



Matematisk statistik  
Stockholms universitet

**Avvikande basområden –  
ett problem vid befolkningsprognos**

Karin Fägerlind

**Examensarbete 2006:20**

**Postal address:**

Matematisk statistik  
Dept. of Mathematics  
Stockholms universitet  
SE-106 91 Stockholm  
Sweden

**Internet:**

<http://www.math.su.se/matstat>



# Avvikande basområden – ett problem vid befolkningsprognos

Karin Fägerlind\*

November 2006

## Sammanfattning

Stockholms län består av 26 komuner, som i sin tur är uppdelade i nästan 1500 så kallade basområden.

På uppdrag av RTK, Regionplane- och trafikkontoret gör företaget Statisticon befolkningsprognoser för hela länet, kommunerna och basområdena.

Basområdenas befolkning kan vara svår att prognostisera på grund av de små befolkningarna, men det finns även andra faktorer som kan påverka prognosens säkerhet. Förekomsten av studenthem eller ålderdomshem är exempel på sådana faktorer, eftersom studenter och boende på olika hem har andra egenskaper än icke-studerande eller gamla som fortfarande bor hemma i sin egen lägenhet. När dessutom de boende på något av nämnda hem flyttar eller dör ersätts de inte av en slumpmässig hyresgäst utan just en ny student eller åldring, med samma typ av egenskaper.

Om de allmänna antagandena för in- och utflyttning som gäller för hela kommunen tillämpas på områden med dessa områden, skulle den åldrande befolkningen på ålderdomshemmen antas så småningom ersättas med inflyttare i alla åldrar (allt eftersom de gamla väntas dö) och studenterna skulle antas få barn och åldras, i stället för att, som i verkligheten, flytta ut efter ett par års studier.

För att kunna prognostisera dessa avvikande basområden på ett mer adekvat sätt måste de hittas, och en metod för detta diskuteras i denna rapport, nämligen diskriminantanalys.

Målet är att finna en funktion som ger varje basområde ett värde. Överstiger värdet en bestämd gräns antas området vara 'avvikande',

---

\*Postal address: Matematisk statistik, Stockholms universitet, SE-106 91, Stockholm, Sweden. E-mail: [to\\_karin@hotmail.com](mailto:to_karin@hotmail.com). Handedare: Anders Björkström

det vill säga det antas innehålla ett ålderdoms- eller studenthem, om inte antas det vara 'normalt'. För att kalibrera en sådan funktion lokaliserades först alla kända ålderdoms- och studenthem i ett begränsat geografiskt område, tyvärr under förutsättningar där misstag var oundvikliga.

Rapporten diskuterar vilka variabler som bör användas och vilka gränser som bör sättas när det kommer till faktorer som befolkningsstorlek och storleken på den 'avvikande' befolkningen, och även resultatet i sig. Slutsatsen är ett påvert resultat i gjorda undersökningar eventuellt kan härledas till den betydande felmarginalen i kalibreringsdatan och att metoden då skulle kunna vara användbar för att finna basområden med ålderdomshem om man mer exakt kunde fastställa ålderdomshemmens läge och storlek, medan metoden inte skulle kunna användas för att hitta studentområden, eftersom de är så pass få och studenternas egenskaper inte skiljer sig så mycket från en normal befolkning.

# Innehåll

<b>1</b>	<b>Sammanfattning</b>	<b>3</b>
<b>2</b>	<b>Abstract in English</b>	<b>4</b>
<b>3</b>	<b>Bakgrund till uppgiften</b>	<b>5</b>
3.1	Kortfattad introduktion till befolkningsprognos . . . . .	5
3.1.1	Nyttan av en bra befolkningsprognos . . . . .	5
3.1.2	Hur man går till väga . . . . .	5
3.1.3	Att beräkna inflyttade . . . . .	6
3.1.4	Befolkningen förändras . . . . .	6
3.1.5	Exempel . . . . .	7
3.2	Basområden med icketytiska egenskaper . . . . .	7
3.2.1	Vad är ett basområde? . . . . .	7
3.2.2	Problemet med icketytiska basområden . . . . .	8
3.3	Beskrivning av uppgiften . . . . .	8
3.3.1	Att finna en metod . . . . .	8
<b>4</b>	<b>Studenter och åldringars egenskaper</b>	<b>9</b>
4.1	Basområdena undersöks . . . . .	9
4.1.1	Basområdenas storlek . . . . .	9
4.1.2	Områden med många i en åldersgrupp . . . . .	9
4.1.3	S-områden och Å-områden . . . . .	10
4.2	Befolkningens egenskaper . . . . .	11
4.2.1	Jämförande undersökning . . . . .	11
4.3	Studenter jämfört med andra unga . . . . .	11
4.3.1	Studenternas flyttbenägenhet . . . . .	13
4.3.2	Till och från Sverige eller utlandet . . . . .	13
4.3.3	Studenternas ålder vid flytt . . . . .	14
4.3.4	Födelsefrekvens . . . . .	14
4.4	Ålderdomshem jämfört med andra gamla . . . . .	15
4.4.1	Ålderdomshemmens könsfördelning . . . . .	15
4.4.2	Ålderdomshemmens dödlighet . . . . .	16
4.4.3	Ålderdomshemmens flyttbenägenhet . . . . .	16
4.4.4	Åldringarnas ålder vid flytt . . . . .	16
4.5	Slutsatser efter första undersökningen . . . . .	17
4.6	Förberedelser för analys . . . . .	17
4.6.1	Kriterium för att ett område ska analyseras . . . . .	17

4.6.2	Kontroll av resultatet . . . . .	17
4.6.3	De tidigare undersökningarnas resultat . . . . .	18
<b>5</b>	<b>Diskriminantanalys</b>	<b>19</b>
5.1	En grafisk modell . . . . .	19
5.1.1	En algebraisk modell . . . . .	19
5.2	Klassificering med flerdimensionell normalfördelning . . . . .	21
5.2.1	Samma kovariansmatris . . . . .	22
5.2.2	Skatta väntevärde och kovarians . . . . .	22
5.3	Tillämpbarhet på det aktuella problemet . . . . .	23
5.3.1	Klassificeringsproblematiken . . . . .	23
5.4	Val av variabler . . . . .	23
5.4.1	Variablernas fördelning . . . . .	24
5.4.2	Variablernas fördelning för de större områdena . . . . .	27
<b>6</b>	<b>Diskriminantanalys på Studentområden</b>	<b>31</b>
6.1	Få studentområden . . . . .	31
6.2	En första analys . . . . .	31
6.2.1	Overfitting . . . . .	32
6.2.2	Val av kriterium . . . . .	33
6.2.3	Val av befolkningsstorlek . . . . .	34
6.3	Kostnad för felklassificering . . . . .	34
6.4	Alternativa sätt att välja variabler . . . . .	35
6.5	Slutsats . . . . .	37
<b>7</b>	<b>Diskriminantanalys på Äldreområden</b>	<b>38</b>
7.1	Gränser för analysen . . . . .	38
7.1.1	Gräns för andelen ålderdomshemboende . . . . .	38
7.1.2	Gräns för befolkningens storlek . . . . .	39
7.2	Välja variabler efter diskriminantfunktionen . . . . .	39
7.3	Slutsats . . . . .	40
7.4	Ålderdomshemmen i framtiden . . . . .	40
<b>8</b>	<b>Slutsatser</b>	<b>41</b>
<b>9</b>	<b>Appendix</b>	<b>43</b>
9.1	Ordlista . . . . .	43
9.2	Litteraturlista . . . . .	43
9.3	Tack . . . . .	43

# Kapitel 1

## Sammanfattning

Stockholms län består av 26 kommuner, som i sin tur är uppdelade i nästan 1500 så kallade basområden.

På uppdrag av RTK, Regionplane- och trafikkontoret gör företaget Statisticon befolkningsprognoser för hela länet, kommunerna och basområdena.

Basområdenas befolkning kan vara svår att prognostisera på grund av de små befolkningarna, men det finns även andra faktorer som kan påverka prognosens säkerhet. Förekomsten av studenthem eller ålderdomshem är exempel på sådana faktorer, eftersom studenter och boende på olika hem har andra egenskaper än icke-studerande eller gamla som fortfarande bor hemma i sin egen lägenhet. När dessutom de boende på något av nämnda hem flyttar eller dör ersätts de inte av en slumpmässig hyresgäst utan just en ny student eller åldring, med samma typ av egenskaper.

Om de allmänna antagandena för in- och utflyttning som gäller för hela kommunen tillämpas på områden med dessa områden, skulle den åldrande befolkningen på ålderdomshemmen antas så småningom ersättas med inflyttare i alla åldrar (allt eftersom de gamla väntas dö) och studenterna skulle antas få barn och åldras, i stället för att, som i verkligheten, flytta ut efter ett par års studier.

För att kunna prognostisera dessa avvikande basområden på ett mer adekvat sätt måste de hittas, och en metod för detta diskuteras i denna rapport, nämligen diskriminantanalys.

Målet är att finna en funktion som ger varje basområde ett värde. Överstiger värdet en bestämd gräns antas området vara 'avvikande', det vill säga det antas innehålla ett ålderdoms- eller studenthem, om inte antas det vara 'normalt'. För att kalibrera en sådan funktion lokaliserades först alla kända ålderdoms- och studenthem i ett begränsat geografiskt område, tyvärr under förutsättningar där misstag var oundvikliga.

Rapporten diskuterar vilka variabler som bör användas och vilka gränser som bör sättas när det kommer till faktorer som befolkningsstorlek och storleken på den 'avvikande' befolkningen, och även resultatet i sig. Slutsatsen är ett påvert resultat i gjorda undersökningar eventuellt kan härledas till den betydande felmarginalen i kalibreringsdatan och att metoden då skulle kunna vara användbar för att finna basområden med ålderdomshem om man mer exakt kunde fastställa ålderdomshemmens läge och storlek, medan metoden inte skulle kunna användas för att hitta studentområden, eftersom de är så pass få och studenternas egenskaper inte skiljer sig så mycket från en normal befolkning.

## Kapitel 2

# Abstract in English

The district of Stockholm includes 26 municipalities which are in their turn divided into almost 1500 so called base areas. The company Statisticon is making population forecasts for RTK, Regionplane- och trafikkontoret, a bureau that is responsible among other things for the planning of infrastructure, habitation, issues of environment and water supply.

The population of the base areas vary, about 10 percent of them are uninhabited, normally they have a few hundred inhabitants. Small populations make forecasting insecure, but there are also other factors e.g. the characteristics of the inhabitants, that obstruct the reliability of the forecast.

Especially base areas with students homes or homes for the aged are interesting in this matter. Students and elderly living in institutions have other qualities than nonstudents and elderly living in normal housing, and above all, when moving out or dying, they are replaced not with random people, but a new student or old person with the same qualities.

If normal expectations were used in predicting the future of these base areas, the population replacing the elderly (who are expected soon to die) would look like that in any other area, and the students in their student flats would raise children and grow old, instead of moving out within the next few years as they actually have to do when graduating or dropping out.

To treat these 'special' areas, it is needed to find them and one method to do so is discussed in this report, more exactly; through discriminant analysis.

The aim is to find a function, which gives each area a value. If the value is above a certain limit, the area is considered containing a home for the aged, or a students' home, when not it is considered 'normal'. To create such a function, a test sample was produced, unfortunately under circumstances where mistakes where unavoidable.

The report discusses which variables to use, and which limits to demand when it comes to the size of the population and the size of the diverging population. Eventually it discusses the worth of the result, since the margin of error was so wide, and establishes that the method could be useful for finding homes for the aged if the test sample could be made more exact, but that such actions would probably not be enough for making the finding of the student homes successful, since the students homes are too few, and their inhabitants not enough diverged from 'common people'. Instead a manual directory is suggested for the student homes in the case of the district of Stockholm.



# Kapitel 3

## Bakgrund till uppgiften

### 3.1 Kortfattad introduktion till befolkningsprognos

#### 3.1.1 Nyttan av en bra befolkningsprognos

Det som skapar all politisk, ekonomisk, social och kulturell dynamik i ett samhälle är de individer som lever i det. Därför kan nyttan av en god prediktion av människors förehavanden knappast överdrivas. Nya vägar, sjukhus, skolor och placering och omflyttning av statliga verk och myndigheter kräver kunskap om *hur många människor som kommer att bo i det aktuella området i framtiden*. Detsamma gäller beslut om skatter och bidrag, utveckling av lokaltrafiken och stiftandet av många lagar. Ett företag som vill öppna en ny filial frågar sig också *hur många potentiella anställda och hur många potentiella kunder det finns i närområdet i framtiden*, oavsett om det gäller en liten butik, ett köpcenter eller en pappersmassfabrik.

#### 3.1.2 Hur man går till väga

Antag att man kontrollerar befolkningen sista december 2005, och vill skatta befolkningen sista december 2006. Man utgår ifrån den befintliga befolkningen i området som ska beräknas. Befolkningen delas in i grupper om kön och ålder. Därefter används historiska data för att skatta ett antal risker, nämligen risken att dö, risken att flytta och (för kvinnor) risken att generera en ny individ, det vill säga att föda barn<sup>1</sup>. Mycket förenklat skattas risken för en individ att exempelvis dö under en period, genom att man ser hur många individer av samma kön och ålder som dog förra perioden, och delar detta antal med det totala antalet individer i denna kön-åldersgrupp, det vill säga

$$\text{risken att dö} = \frac{\text{antal döda}}{\text{totala antalet individer}}$$

Generellt sett blir en skattning säkrare ju större data den baseras på. Därför kan det vara bättre att skatta riskerna med en linjärkombination av flera års

---

<sup>1</sup>Av beräkningsmässiga skäl negligeras mannens del i processen

data, det vill säga, för en kön-åldersgrupp

$$\text{risken att dö} = \sum_{i=1}^n \lambda_i \frac{\text{antal döda period } i}{\text{totalt antal individer period } i}$$

$i$ =index som anger år  $i = 1 \dots n$   $\sum_{i=1}^n \lambda_i = 1$

På detta sätt beräknas risken att dö, risken att flytta ut och risken att få barn för alla grupper om kön och ålder. Därefter beräknas de inflyttade.

### 3.1.3 Att beräkna inflyttade

Naturligtvis är det så att alla individer i hela världen, oavsett var de bor, har en risk att flytta till Stockholms län under kommande år. Men beräkningar på dessa risker och sex miljarder människor låter sig inte göras. Inflyttningen till ett område kan istället i många fall ses som en poissonprocess med en inflyttningsintensitet. Avseende Stockholms län vore det felaktigt att tala om en poissonprocess för inflyttarna eftersom det föreligger en betydande bostadsbrist (för en poissonprocess förutsätts ett 'oändligt väntrum'), i stället antas inflyttningen bero av bostadsbyggandet.

