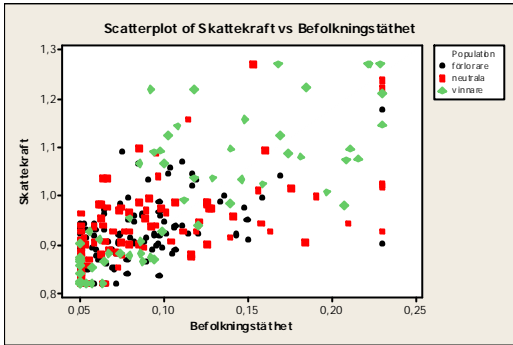
Appendix 4: År 2002





Appendix 5: År 2003





Appendix 6

```
TITLE 'DA med X1 X2 X3 lika apiori år 2000';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.3333 '2'=0.3333 '3'=0.3333;
CLASS pop;
VAR X1 X2 X3;
RUN;
TITLE 'DA med transX1 transX2 transX3 lika apriori år 2000';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.3333 '2'=0.3333 '3'=0.3333;
CLASS pop;
VAR transX1 transX2 transX3;
RUN;
TITLE 'DA med X1 X2 X3 viktade apiori år 2000';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.246528 '2'=0.333333 '3'=0.420139;
CLASS pop;
VAR X1 X2 X3;
RUN;
TITLE 'DA med transX1 transX2 transX3 viktade apriori år 2000';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.246528 '2'=0.333333 '3'=0.420139;
CLASS pop;
VAR transX1 transX2 transX3;
RUN;

TITLE 'DA med X1, X2 och X3 lika apriori år 2001';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.3333 '2'=0.3333 '3'=0.3333;
CLASS pop;
VAR X1 X2 X3;
Run;
TITLE 'DA med transX1, transX2 och transX3 lika apriori år 2001';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.3333 '2'=0.3333 '3'=0.3333;
CLASS pop;
VAR transX1 transX2 transX3;
Run; ;;;
TITLE 'DA med X1, X2 och X3 viktad apriori år 2001';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.315972 '2'=0.319444 '3'=0.364583;
CLASS pop;
VAR X1 X2 X3;
Run;
TITLE 'DA med transX1, transX2 och transX3 viktad apriori år 2001';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.315972 '2'=0.319444 '3'=0.364583;
CLASS pop;
VAR transX1 transX2 transX3;
Run;
```

```

TITLE 'DA med X1 X2 X3 lika apiori år 2002';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.3333 '2'=0.3333 '3'=0.3333;
CLASS pop;
VAR X1 X2 X3;
RUN;
TITLE 'DA med transX1 transX2 transX3 lika apriori år 2002';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.3333 '2'=0.3333 '3'=0.3333;
CLASS pop;
VAR transX1 transX2 transX3;
RUN;
TITLE 'DA med X1 X2 X3 viktade apiori år 2002';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.295139 '2'=0.406250 '3'=0.298611;
CLASS pop;
VAR X1 X2 X3;
RUN;
TITLE 'DA med transX1 transX2 transX3 viktade apriori år 2002';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.295139 '2'=0.406250 '3'=0.298611;
CLASS pop;
VAR transX1 transX2 transX3;
RUN;

TITLE 'DA med X1 X2 X3 lika apiori år 2003';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.3333 '2'=0.3333 '3'=0.3333;
CLASS pop;
VAR X1 X2 X3;
RUN;
TITLE 'DA med transX1 transX2 transX3 lika apriori år 2003';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.3333 '2'=0.3333 '3'=0.3333;
CLASS pop;
VAR transX1 transX2 transX3;
RUN;
TITLE 'DA med X1 X2 X3 viktade apiori år 2003';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.201389 '2'=0.364583 '3'=0.434028;
CLASS pop;
VAR X1 X2 X3;
RUN;
TITLE 'DA med transX1 transX2 transX3 viktade apriori år 2003';
PROC DISCRIM ALL pool=test crossvalidate;
priors '1'=0.201389 '2'=0.364583 '3'=0.434028;
CLASS pop;
VAR transX1 transX2 transX3;
RUN;

```