Man skattar boendetätheten (*antal boende/bostad*) för de befintliga bostäderna i området och antar den vara konstant för kommande period. Man skattar hur många bostäder som kommer att rivas och räknar av dem från de befintliga bostäderna.

Därefter skattas antalet bostäder under uppbyggnad och antalet fritidshus som kommer permanentas, och boendetätheten i dessa. Fram tills för några år sedan antogs boendetätheten i de nya bostäderna vara densamma som för de befintliga bostäderna, men RTK har nu utvecklat en modell för att skatta boendetätheten med avseende på belägenhet, storlek, ägarkategori (om det rör sig om äganderätt, bostadsrätt eller hyresrätt), samt fastighetstyp (flerbostadshus, småhus och övriga). För att skatta befolkningsstorleken nästa period räknar man ut den befintliga boendetätheten\*antalet befintliga bostäder+den nya boendetätheten\*antalet nya bostäder.

Det skattade antalet inflyttade blir befolkningsstorleken minus antal döda och avflyttade plus antalet födda. Kön- och åldersfördelningen för de inflyttade skattas till exempel med (eventuellt justerade) tidigare data och man slumpar ut de inflyttade efter dessa fördelningar.

### 3.1.4 Befolkningen förändras

Först beräknas befolkningen för hela länet och därefter för kommuner och basområden, och man måste alltså kontrollera att summeringarna stämmer, det vill säga; lägger man ihop kommunernas beräknade döda, födda, in- och utflyttade måste de summeras till de antal som beräknats för hela länet. En annan viktig aspekt är att en population människor inte beter sig som bakterier i en odling. Att vi lever längre och längre är ett tydligt fenomen, och sådana trender måste nog beaktas. Än svårare är det att beräkna risken att flytta och föda barn, här kan man se starka trender som dessutom växlar snabbt och ibland oväntat. Även om det bara är nästa års befolkning som ska predikteras kan felet bli stort om man endast skattar risken med förra årets observationer. För en längre period är det naturligtvis ännu viktigare att göra en bra skattning av trenderna. Hur detta i praktiken går till, avhandlas inte i denna rapport.

### 3.1.5 Exempel

Antag att vi vill räkna ut antal 31-åriga kvinnor i Stockholms län den 31:a december 2006. Antag att det fanns 15 813 30-åriga kvinnor i länet den 31:a december 2005 (de kommer ju att fylla 31 under året). Antag vidare att risken att dö under det närmsta året är skattat till 0.00045, risken att få barn under året skattad till 0.13 och risken att flytta ut ur länet under året skattad till 0.05 för kvinnor i denna ålder. Med dessa siffror får vi att

$$\text{Antalet dödsfall} = 15813 * 0.00045 \approx 7$$

$$\text{Antalet utflyttade} = 15813 * 0.05 \approx 791$$

$$\text{Antalet nyfödda} = 15813 * 0.13 \approx 2056$$

De 30-åriga kvinnorna förväntas alltså under året ha utvecklats till  $15813 - 7 - 791 = 15015$  31-åriga kvinnor och 2056 spädbarn. Man beräknar så hur alla de olika kön-åldersgrupperna kommer att utvecklas under året. Därefter beräknar man de inflyttande i varje kön-åldersgrupp, för de 30-åriga kvinnorna närmare bestämt så: Antag att det under året kommer göras plats åt ytterligare 10 000 människor i länet. Antag dessutom att beräkningar enligt ovan givit att totalt 137 000 personer flyttar eller dör och 25 000 barn föds i länet. Antalet inflyttade ges av *antalet nya bostäder \* boendetätheten + antalet utflyttade och döda - antalet födda* det vill säga; i detta fall  $10000 + 137000 - 250000 = 122000$  inflyttare. Dessa inflyttare låter man fördela sig så nära den skattade fördelningskurvan för inflyttare som möjligt. Säg att man skattat att 6,7 promille av de inflyttade är kvinnor som var 30 år vid förra årets slut, då blir det skattade antalet inflyttade kvinnor i denna åldersgrupp 817 stycken och vi har vid årets slut  $15015 + 817 = 15832$  31-åriga kvinnor. Denna procedur genomför man för varje grupp med individer i samma ålder av samma kön, och får därigenom den prognostiserade befolkningen nästa år (*ursprungsbefolkning - döda + nyfödda - utflyttade + inflyttade*). För att göra en tvåårsprognos utgår man från den predikterade prognosen och använder de risker som man skattat för nästnåsta år och så vidare.

## 3.2 Basområden med icketytiska egenskaper

### 3.2.1 Vad är ett basområde?

Stockholms län består av 26 kommuner och kommunerna delas i sin tur in i mindre geografiska områden som kallas basområden. Att kalla de mindre områdena basområden är inte praxis i hela Sverige utan gäller just i Stockholms län. Basområdena är olika många i olika kommuner, är olika stora och har olika antal invånare. Företaget Statisticon gör bland annat befolkningsprognoser för Stockholms län på uppdrag av RTK (Regionplane- och trafikkontoret). Befolkningsprognosen gäller även kommunerna och basområdena. Man går tillväga som så att man skriver fram befolkningen och sedan räknar man av det förväntade antalet döda och utflyttade, och låter förväntat antal kvinnor föda barn, så att summorna för antalet döda, ut- och inflyttade och födda stämmer för kommunen och hela länet.<sup>2</sup> Detta tillvägagångssätt skulle inte vara något problem

<sup>2</sup>Befolkningsprognos 2004

om invånarstrukturen i varje basområde någorlunda liknade invånarstrukturen i hela kommunen, det vill säga om det bodde människor 'i alla åldrar' i varje basområde. Men så är *inte* fallet. Det finns basområden där det uteslutande bor studenter eller åldringar, helt enkelt för att de enda bostäder som finns i området är studentbostäder eller ålderdomshem. Det finns också områden som består av både studentbostäder och ålderdomshem, det finns basområden som innehåller vanliga bostäder men också en stor del studenthem eller ålderdomshem. Kanske finns det också andra typer av boenden som avviker, exempelvis seniorboenden och andra typer av ungdomsboenden.

Även i dessa små områden är en bra prognos eftersträvansvärd eftersom den framtida befolkningen har betydelse för planeringen av kommunal och privat service och infrastruktur.

### 3.2.2 Problemet med icketyypiska basområden

Problemet med dessa avvikande basområden är att om man gör en prognos enligt beskrivningen i exempel 3.1.5 och med liknande antaganden på ett avvikande basområde så får vi ett resultat som är riktigt dåligt. Unga kvinnliga studenter föder inte alls barn i samma utsträckning som arbetande kvinnor i samma ålder. In- och utflyttningen i studentbostäder är högre, och framför allt; de som flyttar in i dessa områden är *studenter respektive åldringar*. De som flyttar in i hela kommunen är med all sannolikhet en blandning av barn, ungdomar, vuxna, medelålders, pensionärer och åldringar. Om vi använder åldersfördelningen för hela kommunens inflyttare för dessa områden kommer vi alltså anta att även en hel del barn och gamla flyttar in i studentområden och att områden med ålderdomshem kommer att ta emot en del unga vuxna. Faktum är att efter ett visst antal iterationer (det vill säga; i det här fallet år) kommer dessa områden befolkningsmässigt inte längre skilja sig från resten av kommunen. Och det är naturligtvis ett helt felaktigt antagande.

## 3.3 Beskrivning av uppgiften

### 3.3.1 Att finna en metod

Uppgiften som jag fått av Statisticon är först och främst att definiera dessa "icketyypiska" basområden, och finna en metod att snabbt söka rätt på dem. Med nästan 1500 basområden i Stockholms län är detta något som måste kunna utföras av en dator. Jag måste alltså undersöka hur dessa basområden befolkningsmässigt ser ut och hur de mer i detalj skiljer sig från de konventionella basområdena. Skillnader i faktorer som medelålder, andel kvinnor, flyttfrekvens, födslofrekvens, andel människor som är äldre eller yngre än en viss ålder, hela åldersfördelningskurvans utseedende eller kanske arean under den är exempel på sådana detaljer. I ett område med både vanliga bostadshus och ett studenthem kommer felet som uppstår när man använder de allmänna antagandena kanske inte att bli så stort. Hur stort felet får vara är inte angivet, utan det ingår i uppgiften att göra en avvägning.

## Kapitel 4

# Studenter och åldringars egenskaper

### 4.1 Basområdena undersöks

#### 4.1.1 Basområdenas storlek

Materialet för uppgiften är de fem senaste årens befolkningsdata i Stockholmsområdet, närmare bestämt ett dataset med följande variabler: Basområde, kommun, ålder, kön, födda, döda, inflyttade och utflyttade. Inflyttade är uppdelade på om de flyttar in från övriga Sverige eller från ett annat land, samma sak gäller de som flyttar ut. Det exakta antalet i varje kön-åldersgrupp är angivet den 31 december varje år, men som mått på befolkningsstorleken föredras medelbefolkningen (befolkningen vid ingång och utgång av varje period delat med 2 det vill säga, det genomsnittliga antalet personer under året). På grund av att vi ej har tillgång till data för ingången av år 2000 kan medelbefolkningen detta år inte beräknas, men data från detta år används för uträkning av medelbefolkningen år 2001. Den undersökta perioden är alltså fyra år, 2001-2004.

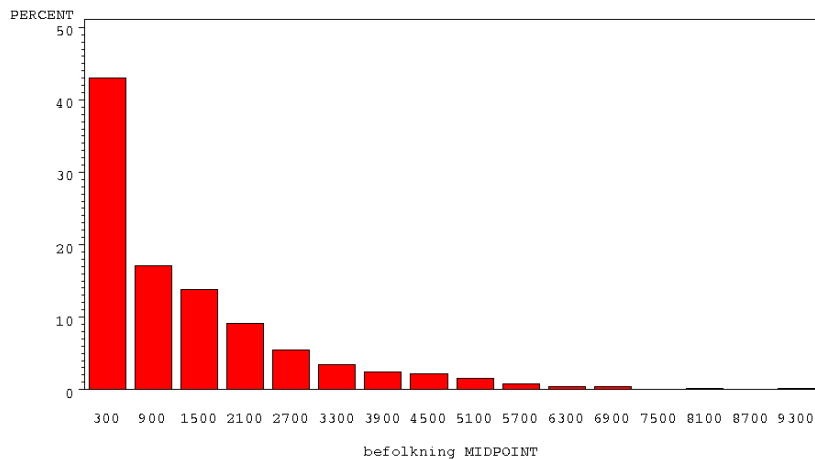
Basområdenas storlek varierar stort, vilket histogram 4.1 visar. 10 % av områdena har inte ens en hel invånare per år. 25 % har färre än 100 invånare. Medianen går vid 829 personer och 90% av basområdena har färre än 3258 invånare per år.

#### 4.1.2 Områden med många i en åldersgrupp

För att kunna skatta studenter respektive åldringars egenskaper används de mest homogena områdena, där det nästan bara bor människor som är i samma ålder. Befolkningen delas in i fem klasser efter ålder.

<b>A</b>	0-18 år	(23,9 %)
<b>B</b>	19-34 år	(21,2 %)
<b>C</b>	35-49 år	(22,1 %)
<b>D</b>	50-64 år	(18,4 %)
<b>E</b>	65-100 år	(14,4 %)

Observera att de som är äldre än 100 år registreras som 100-åringar i datan.



Figur 4.1: Histogram över basområdenas befolkningsstorlek

Inom parentes står den procentuella andelen av den totala befolkningen i varje åldersgrupp. Uppdelningen förklaras genom att det i första hand är tänkt att sålla ut studenter och åldringar. Många börjar på universitetet direkt efter sina gymnasiestudier, det vill säga; det är de fyller 19, och majoriteten av studenterna har slutat studera vid senast 34 års ålder. Antagligen kan denna övre gräns skjutas nedåt, men det är i detta tidiga skede värt att vara generös för att inte riskera att missa något. Gränsen vid 50 år sattes för att se om något seniorboende skulle falla ut, där gränsen för inflyttning ofta är 50 år.

Den som inte kan ta hand om sig själv hänvisas till handikappomsorgen upp till 65 års ålder och därefter äldreomsorgen, därför har den översta gränsen satts vid 65 år. Antagligen är majoriteten av dem som flyttar till ett ålderdomshem mycket äldre än så, men med samma resonemang som i fallet studenter har gränsen satts generöst. De basområden som har mer än 80% av sina invånare i en och samma klass anses vara ett homogent område i denna undersökning. Naturligtvis sorteras då fram en hel del områden med ett par enstaka invånare, som råkar vara i samma ålder. Om det är slump eller inte är omöjligt att säga bara genom att titta på data, och helt ointressant. För att undvika dem kontrolleras endast områden med fler än 5 invånare i snitt per år. Intressant nog är det bara basområden med mer än 80 % av invånarna i klass B och E som faller ut.

### 4.1.3 S-områden och Å-områden

För enkelhetens skull kallar vi områden som innehåller studenthem för S-områden och områden som innehåller ålderdomshem för Å-områden. Områden som nu sållats ut, där invånarna åldersmässigt är nästan helt homogena kallas 'rena' S-områden och Å-områden. Det finns två rena Å-områden och de har tillsammans i genomsnitt 430 invånare per år. Det finns fem rena S-områden och de har i genomsnitt 645 ivånare per år.

## 4.2 Befolkningens egenskaper

De egenskaper som är typiska för S-områden och Å-områden undersöks. Framför allt jämförs och graderas kvoter av olika slag, vanligtvis med totalbefolkningen som nämnare, för alla basområden. I de fall då alla rena S- respektive Å-områden hamnar under 10-percentilen eller över 90-percentilen anses denna kvot vara en typisk egenskap, och andra egenskaper i anslutning till täljaren undersöks vidare. Små basområden får ofta extremvärden på grund av sin ringa totala befolkning, därför undersöks endast områden med minst 5 invånare i snitt per år (20 invånare ackumulerat). Oberoende antas mellan boende på hem och övriga befolkningen. Det vill säga; det antas att de som bor i närheten av ett ålderdomshem eller studenthem inte skiljer sig från övriga befolkningen. Därför är det meningslöst att undersöka områdenas hela befolkning, utan undersökningen begränsas till respektive åldersgrupp.

### 4.2.1 Jämförande undersökning

De flesta egenskaper hos befolkningen är starkt bundna till ålder. De allra flesta skaffar barn mellan 20 och 40 och dör i alla fall efter pensionen. Många flyttlass går också vid bestämda händelser i livet: man flyttar hemifrån, efter ett tag flyttar många ihop med en partner och sedan flyttar familjen till större bostad när barnen föds. Egenskaper som bara beror på åldern är mindre intressanta, eftersom åldern är en av de variabler som redan ingår i grunddata och inte behöver sökas reda på. En större utmaning är att finna de egenskaper som *inte* beror på åldern, för att med dem kunna finna studentområden och ålderdomshem som inte är fullt så uppenbara som de i de rena områdena.

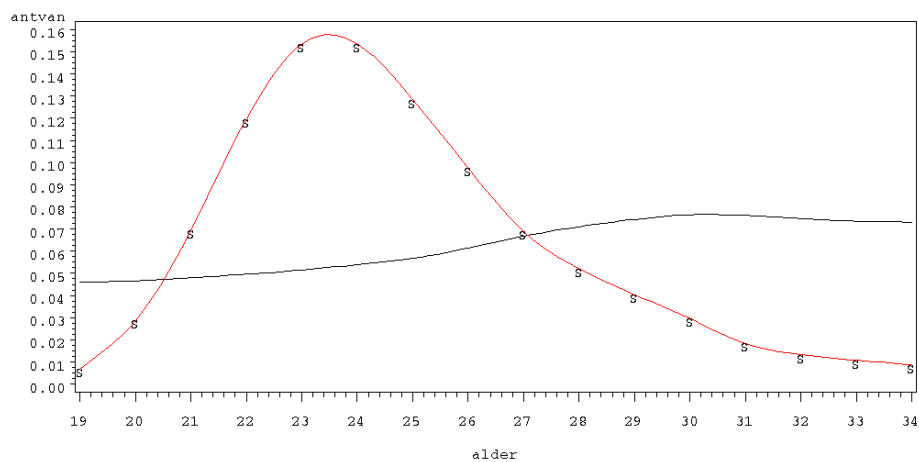
Sådana jämförelser låter sig emellertid inte göras så lätt. Vid jämförelser av åldersfördelningskurvorna för de rena S- och Å-områdena och motsvarande åldersklasser (B respektive E) i de övriga områdena, ser man att de ser helt olika ut, figur 4.2 och 4.3 (studenternas kurva är märkt med 's', ålderdomshemmens kurva med '+'). Olikheter som framträder vid jämförelse av grupperna rakt av kan alltså mycket väl bero på att det är en skillnad i åldersstrukturen, det vill säga; dessa olikheter kan även de bero på åldern.

Detta problem skulle antingen kunna lösas genom att man jämför mindre åldersgrupper i de avvikande områdena med motsvarande grupper i de vanliga områdena, eller genom någon typ av regressionsanalys. Båda lösningarna faller emellertid på bristen på data, och undersökningarna har utförts på hela de aktuella åldersgrupperna, dock med detta dilemma 'i bakhuvudet'.

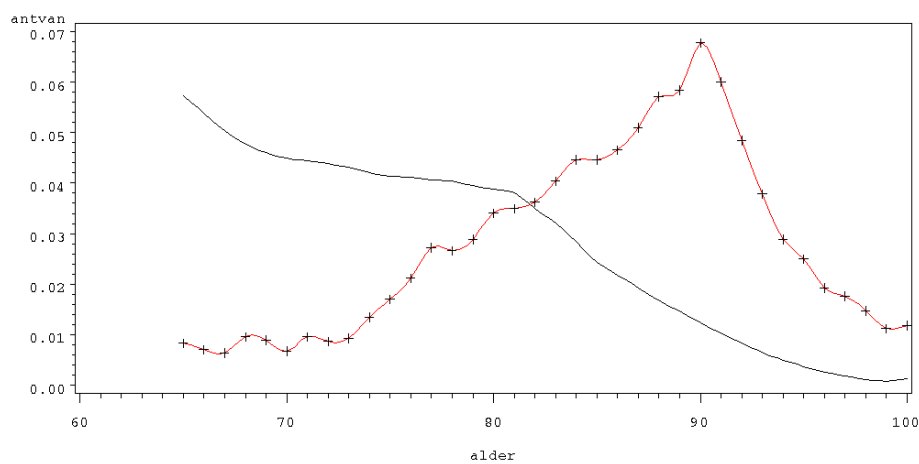
## 4.3 Studenter jämfört med andra unga

Först undersöks befolkningens spridning i de fem S-områdena, för att se om åldersgränserna för S-områden behöver justeras. 5-percentilen går vid 20 år och 95-percentilen vid 34 år, och detta föranleder ingen ändring av gränserna.

Om vi antar att studenters benägenhet att dö inte skiljer sig från icke-studerande ungdomar, är de faktorer som kan tänkas skilja grupperna inflyttning, utflyttning, och barnafödande. Det finns 1181 områden i länet med minst än 20 människor totalt i denna åldersgrupp. Dessa undersöks:



Figur 4.2: Studenter jämfört med andra unga



Figur 4.3: Boende på ålderdomshem jämfört med andra gamla



### 4.3.1 Studenternas flyttbenägenhet

Den som flyttar in i en studentbostad flyttar ofta vidare inom bara något år. Detta dels för att man bara får bo i bostaden så länge man faktiskt studerar, dels för att standarden ofta är låg och hyresgästen därför flyttar senast då hon eller han slutar studera och därmed får en bättre ekonomi. Detta resonemang leder till en misstanke om att in- och utflyttningen när det gäller studentbostäderna är mer frekvent än när det gäller vanliga bostäder.

Först undersöks hur befolkningen i studentområdena förändras under de fyra åren. Ett nybyggt bostadsområde får naturligt en hög inflyttningsfrekvens. Studentområdena har tillsammans 345,5 invånare i medelbefolkning 2001 och hela 865 invånare som medelbefolkning 2004, alltså mer än en fördubbling! Det är framför allt två områden som ökar, ett från 24 till 519,5 personer, det andra från en person till 26 personer. En hög inflyttning är alltså att vänta.

Som mått för att undersöka in- och utflyttningen används  $\text{inflyttningsandel} = \text{antal inflyttade} / \text{hela befolkningen}$  och  $\text{utflyttningsandel} = \text{antal utflyttade} / \text{hela befolkningen}$ . Observera att siffrorna för varje år helt enkelt lagts ihop, så att invånare som bor i området under hela perioden registreras fyra gånger, medan utflyttarna bara registreras det år de flyttar. Detta är alltså inget korrekt frekvensmått, utan bara ett mått för att enkelt kunna jämföra grupperna. När basområdena rankas efter inflyttningsandelen hos människor mellan 19 och 34 år placerar sig de rena S-områdena på första, elfte, 17:e, 23:a och 65:e plats. Det är gott och väl bland de 10% mest extrema värdena. De rena S-områdenas gemensamma inflyttningsandel är 72,6 % mot den totala inflyttningsandelen 29,0 %.

När områdena rankas efter utflyttningsandelen visas ett liknande mönster med ett undantag: ett område hamnar på 816:e plats, närmare bestämt det som ökade från en till 26 personer, det har bara 7 utflyttare på 30,5 ackumulerade invånare. Det visar sig att av dessa 30,5 invånare utgör 25,5 medelbefolkningen för de 51 personer som flyttade in i området under det sista året i datan. De har alltså inte hunnit bo där tillräckligt länge för att flytta.

Det område som ökade från 24 personer till 519,5, hade redan 2002 en medelbefolkning på 247,5 personer och därifrån har folk mycket väl hunnit flytta, utflyttningsandelen placerar sig på 22:a plats i rankningen! (De andra rena S-områdena placerade sig på tredje, 10:e och 25:e plats, och utflyttningsandelarna blev för S-områdena 50,3 % och totalt 26,4 %.)

På grund av att en hög inflyttningsandel kan bero på nybyggnation likaväl som på en flyttbenägen befolkning, och att en låg utflyttningsandel även det lika gärna kan bero på nybyggnation som på en mindre flyttbenägen befolkning undersöks i stället summan av dessa kvoter. Misstanken är ju att S-områdena har en hög inflyttnings- och utflyttningsandel. Och mycket riktigt: De rena S-områdena placerar sig på 2:a, 8:e, 11:e, 13:e och 33:e plats, bland de 3 % högsta värdena.

### 4.3.2 Till och från Sverige eller utlandet

Att utbilda sig på universitet och högskolor i Sverige är, till skillnad från många andra länder, gratis, och Stockholms län har flera välkända lärosäten. Följdaktligen kommer många från utlandet för att utbilda sig här, och åker tillbaka efter utbildningens slut, eller kanske efter bara en eller två terminer.

Ser man till statistiken håller dock inte resonemanget.

Kvoten *antal utrikes inflyttade/hela befolkningen* är visserligen högre än genomsnittet, men de fem basområdets värden har en stor spridning, det samma gäller *antal utrikes utflyttade/hela befolkningen*.

S-områdena har som sagt en stor andel inflyttare, och därför undersöks även proportionen utrikes inflyttare av det totala antalet inflyttade. Denna kvot är lägre för de rena S-områdena än för de vanliga områdena, 6,7 % jämfört med 7,1 % (även om det mest extrema området, där 87 % av de inflyttade kom från utlandet, är ett rent S-område). Även det omvända, andelen utflyttare som flyttar ut ur Sverige, är genomsnittet för de rena S-områdena lägre, 2,2 % jämfört med det totala genomsnittet 3,8 %. Skillnaden är dock inte så extrem att en låg andel in-och-utflyttningar till utlandet skulle kunna användas som en indikator för studentbostäder, i synnerhet inte som spridningen är så stor.

### 4.3.3 Studenternas ålder vid flytt

Ungdomar är generellt sett flyttbenägna, i synnerhet i bostadsbristens Stockholm. De flyttar visserligen hemifrån sent, men sedan väntar ofta flera år med kortfristiga boenden. Snittåldern för en inflyttare totalt i åldersgruppen 19-34 är 26,8 år, snittåldern för en utflyttare 27,0. I ett studentområde kan man förvänta sig en större skillnad mellan in- och ut-flyttarnas ålder. De som är i början av sina studier flyttar in, de som är i slutet av studierna flyttar ut, och det bör finnas en åldersskillnad mellan dessa. Detta resonemang bekräftas delvis med en undersökning. Den genomsnittliga inflyttaren i ett rent S-område är 24,0 år, den genomsnittliga utflyttaren är 25,3 år, det gör skillnaden dem emellan nästan åtta gånger så stor som den för hela länet. Men spridningen är lite för stor inom S-områdena, både när det gäller genomsnittliga in- och -utflyttningens åldrar och skillnaden dem emellan, för att det ska vara användbart i fortsatta undersökningar.

### 4.3.4 Födelsefrekvens

Antalet födda barn per fertil kvinna undersöks. 90 % av barnafödarna är mellan 23 och 39 år. Varje år får 9,0 % av kvinnorna i denna ålder barn, i de vanliga områdena vill säga. I de rena S-områdena får bara 1,3 % av de jämngamla kvinnorna barn. Denna jämförelse kan tyckas missvisande, eftersom de rena S-områdena till minst 80 % består av individer som är under 35 år, och alltså har långt färre tänkbara mödrar mellan 35 och 39 år än de vanliga områdena. Detta gör också att mödrarna i de rena S-områdena generellt sett är yngre än mödrarna i de vanliga områdena. När åldern för dem som föder barn istället jämförs inom denna åldersgrupp (19-34 år), är studenter som får barn fortfarande yngre än icke-studenter (median 26 år för studenter och 30 för icke-studenter), men detta beror antagligen på skillnader i åldersstrukturerna.

Antalet födda barn per år och kvinna i åldern 19-34 år i områden med fler än minst 2,5 kvinnor i snitt per år undersöks och områdena rankas efter denna kvot. Även här räknas kvinnorna en gång per år, medan födda barn bara räknas de år de föds. 1169 områden har minst 10 registrerade kvinnor i denna åldersgrupp totalt. Ett rent S-område har endast 7,5 kvinnor ackumulerat och är således inte med i undersökningen (Detta område hade inga födda). När områdena rankas

efter antalet barn per kvinna och år, hamnar de fyra undersökta studentområdena på plats nummer 525, 1124, 1130 och 1137. (Området med det högre antalet barn per kvinna sammanfaller inte med det område som avvek i undersökningen om utflyttningsfrekvens.) Här är det motiverat att definiera området på plats nummer 525 som en 'outlier'. De tre lägst placerade områdena hamnar inom 4-percentilen och det bortsållade området följer mönstret. Det avvikande området hade 4 födda barn på 43 kvinnor, och detta underlag är för lite för att inte anse låg födelsefrekvens vara en typisk egenskap för studenter. Totalt är födelseandelen 8,4 %, för studenter även i denna åldersgrupp 1,3 %.

## 4.4 Ålderdomshem jämfört med andra gamla

Rent formellt kan man få flytta till ett ålderdomshem redan vid 65 års ålder. Det är dock inte särskilt vanligt. Vid en undersökning av de två rena Å-områdena ges att 10-percentilen går vid just 65 års ålder, men det beror på att den dras ner av det ena området som har några yngre invånare. För det andra området går 10-percentilen vid 76 års ålder. Även inflyttarna undersöks. 10-percentilen går vid 70 % totalt, och vid 75 % för området utan yngre invånare. För säkerhets skull jämfördes områdenas äldre invånare, för att se om det fanns en strukturell skillnad mellan dem. Om till exempel det ena boendet vore ett servicehem, där invånarna inte tar emot lika mycket hjälp som på ett ålderdomshem, skulle det också kunna förklara skillnaden. Men när invånare under 75 år sorteras bort så är de två områdena mycket lika i sin åldersstruktur. Detta ger skäl att anta att området med äldre invånare är mer ett renodlat Å-område, och åldergränsen för de framtida undersökningarna av åldringarna sätts till 72 år som är 5-percentilen av de inflyttade i detta område. Det finns 1117 områden med invånare i denna åldersklass.

Faktorer som kan tänkas skilja åldringar på ålderdomshem från åldringar som bor hemma (med eller utan hemtjänst) är könsfördelning, dödsfrekvens, inflyttningsfrekvens och utflyttningsfrekvens. Även i de följande undersökningarna är siffrorna varje år sammanslagna så att befolkningen räknas fyra gånger medan de som flyttar och dör endast räknas det år då de flyttar eller dör.

### 4.4.1 Ålderdomshemmens könsfördelning

Befolkningen över 72 år undersöks och områdena rankas efter sin andel kvinnor. De två rena Å-områdena hamnar på 41:a och 45:e plats, det vill säga inom de 10 % högsta värdena. Andelen kvinnor är 79,3 % för vanliga områden och för rena Å-områdena 84,7 %. I den äldre generationen är parförhållandet norm och där är mannen också i de flesta fall är äldre än kvinnan. Eftersom kvinnor dessutom generellt sett lever längre än män, är kvinnan ofta frisk nog att själv vårda sin man när han blir gammal och sjuk. När kvinnan sedan går samma öde till mötes är mannen redan död och hon får då vård på ett ålderdomshem. Men detta är inte hela förklaringen. Som vi redan sett är åldersstrukturen för Å-områdena högerförskjuten jämfört med de vanliga områdena. Kanske beror den höga andelen kvinnor i rena Å-områdena mest på det faktum att befolkningen där är äldre.

#### 4.4.2 Ålderdomshemmens dödlighet

För att få komma till ett ålderdomshem måste man vara oförmögen att ta hand om sig själv. Därför kan man misstänka att åldringar på ålderdomshem är svagare och sjukare än de som bor hemma och därmed har större sannolikhet att dö. Men är man riktigt sjuk läggs man in på sjukhus, och många boende på ålderdomshem är där på grund av olika demenssjukdomar, som inte gör den gamla fysiskt sjuk. Vid rankning av kvoten *döda/hela befolkningen* placerar sig de två rena Å-områdena på 21:a respektive 62:a plats, det vill säga, inom de 10 % mest extrema värdena, men det är svårt att avgöra om det beror på att de är sjukare eller bara äldre. Dödsandelen är för vanliga områden 6,6 % och för de rena Å-områden 20,2 %.

Genomsnittsdödsåldern och genomsnittsåldern kontrolleras, och även skillnaden dem emellan. Genomsnittsdödsåldern befinner sig inom de 5 % högsta värdena och genomsnittsåldern inom de 1,5 % högsta värdena för de båda rena Å-områdena, och differansen mellan snittåldern och dödsåldern var inom den lägsta kvartilen. (För vanliga områden var medelåldern 80,2 år och genomsnittliga dödsåldern 84,7. I de rena Å-områdena var medelåldern 86,6 år och genomsnittliga dödsåldern 88,8 år.)

#### 4.4.3 Ålderdomshemmens flyttbenägenhet

Först undersöks befolkningsstorleken år för år, för att se om något av de rena Å-områdena verkar vara nybyggt, det skulle i så fall förklara en hög inflyttningsfrekvens. Medelbefolkningen i Å-områdena ökar med 48 personer under perioden 2001-2004, i princip står det ena området för hela förändringen, där det rör sig om en förändring med ungefär en fjärdedel.

Samma mått på flyttbenägenheten används för de gamla som för studenterna, det vill säga:  $in(ut)flyttningsandel = \frac{antalet\ flyttade}{hela\ befolkningen}$ . För de rena Å-områdena visade sig inflyttningsandelen vara beaktansvärd, områdena placerades på 10:e och 38:e plats, det vill säga, inom 4-percentilen. Inflyttningsandelen var 25,5 % för Å-områdena mot den totala inflyttningsandelen 3,8 % . När det gällde utflyttningsandelen lade sig de två rena Å-områdena på varsin sida om det totala snittet på 4,1% ( 2,4 % respektive 4,4% ) och när det gäller in- och utflyttning till och från utlandet är det svårt att säga något eftersom Å-områdena hade varken eller, liksom långt över hälften av områdena.

Antalet inflyttade per utflyttad rankades. De rena Å-områdena placerades på 20:e och 62:a plats, gott och väl ovanför 90-percentilen. De översta 16 områdena hade överhuvudtaget inga utflyttade vilket gör resultatet ännu starkare. De vanliga områdena hade 1,1 inflyttade per utflyttad, medan de rena Å-områdena hade hela 13,9 inflyttade per utflyttad!

#### 4.4.4 Åldringarnas ålder vid flytt

Den genomsnittliga åldern för inflyttare och utflyttare rankas. För den genomsnittliga åldern för inflyttare placerar sig de två rena Å-områdena på 40:e och 113:e plats, precis under 90-percentilen alltså. För utflyttade placerade sig de rena Å-områdena på 19:e och 23:e plats, bland de 2 % högsta värdena! För de vanliga områdena var både den genomsnittliga in- och utflyttningsåldern 81,4

år, för de rena Å-områdena var den genomsnittliga inflyttningsåldern 85,9 år och den genomsnittliga utflyttningsåldern 87,3 år.

Skillnaden mellan genomsnittlig ut- och inflyttningsålder, mellan genomsnittlig inflyttningsålder och medelålder, och mellan medelålder och genomsnittlig utflyttningsålder jämfördes utan att något intressant framkom.

## 4.5 Slutsatser efter första undersökningen

De första undersökningarna av basområdena har givit att de typiska egenskaperna för studentområdena är in- och utflyttningsandel och födelsefrekvens och att de typiska egenskaperna för ålderdomshem är inflyttningsandel, antal inflyttad per utflyttad och en hög ålder hos utflyttarna. Eventuellt kan könsfördelningen, dödsåldern och snittåldern också vara en indikator. Egenskaperna gäller alltså för ungdomar respektive gamla.

## 4.6 Förberedelser för analys

### 4.6.1 Kriterium för att ett område ska analyseras

Även om det skulle visa sig att det generellt sett lönar sig att använda särskilda antaganden för S- och Å-områden, gäller detta inte de allra minsta basområdena, där det bara bor några få människor. Ett så litet underlag ger osäkra skattningar oavsett vilka antaganden man gör, och ger ofta extremvärden som man inte vill ge samma vikt i analysen som de värden som genereras av de stora, 'stabila' områdena. Men var gränsen för ett 'alltför litet' underlag går är inte givet. I stället för att sätta en gräns för hela befolkningen sätts en gräns för befolkningen i respektive åldersgrupp. Först antas en ganska låg gräns, senare kommer även justeringar uppåt av denna gräns att utforskas. Gränsen för att det överhuvudtaget ska vara meningsfullt att utföra en analys för att lokalisera S-områden sätts till minst 20 ungdomar i åldersklassen (B) i snitt per år, och minst lika många gamla för att det ska vara någon mening att analysera och lokalisera ålderdomshem. Resten av befolkningen antas vara oberoende av äldreboenden och studenthem i närheten.

### 4.6.2 Kontroll av resultatet

För att kontrollera resultatet av den klassificeringsmetod som används, måste ett antal Å- och S-områden lokaliseras. Detta sker så att säga manuellt, genom att kommunerna tillhandahåller adresslistor till sina äldreboenden och bostadsförmedlingarna till sina studentbostäder. Adresserna har sedan lokaliserats geografiskt på kartor över basområdena, (inte genom adresslistor på vilka gator som tillhör varje basområde) och detta medför en felmarginal. Inaktuella eller tvärtom –alltför nya– listor gör felet än större. På grund av tidsbrist och svårigheter att få tag i data har inte alla kommuner kunnat undersökas utan endast Stockholm, Södertälje, Solna, Lidingö, Danderyd, Huddinge, Haninge, Botkyrka och Täby. Dessa kommuner innehåller 601 områden med minst 20 invånare över 72 år, och 634 områden med ungdomar mellan 19 och 34 år. Till ålderdomshem har äldreboenden, sevicehus, seniorboenden med åldersgränsen 65 år, gruppboenden för äldre och sjukhem räknats. Detta trots att det är osäkert om alla

som får en plats verkligen skriver sig på hemmet, och trots att inte alla boende på sjukhem är gamla. Till studenthem har alla typer av boenden där minst en hushållsmedlem måste vara studerande räknats.

### 4.6.3 De tidigare undersökningarnas resultat

Av de fem rena S-områdena tillhör ett Upplands Väsby kommun och är ej undersökt. Tre är redan hittade S-områden (Tempus i Åkeshov, Mårtensdal i Hammarby Sjöstad och Hotellet i Huddinge) och det femte är Albano. Sökningar på internet indikerar att det ska finnas 'tillfälliga bostäder' i Albano och eventuellt ska byggas fler där. Albano är det område som ökade sitt antal boende från 2,5 till 26 boende mellan 2003 och 2004. De båda rena Å-områdena visar sig också verkligen innehålla ålderdomshem. Områdena rankas också efter de egenskaper som ansågs typiska för student- respektive ålderdoms-områden, för att bekräfta att de verkligen är typiska. Andelen avvikande områden (S- och Å-områden) i percentilerna för de typiska egenskaperna bekräftar antagandet.

**Studentområden** Typiska egenskaper: inflyttningsandel+utflyttningsandel, samt födelseandel.

Andel S-områden totalt	6,6 %
Andel S-områden i födelseandelens 10-percentil	26,0 %
Andel S-områden i födelseandelens 5-percentil	35,3 %
Andel S-områden i in-och utflyttningsandelens 90-percentil	28,6 %
Andel S-områden i in-och utflyttningsandelens 95-percentil	39,5 %

**Ålderdomshemsområden** Typiska egenskaper: genomsnittsålder för utflyttare, inflyttningsandel och antal inflyttade per utflyttad.

Andel Å-områden totalt	21,3 %
Andel Å-områden i utflyttningsålderns 90-percentil	41,7 %
Andel Å-områden i utflyttningsålderns 95-percentil	43,6 %
Andel Å-områden i inflyttningsandelens 90-percentil	59,7 %
Andel Å-områden i inflyttningsandelens 95-percentil	63,9 %
Andel Å-områden i inflyttade/utflyttades 90-percentil	68,3%

# Kapitel 5

## Diskriminantanalys

### 5.1 En grafisk modell

Diskriminantanalys är en metod som används för att separera kända grupper från varandra, och för att klassificera observationer vars grupp man inte säkert känner till. Analysen används ofta när kausala samband är svåra att hitta och tolka.<sup>1</sup> Antag att variablerna i ett känt dataset utgör axlar i ett rum. Observationer i setet placerar sig då på den position i rummet som motsvarar dess värden för de olika variablerna. Antag att observationerna tillhör antingen den ena eller den andra av två grupper. Om dessa grupper skiljer sig åt med avseende på variablerna i setet kommer deras positioner bilda två kluster i rummet, och mellan dessa kluster kan en gräns dras upp. Kan gränsen beskrivas som ett hyperplan, rör det sig om linjär diskriminantanalys. Måste den snarare beskrivas som en krökt yta, rör det sig om icke-linjär diskriminantanalys. Vilken som används beror på gruppernas inbördes kovariansmatriser.

Analysen används för att se om det verkligen är en signifikant skillnad mellan grupperna, exempelvis ser man hur många observationer som hamnar på 'fel' sida gränsen. Resultatet kan användas för att kategorisera observationer vilkas grupptillhörighet man ej känner till.

Metoden förutsätter att observationerna har en multivariat normalfördelning, men är dock känd för att vara relativt robust mot mindre avvikelser från detta antagande. Eftersom grupptillhörigheten för det set som analyseras är känd, kan man omgående få ett facit på hur väl metoden fungerar för de data man har, exempelvis genom att utsluta en observation i taget, utföra analysen och se vilken grupp den enligt analysen skulle placeras i. Felkvoten i en sådan korsvalidering ger en god skattning om metodens felkvot totalt (dock ej helt väntevärdesriktig.)<sup>2</sup>

#### 5.1.1 En algebraisk modell

Låt  $\pi_1$  och  $\pi_2$  vara två populationer (grupper). Låt vidare  $\mathbf{x} = (x_1 \ x_2 \ \dots \ x_k)$  vara en observation med  $k$  variabler tillhörande  $\pi_i$ .

Låt  $f_i(\mathbf{x})$  vara täthetsfunktionen för populationen  $\pi_i$ . Låt  $\Omega$  vara utfallsrummet

<sup>1</sup>A Handbook of Statistical Analyses using SAS

<sup>2</sup>A Handbook of Statistical Analyses using SAS

för observationerna.  $\Omega$  ska då på ett optimalt sätt delas i två (och endast två) rum,  $R_1$  och  $R_2$  där observationerna tillhörande  $R_1$  klassificeras som tillhörande  $\pi_1$  och  $R_2$  tillhörande  $\pi_2$ .

Sannolikheten för att en observation  $\mathbf{x} \in \pi_1$  felklassificeras är då lika med sannolikheten att  $\mathbf{x}$  befinner sig i  $R_2$ . Denna sannolikhet betecknas  $P(2 | \pi_1)$ . Sannolikheten att en observation  $\mathbf{x} \in \pi_2$  felplaceras är då lika med sannolikheten att  $\mathbf{x}$  befinner sig i  $R_1$  och betecknas  $P(1 | \pi_2)$ .

$$P(2 | \pi_1) = \int_{R_2} f_1(\mathbf{x}) d\mathbf{x}$$

$$P(1 | \pi_2) = \int_{R_1} f_2(\mathbf{x}) d\mathbf{x}$$

Låt  $p_1$  vara sannolikheten för att en godtycklig observation ska tillhöra  $\pi_1$  och  $p_2$  vara sannolikheten att den tillhör  $\pi_2$ . Därmed är sannolikheten att en godtycklig observation felplaceras

$$\begin{aligned} P(\text{felplacerad observation}) &= P(2 | \pi_1) * p_1 + P(1 | \pi_2) * p_2 = \\ &= \int_{R_2} f_1(\mathbf{x}) d\mathbf{x} p_1 + \int_{R_1} f_2(\mathbf{x}) d\mathbf{x} * p_2 \end{aligned} \quad (5.1)$$

Eftersom  $\Omega = R_1 \cup R_2$  gäller följande:

$$\begin{aligned} 1 &= \int_{\Omega} f_2(\mathbf{x}) d\mathbf{x} = \int_{R_1} f_2(\mathbf{x}) d\mathbf{x} + \int_{R_2} f_2(\mathbf{x}) d\mathbf{x} \iff \\ &\iff \int_{R_1} f_2(\mathbf{x}) d\mathbf{x} = 1 - \int_{R_2} f_2(\mathbf{x}) d\mathbf{x} \end{aligned}$$

Detta resultat används i ekvation 5.1 och ger då att

$$\begin{aligned} P(\text{felplacerad observation}) &= p_1 * \int_{R_2} f_1(\mathbf{x}) d\mathbf{x} + p_2 * (1 - \int_{R_2} f_2(\mathbf{x}) d\mathbf{x}) = \\ &= p_1 * \int_{R_2} f_1(\mathbf{x}) d\mathbf{x} + p_2 - p_2 * \int_{R_2} f_2(\mathbf{x}) d\mathbf{x} = \\ &= \int_{R_2} [p_1 * f_1(\mathbf{x}) - p_2 * f_2(\mathbf{x})] d\mathbf{x} + p_2 \end{aligned} \quad (5.2)$$

Ekvation 5.2 är sannolikheten för att en godtycklig observation felklassificeras och denna sannolikhet ska minimeras. Sannolikheten  $p_2$  är konstant och täthetsfunktionen alltid positiv. Gränsen för  $R_2$  ska väljas så att volymen under grafen  $[p_1 * f_1(\mathbf{x}) - p_2 * f_2(\mathbf{x})]$  blir så liten som möjligt. Det vill säga; så länge ytan är negativ skall området klassificeras som  $R_2$ .

$$R_2 : \quad p_1 * f_1(\mathbf{x}) - p_2 * f_2(\mathbf{x}) \leq 0 \iff \frac{f_2(\mathbf{x})}{f_1(\mathbf{x})} \geq \frac{p_1}{p_2}$$

och eftersom  $R_2$  och  $R_1$  helt uppfyller  $\Omega$  gäller

$$R_1 : \quad \frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} > \frac{p_2}{p_1} \quad (5.3)$$



## 5.2 Klassificering med flerdimensionell normalfördelning

Låt  $\pi_1$  och  $\pi_2$  vara två flerdimensionellt normalfördelade populationer med  $n_1$  och  $n_2$  observationer å  $m$  variabler. Låt  $f_1(\mathbf{x})$  och  $f_2(\mathbf{x})$  vara täthetsfunktionerna för  $\pi_1$  och  $\pi_2$ . Antag att de två fördelningarna har kovariansmatris  $M_1$  respektive  $M_2$ , och väntevärdesvektor  $\mu_1$  respektive  $\mu_2$ .

Då en täthetsfunktionen för en observation  $\mathbf{x}$  av

$$f_i(\mathbf{x}) = \frac{1}{(2\pi)^{k/2} |M_i|^{1/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu_i)' M_i^{-1} (\mathbf{x} - \mu_i) \right\}$$

för  $i = 1, 2$

Vi vill använda resultatet 5.3 och utvecklar kvoten:

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} = \frac{(2\pi)^{m/2} |M_2|^{1/2}}{(2\pi)^{m/2} |M_1|^{1/2}} * \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu_1)' M_1^{-1} (\mathbf{x} - \mu_1) - \left( -\frac{1}{2} (\mathbf{x} - \mu_2)' M_2^{-1} (\mathbf{x} - \mu_2) \right) \right\}$$

klammern utvecklas som följer:

$$\begin{aligned} & \left\{ -\frac{1}{2} \left( (\mathbf{x}' M_1^{-1} \mathbf{x} - 2\mu_1' M_1^{-1} \mathbf{x} + \mu_1' M_1^{-1} \mu_1) - (\mathbf{x}' M_2^{-1} \mathbf{x} - 2\mu_2' M_2^{-1} \mathbf{x} + \mu_2' M_2^{-1} \mu_2) \right) \right\} \\ &= \left\{ -\frac{1}{2} (\mathbf{x}' M_1^{-1} \mathbf{x} - \mathbf{x}' M_2^{-1} \mathbf{x} - 2\mu_1' M_1^{-1} \mathbf{x} + 2\mu_2' M_2^{-1} \mathbf{x} + \mu_1' M_1^{-1} \mu_1 - \mu_2' M_2^{-1} \mu_2) \right\} \\ &= \left\{ -\frac{1}{2} (\mathbf{x}' (M_1^{-1} - M_2^{-1}) \mathbf{x} - 2(\mu_1' M_1^{-1} - \mu_2' M_2^{-1}) \mathbf{x} + \mu_1' M_1^{-1} \mu_1 - \mu_2' M_2^{-1} \mu_2) \right\} \end{aligned}$$

Vi inför nu konstanten  $\vartheta = \frac{|M_2|^{1/2}}{|M_1|^{1/2}} * \exp \left\{ \frac{1}{2} (\mu_2' M_2^{-1} \mu_2 - \mu_1' M_1^{-1} \mu_1) \right\}$  och får då att

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} = \vartheta * \exp \left\{ -\frac{1}{2} (\mathbf{x}' (M_1^{-1} - M_2^{-1}) \mathbf{x}) + (\mu_1' M_1^{-1} - \mu_2' M_2^{-1}) \mathbf{x} \right\}$$

Vi använder 5.3 och logariterar uttrycket. Vi får att en observation  $\mathbf{x}$  bör allokeras till  $R_1$  om följande gäller:

$$\ln \vartheta - \frac{1}{2} (\mathbf{x}' (M_1^{-1} - M_2^{-1}) \mathbf{x}) + (\mu_1' M_1^{-1} - \mu_2' M_2^{-1}) \mathbf{x} > \ln p_2 - \ln p_1 \quad (5.4)$$

Annars allokeras  $\mathbf{x}$  till  $R_2$

Detta är den så kallade diskriminantfunktionen. Den kan också skrivas på formen:

$$\alpha_1 * x_1 + \alpha_2 * x_2 + \dots + \alpha_m * x_m + \beta_1 * x_1^2 + \beta_2 * x_1 * x_2 \dots \beta_m * x_1 * x_m \dots \kappa_1 * x_m^2 + C > 0$$

där  $C$  är en konstant. Här ser man tydligare att funktionen är icke-linjär.

### 5.2.1 Samma kovariansmatrix

Om  $\pi_1$  och  $\pi_2$  kan antas ha samma kovariansmatrix  $M_1 = M_2 = M$  går resultatet 5.4 att förenkla genom att

$$-\frac{1}{2} (\mathbf{x}'(M_1^{-1} - M_2^{-1})\mathbf{x}) = 0$$

och

$$(\mu_1' M_1^{-1} - \mu_2' M_2^{-1})\mathbf{x} = (\mu_1 - \mu_2)' M_2^{-1} \mathbf{x}$$

Konstanten  $\vartheta = \frac{|M_2|^{1/2}}{|M_1|^{1/2}} * \exp \left\{ -\frac{1}{2} (\mu_2' M_2^{-1} \mu_2 - \mu_1' M_1^{-1} \mu_1) \right\}$  kan i detta fall skrivas som

$$1 * \exp \left\{ -\frac{1}{2} ((\mu_1 - \mu_2)' M^{-1} (\mu_1 + \mu_2)) \right\}$$

och resultat 5.3 ger då att vi ska allokeras en observation  $\mathbf{x}$  till  $R_1$  om följande gäller:

$$\left\{ -\frac{1}{2} ((\mu_1 - \mu_2)' M^{-1} (\mu_1 + \mu_2)) + ((\mu_1 - \mu_2)' M^{-1}) \mathbf{x} \right\} > \ln p_2 - \ln p_1 \quad (5.5)$$

Annars allokeras  $\mathbf{x}$  till  $R_2$ . Denna diskriminantfunktion blir linjär, vilket syns tydligt om man skriver olikheten 5.5 på formen

$$\alpha * x_1 + \beta * x_2 + \dots + \kappa * x_k + \text{konstant} > 0$$

### 5.2.2 Skatta väntevärde och kovarians

Låt  $\mathbf{X}_1$  och  $\mathbf{X}_2$  vara två matriser som representerar observationer av två populationer  $\pi_1$  och  $\pi_2$  i diskriminantanalysen. Vi har  $n_1$  och  $n_2$  observationer á  $m$  variabler, vilket ger att  $\mathbf{X}_1$  är en  $(n_1 * m)$ -matrix och  $\mathbf{X}_2$  är en  $(n_2 * m)$ -matrix. Den skattade väntevärdesvektorn för  $\pi_i$  är  $\widehat{\mu}_i$  och dess element  $\widehat{\mu}_{ij}$  skattas som följer

$$\widehat{\mu}_{ij} = \frac{1}{n_i} \sum_{k=1}^{n_i} x_{ijk}$$

$$i = 1, 2$$

$$j = 1 \dots m$$

$$k = 1 \dots n_i$$

Kovariansmatrix skattas på följande sätt (för läslighetens skull har indexeringen för population, i lagts bort):

$$\widehat{S}_{jl} = \frac{1}{n-1} \sum_{k=1}^n (x_{jk} - \mu_{jk})(x_{lk} - \mu_{lk})$$

$$k = 1 \dots n \quad j = 1 \dots m \quad l = 1 \dots m$$

Om skäl finns för att anta att populationerna har samma kovariansmatrix skattas denna med en polad kovariansmatrix

$$S_\rho = \left[ \frac{n_1 - 1}{n_1 + n_1 - 2} \right] S_1 + \left[ \frac{n_2 - 1}{n_1 + n_1 - 2} \right] S_2$$

## 5.3 Tillämpbarhet på det aktuella problemet

Analysen i detta fall måste utföras i två separata delar, en för att klassificera områden med studentområden, och en för att klassificera ålderdomshem. De två grupperna är 'vanligt område' och 'studenthemsområde', respektive 'vanligt område' och 'äldreområde'.

Det finns visserligen metoder där observationer klassificeras i fler än två klasser, men eftersom oberoende med andra åldersklasser antagits är utfallsrummen för ålderdomshemsanalysen och studenthemsanalysen inte desamma. (Variablerna för åldringar äldre än 72 år används inte i analysen för studenter och vice versa). Vilka variabler som bör användas är inte självklart. Att göra en analys på hela det ursprungliga datasetet är av beräkningsskäl omöjligt, därför måste ett nytt dataset med utvalda variabler konstrueras.

### 5.3.1 Klassificeringsproblematiken

Att dela in områden i 'vanliga' och 'avvikande' kan vid en första anblick verka trivialt. Ett avvikande område är helt enkelt ett område som har ett student- eller ålderdomshem. Vid närmare eftertanke kanske denna indelning inte är så självklar. Ett område med många invånare och endast en liten 'institution' kommer inte systematiskt skilja sig från ett vanligt område och befolkningsutvecklingen kan utan problem predikteras med vanliga antaganden. Kanske borde S-områden och Å-områden omdefinieras så att de i stället för att innehålla ett studenthem eller ålderdomshem, ska ha minst en viss kvot av sina invånare boende på studenthem eller ålderdomshem. I så fall borde denna kvot bestämmas genom att undersöka hur olika gränser påverkar prediktionen, och det kan ej genomföras här. Däremot kan vi undersöka hur olika gränser påverkar klassificeringen. Härtill kommer också ett praktiskt problem: Svårigheten att uppskatta hur många som bor på respektive institution. I flera fall medför felplaceringar och gissningar att andelen boende på institution överstiger det totala antalet boende i området, vilket väl säger något om tillförlitligheten i skattningarna.

## 5.4 Val av variabler

Vid den tidigare undersökningen av studentområden framkom att inflyttnings- och utflyttningsandel och födda barn per kvinna är intressanta faktorer för vidare analys. Därtill läggs åldersfördelning, i form av 10- och 90-percentiler, kvartiler och median, för invånarnas, mödarnas, inflyttarnas och utflyttarnas ålder och andelen unga (19-34 år) av den totala befolkningen i området. För åldersgruppen 19-34 år undersöks alltså sammanlagt 24 förklarande variabler, de förklarar om ett område är ett S-område, eller ett vanligt område.

För att välja variabler till analysen av de äldre utgår vi från dem som visade sig vara typiska (stycke 4.5). Det konstateras att de områden som undersöktes verkligen är Å-områden, och att de variabler som var typiska för dem även är typiska även för de äldreboenden som nu tillkommit. Följande variabler väljs ut för analys av befolkningen över 72 år (för att finna äldreboenden): medianen och kvartiler för total ålder, kvinnornas ålder, männens ålder, dödsåldern och utflyttningsåldern. Därtill andelen kvinnor, andelen inflyttade, utflyttade, döda, antalet inflyttade per utflyttad, befolkningsandelen över 72 år och den totala

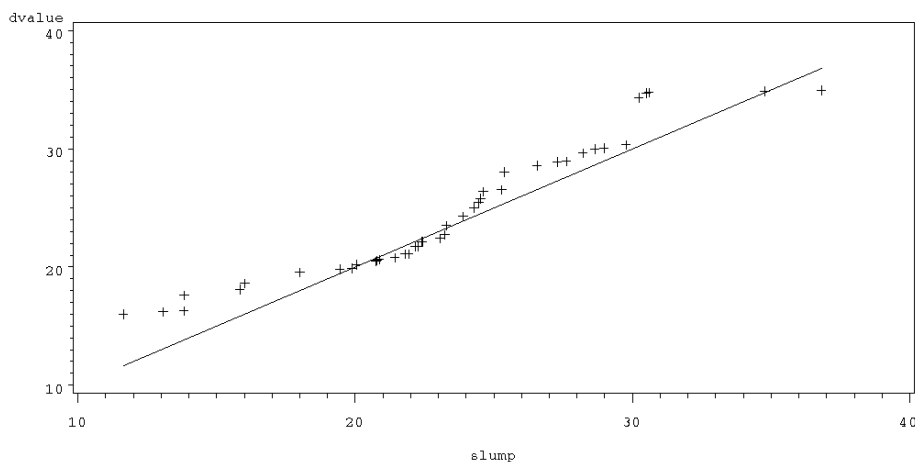
befolkningen (sammanlagt 22 förklarande variabler).

### 5.4.1 Variablernas fördelning

Diskriminantanalys förutsätter multivariat normalfördelning, och även om metoden anses vara robust mot avvikelser från denna förutsättning, undersöks variablerna enligt följande metod: Antag att  $\mathbf{x}_1 \dots \mathbf{x}_n$  är multivariat normalfördelade observationer á  $m$  variabler. Låt  $d_k = (\mathbf{x}_k - \bar{\mathbf{x}})' \mathbf{S}^{-1} (\mathbf{x}_k - \bar{\mathbf{x}})$ . Då är  $d_1 \dots d_n$   $\chi^2$ -fördelade med  $m - 1$  frihetsgrader.<sup>3</sup>

$k = 1 \dots n$

Om man låter slumpa fram  $n$  stycken  $\chi^2(m - 1)$ -fördelade värden och storleksordnar dessa, och plotar dem mot i storleksordning motsvarande  $d_k$ -värde skall dessa punkter forma något som kan tolkas som en rät linje med lutningen 1. Avvikelser från en rät linje indikerar avvikelser från normalfördelning. Figur 5.1 visar studenternas d-värden. Det är alltså 42 observationer som, plottade mot  $\chi^2$ -fördelade slumpstal, lägger sig fint runt hjälplinjen med lutningen 1.

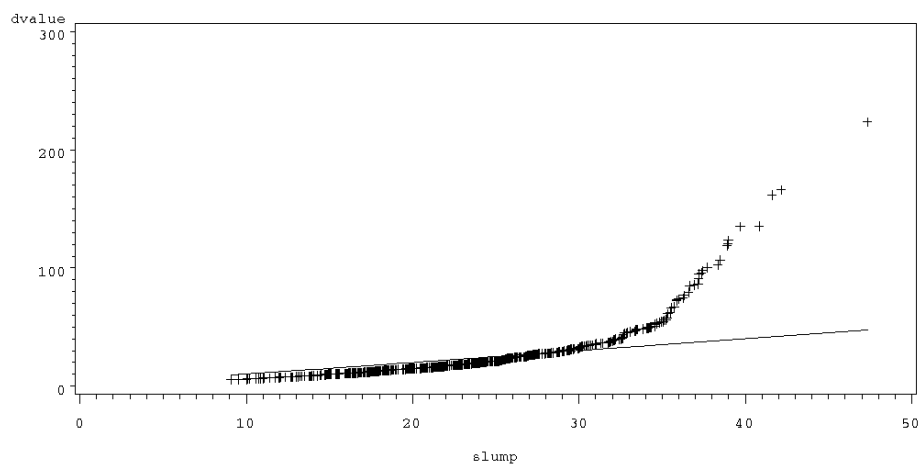


Figur 5.1: Studenternas d-värden

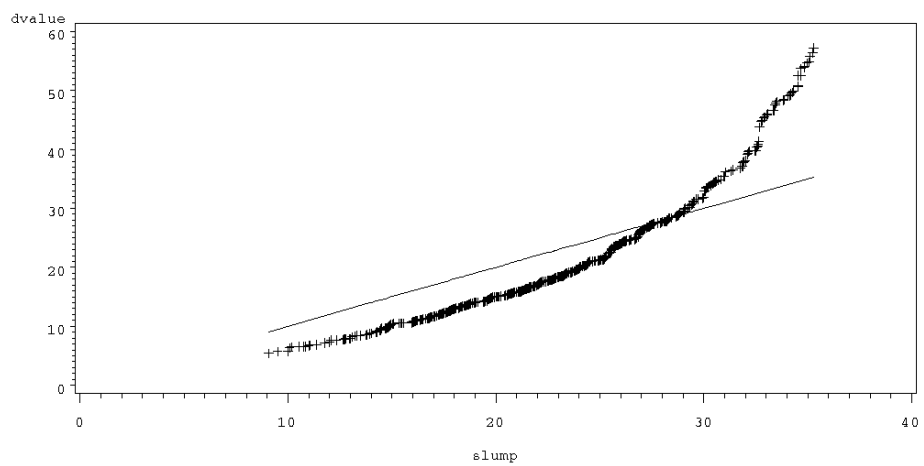
Figur 5.2 visar hur värdena för de 592 områden som inte innehåller studentområden följer den räta linjen till en början, för att sedan vika av med en helt ny lutning. På grund av de höga värdena avvikelse är det svårt att se de lägre värdenas anpassning tydligt. Bortser vi från de 29 högsta värdena, får vi figur 5.3, där man tydligt ser att de unga icke-studerandes d-värden snarast är exponentiellt växande jämfört med de mot  $\chi^2$ -fördelade slumpstalen.

Figur 5.4 visar Å-områdenas 128 d-värden plottade mot  $\chi^2$ -fördelade slumpstal. Även de kröker av uppåt, men när man betraktar figur 5.5 där de 11 högsta d-värdena tagits bort har kurvan visserligen inte lutningen 1, men skulle i pincip kunna tolkas som rät med en 'knäck'.

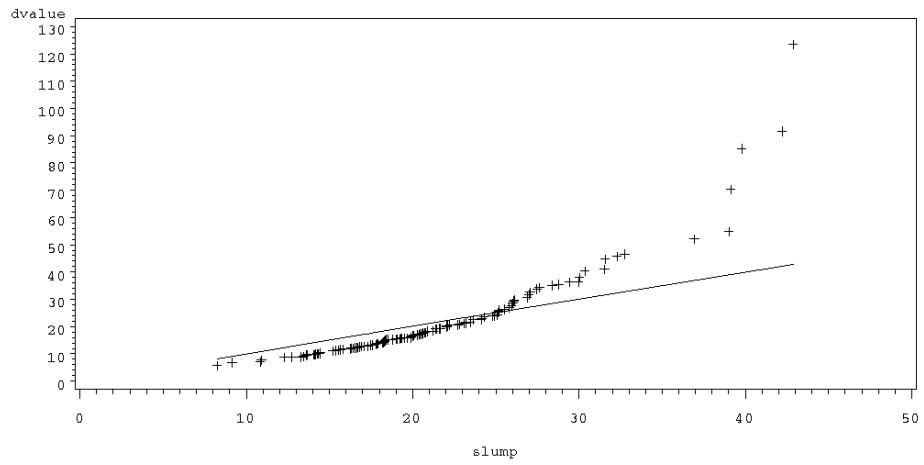
<sup>3</sup>Applied Multivariate Statistical Analysis



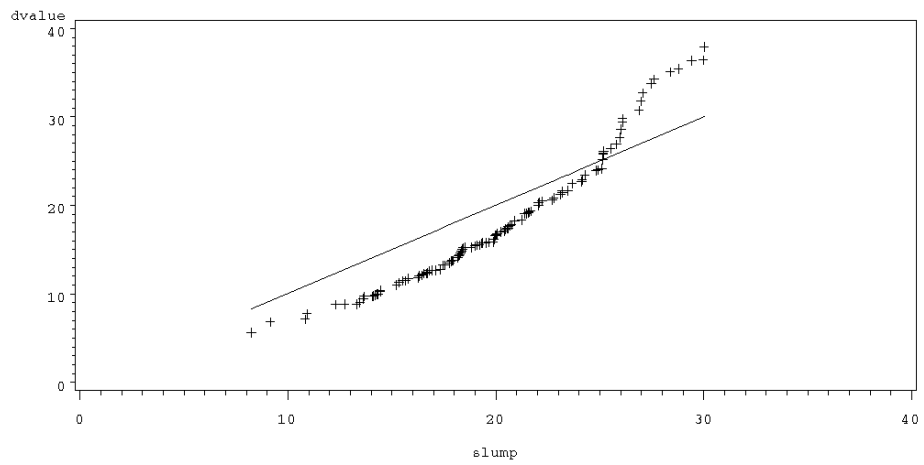
Figur 5.2: Unga icke-studenters d-värden



Figur 5.3: Unga icke-studenters lägre d-värden

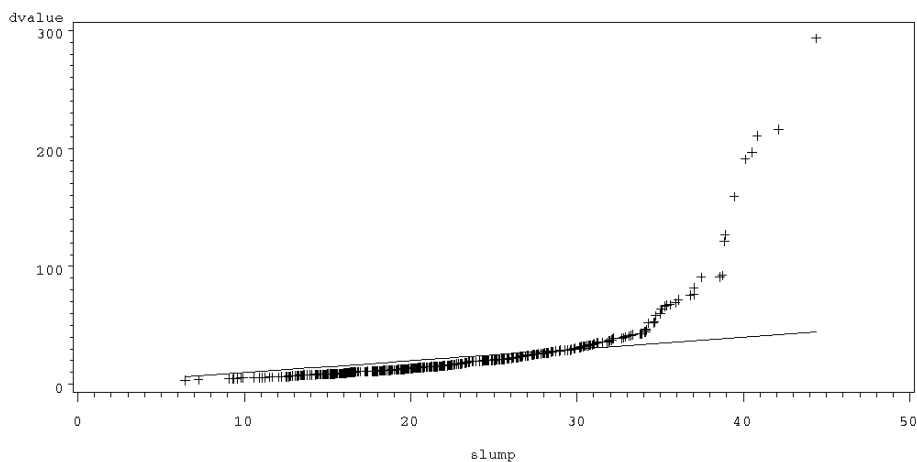


Figur 5.4: Å-områdenas d-värden

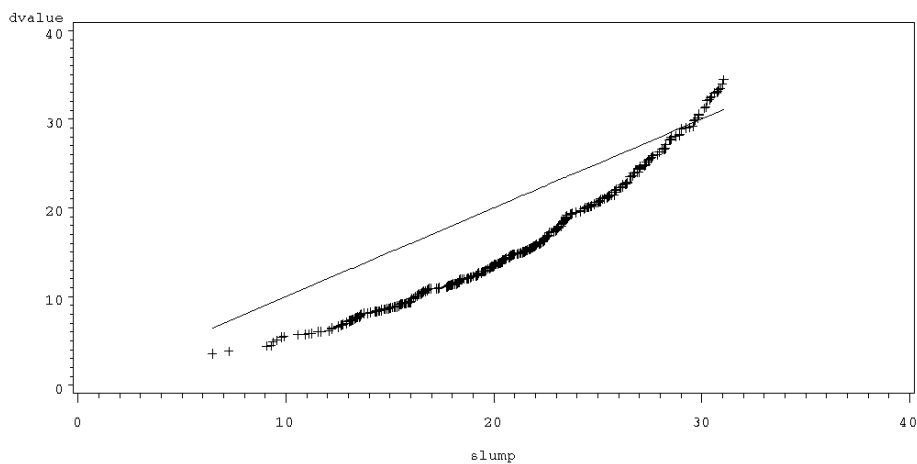


Figur 5.5: Å-områdenas lägre d-värden

De vanliga områdenas d-värden för de variabler som är intressanta i avseende att hitta ålderdomshem, visas i figur 5.6. Återigen en långt ifrån rät linje, och betraktar vi de 420 lägsta d-värdena syns att de är (precis som för de variabler som är intressanta när de gäller att finna studentområden) exponentiellt växande när de plottas mot  $\chi^2$ -fördelade slumpstal.



Figur 5.6: Vanliga områdets äldre invånares d-värden



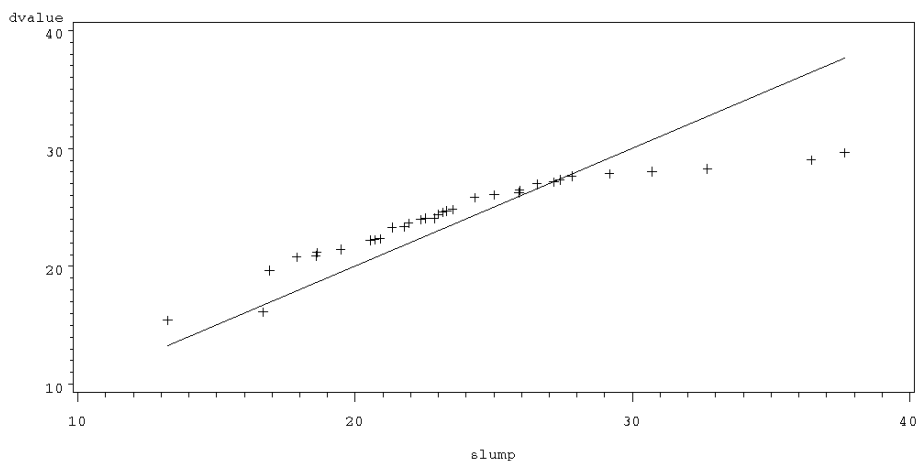
Figur 5.7: Vanliga områdets äldre invånares lägre d-värden

#### 5.4.2 Variablernas fördelning för de större områdena

I undersökningen ovan har alla områden som överhuvudtaget har invånare i den aktuella åldersgruppen behandlats. Graferna har visserligen en krökt form,

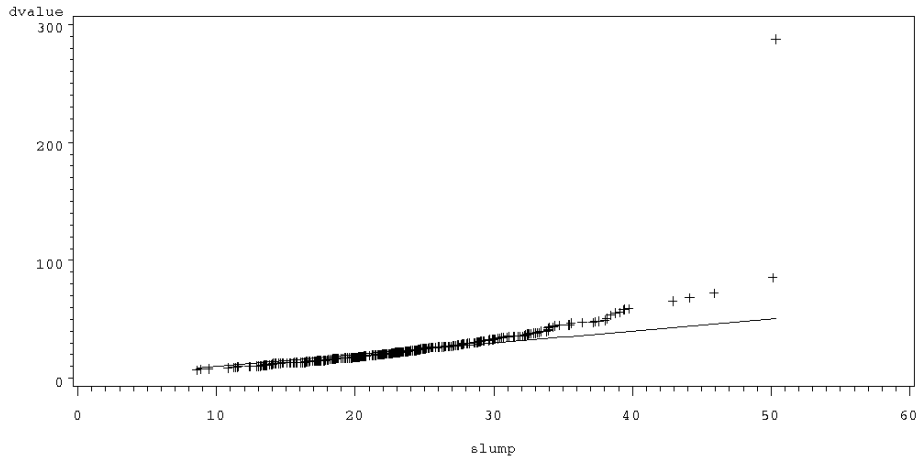
men det som är mest i ögonfallande är att områdenas högsta d-värden är mycket högre än de högsta  $\chi^2$ -värdena. D-värdet är ju i någon mening summan av avvikelserna från väntevärdet  $\mu$ . De största avvikelserna borde återfinnas i de befolkningsmässigt minsta områdena, som naturligt ofta får en homogen befolkning med extrema värden. De befolkningsmässigt minsta områdena är ju, som tidigare konstaterats, de som är minst intressanta att undersöka i detta arbete, eftersom prognosen kommer att vara osäker hur den än utförs. Till en följd av detta undersöks nu de större områdenas d-värden.

D-värdena för de variabler som ansetts intressanta för undersökningen av S-områden räknas ut för områden större än 200 invånare, och plottas mot en ny omgång  $\chi^2$ -fördelade slumpstal. För S-områdena (figur 5.8) är resultatet längre bort från en rät linje, men de vanliga områdena (figur 5.9) får ett rätare resultat.



Figur 5.8: S-områden med över 200 invånare

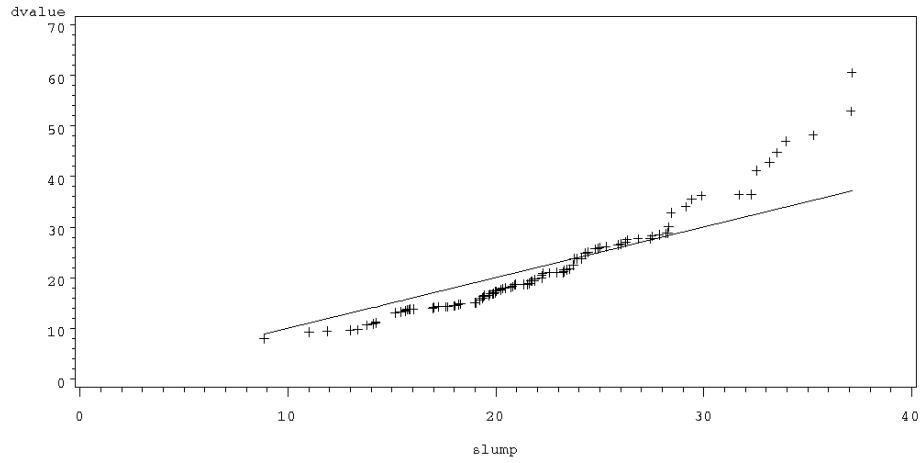




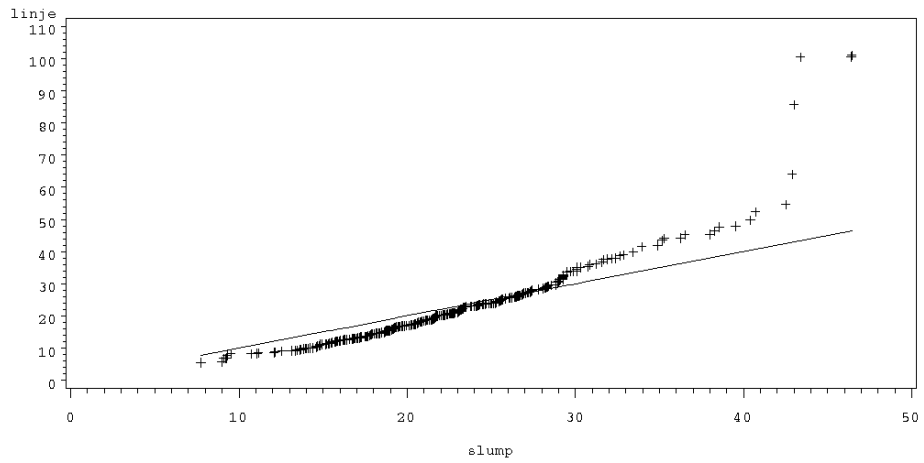
Figur 5.9: Vanliga områden med över 200 invånare

D-värdena för de variabler som ansetts intressanta för undersökningen av Å-områden räknas ut för Å-områden större än 150 invånare och vanliga områden större än 80 invånare, och plottas mot en ny omgång  $\chi^2$ -fördelade slumpstal. För både Å-områdena (figur 5.10) och de vanliga områdena (figur 5.11) får vi ett rätare resultat, om än inte med lutningen 1.

De olika gränserna för Å-områden, S-områden och vanliga kan tyckas inkonsekvent, men meningen framför allt är att se om det finns en tendens mot större överensstämmelse med  $\chi^2$ -fördelningen, när man tar bort de 15-20 % minsta områden. Gränserna för vilka områden som är användbara i analysen kommer undersökas mer senare. Alla kuvor blir rätare utom den för S-områdena, när man utesluter de mindre områdena. Att variablerna är perfekt multivariat normalfördelade kan nog uteslutas utifrån dessa kurvor, men förhoppningsvis är det en rimlig approximation.



Figur 5.10: Å-områden med över 150 invånare



Figur 5.11: Vanliga områden med över 80 invånare

## Kapitel 6

# Diskriminantanalys på Studentområden

### 6.1 Få studentområden

Det första som konstateras är att i de åtta kommuner som undersökts (Stockholm, Solna, Lidingö, Haninge, Huddinge, Södertälje, Täby och Botkyrka) har endast funnits 42 studentområden med tillräckligt stor befolkning. (Många studentområden uppfördes i slutet på denna period eller under 2005-2006). Det är endast 6,6 % av det totala antalet områden med fler än 80 invånare ackumulerat i dessa kommuner. Kanske är detta en för liten del för att diskriminantanalys ska kunna ge tillförlitliga resultat?

### 6.2 En första analys

Till en början är den  $\{0, 1\}$ -variabel som ska undersökas helt enkelt:

typ 1: innehåller studentområde  
typ 0: innehåller ej studentområde

Det visar sig genom ett f-test att det är signifikant skillnad mellan de två typernas kovariansmatriser, och därför bör en kvadratisk modell väljas. När områdena sedan klassificeras enligt gjorda analys, ser resultatet ut så här:

	klassificerad som 0	klassificerad som 1
tillhörande 0	581	11
tillhörande 1	6	36

Resultatet är ju strålande, bara  $11 + 6$  av totalt 634 områden klassificerades inkorrekt. Det betyder att 97,3 % av områdena placerades korrekt. Men om man betänker att även gissningen 'klassificera alla områden som typ 0' skulle ge att 93,3 % av områdena placerades korrekt, och att funktionen har framtagits från det dataset som sedan klassificerats så säger denna tabell kanske inte så mycket.

I stället utesluts ett område i taget, analysen genomförs och därefter klassificeras det uteslutna området. Detta kallas korsvalidering och ger följande tabell.

	klassificerad som 0	klassificerad som 1
tillhörande 0	580	12
tillhörande 1	35	7

Nu är hela 83,3 % av studentområdena felklassificerad och endast 92,6 % av alla områden är korrekt klassificerade vilket gör en gissningen 'klassa alla som 0' bättre totalt sett.

### 6.2.1 Overfitting

Vi har att göra med fenomenet overfitting. Lösningen att ta bort ovesäntliga variabler. Genom en stegvis valprocedur i SAS, där i varje steg den variabel som har högst förklaringsgrad väljs ut och därefter undersöks tillsammans med de variabler valda i ett tidigare steg, och där de variabler som är signifikanta på en viss nivå behålls som utvalda, kan detta ske på ett strukturerat sätt. Några vanliga signifikansgränser undersöks:

Signifikanta variabler (gräns 0,15): inflyttningsålderns tiopercentil och median, befolkningen nedre ålderskvartil, utflyttningsålderns tiopercentil, median och 90-percentil, mödrarnas ålders tiopercentil och nedre kvartil samt andelen unga, totala antalet unga och inflyttningsandelen.

	klassificerad som 0	klassificerad som 1
tillhörande 0	565	27
tillhörande 1	26	16

Signifikanta variabler (gräns 0,1): andelen unga, inflyttningsandelen, befolkningens nedre ålderskvartil, totala antalet unga och utflyttningsålderns 10-percentil och median.

	klassificerad som 0	klassificerad som 1
tillhörande 0	572	20
tillhörande 1	26	16

Signifikanta variabler (gräns 0,05): andelen unga, totala antalet unga, inflyttningsålderns 10-percentil och nedre kvartil, utflyttningsålderns median, mödrarnas ålders tiopercentil och nedre kvartil, inflyttningsandelen.

	klassificerad som 0	klassificerad som 1
tillhörande 0	568	24
tillhörande 1	24	18

Signifikanta variabler (gräns 0,01): utflyttningsålderns övre kvartil, andelen unga och inflyttningsandelen.

	klassificerad som 0	klassificerad som 1
tillhörande 0	576	16
tillhörande 1	31	11

Som synes verkar det inte göra någon större skillnad på resultatet vilken signifikansgräns vi väljer. Det bästa resultatet ges av signifikansgränsen 0,10. Men övermättade modeller ska undvikas och eftersom signifikansgränsen 0,01 ger ett relativt bra resultat görs nästa undersökning för båda signifikansnivåerna.

## 6.2.2 Val av kriterium

För att vara ett S-område måste en viss andel av invånarna mellan 19-34 år bo i studenthem. Denna andel ändras stegvis, och för varje steg kommer två S-områden omdefinieras till 'vanliga'.

signifikansnivå 0,10							signifikansnivå 0,01				
and.S	S-omr	SR	SF	VR	VF	fel	SR	SF	VR	VF	fel
>0,02	40	14	26	575	19	<b>45</b>	11	29	578	16	<b>45</b>
>0,035	38	12	26	577	19	<b>45</b>	11	27	581	15	<b>42</b>
>0,050	36	11	25	581	17	<b>42</b>	12	24	581	18	<b>42</b>
>0,055	34	13	21	579	21	<b>42</b>	11	23	585	15	<b>38</b>
>0,061	32	13	19	586	16	<b>35</b>	11	21	587	15	<b>36</b>
>0,075	30	14	16	589	15	<b>31</b>	10	20	589	15	<b>35</b>
>0,095	28	14	14	595	11	<b>25</b>	10	18	591	15	<b>33</b>
>0,200	26	12	14	598	10	<b>24</b>	12	14	592	16	<b>30</b>
>0,215	24	11	13	601	9	<b>22</b>	10	14	602	8	<b>22</b>
>0,240	22	11	11	603	9	<b>20</b>	11	11	597	15	<b>26</b>
>0,260	20	10	10	606	8	<b>18</b>	11	9	599	15	<b>24</b>

and. S: Andelen som måste bo på studenthem för att området ska räknas som typ 1.

S-omr.: Antalet S-områden.

SR: Rättklassificerad S-områden

SF: Felklassificerade S-områden

VR: Rättklassificerade vanliga områden

VF: Felklassificerade vanliga områden

Fel: Antalet felklassificerade områden

När något av de fettade talen understiger talet i kolumnen 'S-omr', betyder detta att analysen gav en bättre klassificering än gissningen 'alla områden är vanliga'. Med signifikansnivå 0,01 inträffar detta endast en gång. Med signifikansnivån 0,10 blir analysen bättre än gissningen när man väljer gränsen 0,095 eller högre för andelen studenter ett område måste ha för att anses vara ett S-område.

### 6.2.3 Val av befolkningsstorlek

Kravet för att ett område överhuvudtaget ska analyseras har hittills varit att minst 20 personer i åldern 19-34 år bor där. De skattade egenskaperna torde vara säkrare ur statistisk synpunkt hos stora områden med många människor i den aktuella åldersgruppen. Men i diskriminantanalysen läggs inte mer tyngd på de större områdena än på de mindre. Kanske kan det vara meningsfullt att höja gränsen för minsta antalet invånare i åldersgruppen 19-34 år, så att små och osäkra områden sorteras bort.

Analysen med olika minsta antal invånare. Signifikansnivå 0,1.

Minsta ant. inv. mellan 19 och 34 år	omr	s-omr.	SR	SF	VR	VF	fel
20	634	42	16	26	572	20	<b>46</b>
50	591	40	13	27	535	16	<b>43</b>
100	524	37	18	19	470	17	<b>36</b>
150	455	35	16	19	405	15	<b>34</b>
200	396	33	18	15	351	12	<b>27</b>
250	357	30	16	14	316	11	<b>27</b>

För en gräns på 100 invånare i åldersgruppen, eller mer, är resultatet bättre än en gissning. Bäst är resultatet för gränsen 200 invånare per år.

### 6.3 Kostnad för felklassificering

Ovan har konstaterats att modellen är i princip värdelös för våra syften, eftersom påståendet 'Alla områden är vanliga' skulle ge ett bättre, lika bra eller möjligen marginellt sämre resultat än analysen. Dock kan man tänka sig en situation där det var viktigare att klassificera S-områdena rätt, än de vanliga områdena. Det brukar modelleras så att felklassificeringar genererar en kostnad beroende på vilket område som felklassificeras. Kostnaderna kan då inberäknas i analysen, så att fler områden klassificeras till den klass vars kostnad för felklassificering är högre. Målet med analysen blir då att hitta den lägsta förväntade kostnaden för en klassificering av ett område. Vi antar att det inte kostar något att klassificera ett område rätt, alltså, endast felklassificeringen innebär en kostnad, som motsvaras av följande formel:

$$E(\text{Kostnad för felklassificerad observation}) = \\ = p(2|\pi_1) * p_1 * C(2|\pi_1) + P(1|\pi_2) * p_2 * C(1|\pi_2)$$

Där  $C(2|\pi_1)$  är kostnaden för att klassificera ett område från population 1 som tillhörande population 2 och  $C(1|\pi_2)$  är kostnaden för att klassificera ett område från population 2 som tillhörande population 1. Jämför med ekvation 5.1. Kostnaden för felklassificering är alltså konstant, precis som sannolikheten för att ett godtyckligt område tillhör den ena eller andra populationen. Tas konstanterna med i utvecklingen 5.1 fås alltså analogt med ekvation 5.3 att ett godtyckligt område  $\mathbf{x}$  ska allokeras till  $R_2$  om

$$\frac{f_2(\mathbf{x})}{f_1(\mathbf{x})} \geq \frac{p_1 * C(2|\pi_1)}{p_2 * C(1|\pi_2)}$$

och till  $R_1$  om

$$\frac{f_1(\mathbf{x})}{f_2(\mathbf{x})} > \frac{p_2 * C(1|\pi_2)}{p_1 * C(2|\pi_1)}$$

I analysen införlivas kostnaderna enklast så att man väljer sannolikheter,  $p_1^*$  för  $\pi_1$  och  $p_2^*$  för  $\pi_2$  så att

$$\frac{p_1^*}{p_2^*} = \frac{p_1^*}{1 - p_1^*} = \frac{p_1 * C(2 | \pi_1)}{p_2 * C(1 | \pi_2)} \Leftrightarrow$$

$$p_2^* = \frac{p_2 * C(1 | \pi_2)}{p_2 * C(1 | \pi_2) + p_1 * C(2 | \pi_1)} \quad p_1^* = \frac{p_1 * C(2 | \pi_1)}{p_2 * C(1 | \pi_2) + p_1 * C(2 | \pi_1)}$$

Exempel: Antag att man kunde kontrollera om ett givet område innehåller ett studenthem. Antag också att man definierat S-områden som ett område där minst 10% av invånarna mellan 19 och 34 bor på studenthem, samt att man tillåter sig mindre fel. I stället för att kontrollera alla områden skulle man då kunna göra en analys med signifikansnivå 0,10 för variablerna och ange priorifördelningar 0,02 för vanliga och 0,98 för S-områden. Med samma indata klassificeras 278 av områdena som S-områden och när man kontrollerar dem finner man 25 av 27 analyserade S-områden. Det vill säga; genom att kontrollera 44 % av områdena finner man 93 % av S-områdena.

I detta exempel har man alltså antagit att kostnaden för felklassificering av S-område är 1105 gånger högre än kostnaden för felklassificering av ett vanligt område!

## 6.4 Alternativa sätt att välja variabler

Diskriminantfunktionen är den funktion som klassificerar godtyckliga områden som antingen S-områden eller vanliga. De variabler som man utför analysen med, är de variabler, med vars värden diskriminantfunktionen skapas. Om man utför analysen med alla variabler, men bara på ett antal (slumpmässigt utvalda) områden upprepade gånger, fås flera olika diskriminantfunktioner. Om man jämför dessa diskriminantfunktioner bör man kunna se att vissa variabler ger relativt likartade konstanter i diskriminantfunktionen, medan andra växlar mellan att ge stora och små bidrag, positiva och negativa. En variabel som för vissa områden har stor betydelse, men för andra liten, eller som till ömsom ger ett negativt bidrag till det slutgiltiga funktionsvärde som avgör klassificeringen, och ömsom ett positivt är inte mycket värt. Det tyder ju på att den variabeln inte alls är karaktäristisk, vare sig för S-områdena eller de vanliga. Eftersom vi har 25 variabler har vi en diskriminantfunktion med  $25 * 25$  konstanter och detta är väldigt omständligt. Vi gör därför en linjär diskriminant analys och har 25 konstanter att undersöka.

Vi utför fem analyser och ser vilka variabler som ger mest stabila konstanter i diskriminantfunktionen. Analysen utförs på hälften av områdena, dessa väljs ut slumpmässigt för varje gång. Därefter sorteras de konstanter som i vissa funktioner är positiva och vissa negativa bort, och även de där det största värdet på konstanten är mer än tre gånger större än det minsta. Kvar blir konstanter för följande variabler ; ålderns lägre kvartil, mödrarnas lägre ålderskvartil och tiopercentil, inflyttningsålderns median, utflyttningsålderns tiopercentil och totalbefolkningen. Gör man en icke-linjär korsvalideringsanalys på dessa variabler får man följande tabell:

	klassificerad som 0	klassificerad som 1
tillhörande 0	572	20
tillhörande 1	30	12

Resultatet är sämre än att gissa att alla områden är 'vanliga'.

En annan tanke är att undersöka de variabler som ger störst faktiskt effekt på diskriminantfunktionsvärdet. Hur stora eller små konstanterna är syns i dessa fem diskriminantfunktioner, men för att veta deras effekt på funktionsvärdet måste man veta storleken på de variabler de tillhör. (Om exempelvis konstanten för utflyttningsandelen är dubbelt så stor som konstanten för totalbefolkningen så kommer ändå konstanten för totalbefolkningen ge större effekt i och med att totalbefolkningen kommer vara minst 80 gånger större än utflyttningsandelen, som ju som störst kan vara 1). Genom att helt enkelt ta den genomsnittliga diskriminantfunktionen och genomsnittliga basområdet med mer än 80 invånare i åldern 19-34 år så får vi reda på den genomsnittliga effekten av konstanterna på diskriminantfunktionsvärdet. Även här görs en linjär analys av beräkningsskäl. De konstanter som ger störst absolut effekt är de för utflyttningsålderns median, nedre kvartil och tiopercentil, inflyttningsålderns median och övre kvartil och födelseålderns nedre kvartil. En icke-linjär korsvalideringsanalys på dessa variabler fås följande resultat:

	klassificerad som 0	klassificerad som 1
tillhörande 0	584	8
tillhörande 1	27	15

Detta är ett mycket bättre resultat de vi får med stegvis val av variabler på alla signifikansnivåer, och bättre än att klassificera alla som S-områden. För att kontrollera att det förhållandevis goda resultatet inte var en slump görs hela förfarandet om igen. Siffrorna ändrar sig lite, men det totala felet är detsamma; 35 områden felklassificeras.

Undersökningen görs om, men denna gång tas först 10 diskriminantfunktioner ut ur analyser på hälften av områdena för att räkna ut den genomsnittliga diskriminantfunktionen. (De variabler som väljs ut är ålderns övre kvartil, inflyttningsålderns median och övre kvartil, utflyttningsålderns median och 90percentil och mödrarnas nedre kvartil.) Resultatet blir dock inte bättre, tvärtom:

	klassificerad som 0	klassificerad som 1
tillhörande 0	584	8
tillhörande 1	29	13

Det undersöks om resultatet kan förbättras med en högre lägsta nivå för andelen studenter hos S-områdena (9,5 %).

	klassificerad som 0	klassificerad som 1
tillhörande 0	601	5
tillhörande 1	19	9

Det undersöks om resultatet kan förbättras med en högre gräns för minsta antalet invånare i åldern 19-34 år (minst 200):



	klassificerad som 0	klassificerad som 1
tillhörande 0	354	9
tillhörande 1	23	10

Slutligen undersöks om kombinationen minst 200 invånare, och minsta andel studenter för S-område 9,5 %. Men genom de nya gränserna har över hälften av S-områdena omdefinierats eller tagits ur analysen:

	klassificerad som 0	klassificerad som 1
tillhörande 0	354	6
tillhörande 1	9	10

## 6.5 Slutsats

Vi har alltså inte lyckats finna någon metod som ger ett särskilt bra resultat. I många fall skulle en klassifering av alla områden till vanliga ge ett bättre resultat, vilket verkligen inte är så konstigt med tanke på hur många fler de vanliga områdena är. Att se vilka variabler som i linjära analyser av slumpmässigt utvalda områden, som med sina respektive konstanter gav det största bidraget till funktionsvärdet, och använda dessa i en icke-linjär analys, gav till att börja med ett relativt bra resultat. Men när nya kriterier för vilka områden som ska kallas S-områden angavs, och nya gränser för befolkningsstorleken, var denna methods fördel framför stegvis val av variabler marginell.

## Kapitel 7

# Diskriminantanalys på Äldreområden

Av de 601 områden som har tillräckligt stor befolkning i åldersgruppen 72-100 år för att analyseras är 128, det vill säga 21,3 %, Å-områden. Den större andelen avvikande områden ger hopp om ett bättre resultat, vilket infrias redan i den första analysomgången med alla områden och alla variabler. 58 S-områden och 37 vanliga områden felklassificerades, summan av dessa är 95, vilket är mindre än de 128 områden som skulle ha felklassificerats om vi antagit att alla vore vanliga. Kovarianserna är signifikant olika, vilket föranleder en icke-linjär analys. Kovarianserna är signifikant olika, vilket föranleder en icke-linjär analys. Områdena klassificeras till två typer, typ 1: Å-område, typ 0: vanligt område. Först undersöks vilka variabler som är mest förklarande enligt stegvis val av variabler.

Signifikansgräns 0,01 (Inflyttningsandel, dödsålderns övre kvartil, totala antalet invånare över 72 år och antalet inflyttare per utflyttad)

	klassificerad som 0	klassificerad som 1
tillhörande 0	449	24
tillhörande 1	77	51

Signifikansgräns 0,15 0,1 och även 0,05 (Inflyttningsandel, dödsålderns övre kvartil, totala antalet invånare över 72 år, antalet inflyttare per utflyttad och dödsandelen)

	klassificerad som 0	klassificerad som 1
tillhörande 0	448	25
tillhörande 1	79	49

### 7.1 Gränser för analysen

#### 7.1.1 Gräns för andelen ålderdomshemboende

Kanske kan ett bättre resultat uppnås genom att ändra gränsen för hur stor del av befolkningen som ska bo på ett ålderdomshem för att det ska räknas som Å-område? Olika definitioner på ålderdomshem undersöks.

### signifikansnivå 0,01

Definition för typ 1	Å-omr.	ÅR	ÅF	VR	VF	fel
>0 % Å-hemsboende	128	51	77	449	24	<b>101</b>
>5 % Å-hemsboende	119	45	74	461	21	<b>95</b>
>10 % Å-hemsboende	110	40	70	476	15	<b>85</b>
>15 % Å-hemsboende	101	39	62	485	15	<b>77</b>
>20 % Å-hemsboende	90	35	55	498	13	<b>68</b>

Definition för typ 1: Andelen som måste bo på ålderdomshem för att området ska räknas som typ 1.

Å-omr.:	Antalet Å-områden.
ÅR:	Rättklassificerad Ålderdomshem
ÅF:	Felklassificerade Ålderdomshem
VR:	Rättklassificerade vanliga områden
VF:	Felklassificerade vanliga områden
Fel:	Antalet felklassificerade områden

### 7.1.2 Gräns för befolkningens storlek

Ett litet underlag ökar risken för extrema värden. Kanske man måste ignorera de mindre områdena för att få ett bättre resultat i klassificeringen av de större? Några olika gränser för befolkningstorleken (över 72 år) undersöks, med signifikansgränsen 0,01 för variablerna och minst 10 % av invånarna över 72 år boende på ålderdomshem.

Minsta ant. inv. över 72 år	omr	Å-omr.	ÅR	ÅF	VR	VF	fel
20	601	110	39	71	475	16	<b>87</b>
40	530	107	60	47	404	19	<b>66</b>
60	483	106	62	44	354	23	<b>65</b>
80	433	102	58	44	315	16	<b>60</b>
100	395	97	61	36	281	17	<b>86</b>

Antalet felklassificerade områden jämförs relativt antalet Å-områden, som ju skulle vara antalet felklassificerade områden om vi helt enkelt gissade att 'alla områden är vanliga'. Analysen är hela tiden bättre än gissningen, men som bäst är den vid minst 80 invånare över 72 år.

## 7.2 Välja variabler efter diskriminantfunktionen

På samma sätt som i stycke 6.4 utförs icke-linjär diskriminantanalys med korsvalidering på de variabler som vid fem linjära diskriminantanalyser på hälften av områdena, slumpmässigt utvalda, visade sig ha de mest stabila konstanterna. Variablerna var ålderns nedre kvartil, totalbefolkningen över 72 år, kvinnornas övre ålderskvartil, männens nedre ålderskvartil och inflyttningsandelen. Resultatet följer:

	klassificerad som 0	klassificerad som 1
tillhörande 0	441	32
tillhörande 1	71	57

Detta är ett knappt bättre resultat än där variablerna väljs ut med stegvis kontroll av förklaringsgrad och signifikansnivå 0,1. När vi även undersöker vilka av konstanterna som har störst absolut effekt på diskriminantfunktionen faller följande variabler ut för analys: Männens nedre ålderskvartil och median, kvinnornas övre och nedre ålderskvartil, ålderns övre och nedre kvartil. Resultatet blev som följer:

	klassificerad som 0	klassificerad som 1
tillhörande 0	442	31
tillhörande 1	87	41

Att göra analys på de mest stabila konstanterna gav ett hyfsat resultat, och det är möjligt att vi skulle kunna skärpa det ytterligare genom att basera det på ett större antal diskriminantfunktioner. Men eftersom metoden är så omständlig och resultatet bara medelmåttigt lämnar vi det därhän.

### 7.3 Slutsats

Jämfört med S-områdena ger diskriminantanalys på Å-områdena mycket bättre resultat. Exempelvis är majoriteten av de områden som klassificeras som Å-områden verkligen Å-områden, till skillnad från analysen av S-områdena, där endast de bästa analyserna hade en majoritet av verkliga S-områden bland de områden som klassificerades som sådana. Vi har även konstaterat att diskriminantanalys på Å-områdena med de variabler vi använt, lyckas klassificera fler områden rätt, än utgångsläget där alla områden ställs under samma antagande. Beroende på hur stor del av områdena man tycker sig behöva hitta, och beroende på hur stort fel man kan acceptera kan diskriminantanalys vara användbart, framför allt med variabler som berör inflyttning, ålder och död.

Om man vill undvika att felklassificera vanliga områden som Å-områden, kan en analysmetod som den i stycke 6.3 vara användbar, eftersom den kan kalibreras, så att minimum av vanliga områden felklassificeras. Den större del av Å-områden som då felklassificeras kan försvaras genom argumentet att de skulle ha 'felklassificerats' även om man inte använt diskriminantanalys, då alla områden skulle behandlats på samma sätt i prognosen.

### 7.4 Ålderdomshemmen i framtiden

Resultatet av denna rapport ska som sagt användas vid befolkningsprognoser. Men även om ett ålderdomshem kan identifieras i ett basområde, så betyder inte det att det kommer stå där för alltid. Hur vi ska ta hand om våra gamla är ett politiskt beslut, och faktum är att under perioden 2001 har antalet platser i äldreboende minskat med 13800 platser i hela Sverige, en minskning med 11 %! I stället vårdas allt fler i sina hem med hjälp av hemtjänst.

Enligt socialstyrelsen har totalt sett (i hela Sverige) 15,5 % av dem över 65 år hemtjänst eller plats på äldreboende, men hemtjänstens andel av dessa har ökat från 7,9 procentenheter till 8,5. Av dem som är över 80 år har 37 % hemtjänst eller plats på äldreboende.

## Kapitel 8

# Slutsatser

Resultatet av analyserna är för osäkra för att det ska vara meningsfullt att här presentera en formel som avgör om ett område innehåller ett studentområde, ett ålderdomshem eller ingetdera. Det skulle kunna ses som ett misslyckande, men trots allt har ju det visat sig att det även med höga felmarginaler då det gäller antal ålderdomshem och studentområden, antalet platser de innehåller och placeringen av dem, med all tänkbar statistisk signifikans råder en skillnad mellan områdena med avseende på ett antal variabler. Om man skulle ha tillgång till ett adressregister, där ålderdomshem och studentbostäder med säkerhet kunde placeras i rätt basområde, anser jag att det skulle vara meningsfullt att använda metoden, åtminstone för att finna ålderdomshem, eller ännu hellre områden där cirka 10% av invånarna i åldersgruppen över 70 eller över 75 år bodde på ålderdomshem, bland områden där minst några tiotal invånare var över 72 år. Skulle man dessutom kunna få reda på mer exakt hur många som var skrivna på de intressanta adresserna, skulle säkerheten i skattningarna öka än mer.

Variablerna i den första undersökningen av de rena Å-områdena och även diskriminantanalysen indikerar att de variabler som är mest intressanta för att sälla ut Å-områden är de som har med flytt- och åldersstrukturer att göra. Kön- och döds- relaterade parametrar har dykt upp sällan eller inte alls i analyserna.

Vad det gäller studentområden är det mer osäkert om diskriminantanalys har något att hämta. I undersökningen har de stora, centrala kommunerna använts, där man kan anta att det bor fler studenter, och därför finns efterfrågan på fler studentbostäder än exempelvis kommuner som Vallentuna eller Österåker. Trots detta är underlaget mycket litet. Frågan är om det ens är intressant att veta var studentområdena är, med tanke på att

- Studenter inte är en så homogen grupp som åldringar på ålderdomshem. En student återvänder till 'normalstadiet' efter bara några år, och många pendlar mellan att bo i studentlägenhet och som inneboende eller i andra hand. Att en människa på ålderdomshem skulle återvända till 'normalstadiet', boende i egen lägenhet, är mindre sannolikt. De äldre som trots allt bor tidsbegränsat på sjukhem och liknande för exempelvis rehabilitering skriver sig antagligen inte på sjukhemmet, eftersom de, till skillnad från de flesta studenter, har en annan, permanent bostad.
- Studenternas åldersgrupp är så mycket större än åldringarnas, till nume-

rären. Att det bor många mellan 20 och 35 i ett område är ett mer normalt tillstånd än att det bor många äldre än 70 år i ett område. Felskattningar inför framtiden är därför inte lika fatala.

- Nu för tiden finns det många tillfälliga studentområden, som barackerna på Kista IP och i Hammarbyhamnen (de senare är redan nermonterade). Att göra analys på dessa områden kräver precision i tiden, eftersom bostäderna rätt som det är kan monteras ned. Prognoser inför framtiden slår lätt fel.

Skulle det ändå anses befogat att särbehandla basområden med studentområden, skulle jag föreslå en manuell sortering, precis sådan jag redan gjort, eftersom det blir så stora fel med de modeller jag kommit fram till. Detta skulle innebära ett stort arbete vid ett tillfälle, men eftersom nybyggnationen inte är så stor, utan det handlar om förändringar på kanske fem områden om året, skulle arbetet med att uppdatera registren årligen därefter inte vara så betungande.

## Kapitel 9

# Appendix

### 9.1 Ordlista

Data	observation, information
Iteration	upprepning av exempelvis en räkneoperation
Outlier	En observation som avviker kraftigt och där avvikelsen antas inte bero på slumpen
Prediktion	statistiskt grundad förutsägelse
Servicehem	I denna text definierat som ett boende med åldersgräns 65 år eller högre där invånarna erbjuds någon form av service, typ lagad mat eller städhjälp.
Sjukhem	Boende för personer med stora medicinska behov. Har räknats till Ålderdomshem, trots att det ej förekommer åldersgräns.
Strukturell skillnad	Skillnad som inte bara beror på slumpen
Ålderdomshem	I denna text definierat som alla typer av boenden med åldersgräns 65 år eller högre för att få flytta in.
Äldreboende	Hem för gamla som inte längre kan ta hand om sig själva.

### 9.2 Litteraturlista

Befolkningsprognos 2004 (rapport) *Statisticcon*  
A Handbook of Statistical Analyses using SAS *Der Everitt* 2002  
Applied Multivariate Statistical Analysis *Johnson Wichern* 2002  
Dagens Nyheter 23/2 2006

### 9.3 Tack

Jag vill tacka Anders Björkström och Anna Carlsund för deras utomordentliga handledning.