

Slutrapport Översyn av Tyko

Version Nummer	Datum	Summarisk beskrivning av ändringar
1	2010-10-15	Utkast
2	2010-10-26	Slutlig

Innehåll

1	Sammanfattning	3
2	Projektet	3
2.1	Projektets bakgrund och uppgift.....	3
2.1.1	Bakgrund	3
2.1.2	Mål	4
2.2	Deltagare	4
3	Kunder och kundönskemål	5
4	Projektarbete och erfarenheter	5
4.1	Vad som var positivt	5
4.2	Vad som kan förbättras	5
5	Resultat	6
5.1	Det tillkomna beståndet och beståndet 1998.....	6
5.2	Möte med KSP:s prognosgrupp.....	6
5.3	Jämförande och utvärdering av grupperingar.....	7
5.4	Variabler.....	7
5.4.1	Fastighetsvariabler	7
5.4.2	Demografiska variabler	8
5.5	Datamaterialet.....	8
5.6	Klustringsmetodik och modell	8
5.6.1	Klusteranalys	9
5.6.2	Chi2-testning.....	9
5.7	Kvalitet i nuvarande gruppering	10
5.8	Nya Grupperingar	11
5.8.1	Flytande gränser.....	11
5.8.2	Uppdelade gränser.....	13
5.9	Förändring av nuvarande Tyko-indelning.....	14
5.10	Kvalitet.....	15
5.11	Klustringsprogram	15
5.12	Tidsplan och kostnad för implementering.....	15
6	Fortsättning	15
7	Resursinsats.....	16
8	Godkännande.....	16
	Bilaga 1. Förslag till utvärdering av ny Tyko-indelning	17
	Bilaga 2. Kommentarer kring fastighetsvariabler.....	19
	Bilaga 3. Beskrivning av klustringsprogram	21

1 Sammanfattning

Den nuvarande TYKO-klassificeringen som används gjordes 1998 av Svenska kommunförbundet och Föreningen för kommunal statistik och planering (KSP). SCB har därefter fungerat som producent av årligen uppdaterat TYKO-material. Eftersom klassificeringen är mer än tio år gammal har det varit aktuellt med en översyn och en nygruppering av TYKO.

Projektet föreslår att man bibehåller nuvarande kod med diverse justeringar i gränser mellan och sammanslagning av vissa grupper.

Med de program, metodik och arbetsdokument projektet producerat kan man kontinuerligt kontrollera de TYKO-grupper som används, exempelvis inför varje produktionsår.

2 Projektet

2.1 Projektets bakgrund och uppgift

2.1.1 Bakgrund

För små regionala områden är det svårt att göra befolkningsprognoser på grund av att de demografiska måtten är baserade på få observationer av flyttningar, födda och döda. Som lösning på detta kan man ta fram mått som baseras på riksdatabaser för olika typområden av bostäder, vilka kan användas oavsett prognosområdets storlek. Framtaganden av sådana typområden har gjorts i olika former under de senaste årtiondena.

Fastighetsklassificeringen TYKO används till detta ändamål och utgör grunden för statistikpaketet TYKOPAK. Paketet erbjuds årligen till kommuner som hjälp i deras arbete med delområdesprognoser och har omkring 40 beställare per år. Klassificeringen omfattar i nuläget 33 TYKO-grupperingar.

Den nuvarande klassificeringen som används gjordes 1998 av Svenska kommunförbundet, Bygghörsningsrådet och Föreningen för kommunal statistik och planering (KSP). SCB har därefter fungerat som producent av årligen uppdaterat TYKO-material.

Eftersom klassificeringen är mer än tio år gammal har det varit aktuellt med en översyn och en nygruppering av TYKO. Behovet har även påtalats av externa aktörer som KSP. Frågan man ställt sig är hur det tillkomna beståndet skiljer sig från det bestånd som låg till grund för analysen 1998. Skillnader i ålderstrukturen för de boende i det äldre beståndet har ändrats

och då är det inte säkert att den nuvarande grupperingen är optimal i demografiskt hänseende.

Det datorprogram som användes vid den föregående klustringen finns inte längre tillgänglig och dokumentationen hur den gick till är inte helt fullständig. Därför har det funnits ett behov av att göra en ny dokumenterad klustring med en metod och ett program som kan användas för att göra upprepande grupperingar.

2.1.2 Mål

Projektets mål har varit att göra en översyn av det gamla TYKO och ta fram en ny TYKO-indelning. Den nya TYKO-grupperingen ska kunna användas i prognossammanhang inom uppdragsverksamheten och implementerbart i nuvarande produktionssystem för TYKOPAK.

Projektet ska

- Göra en genomgång och bedömning av kvalitet i nuvarande variabler och klassificering.
- Föreslå eventuella nya variabler som ska vara med i klustringsmodellen för en ny klassificering.
- Jämföra minst två olika modeller och motivera till ställningstagande av vald modell.
- Leverera en ny TYKO-klassificering som är jämförd med nuvarande klassificering.
- Framställa ett dokumenterat klassificeringsprogram.
- Göra en kostnadsuppskattning och tidsplan för implementering av ny TYKO-gruppering.
- Ta fram ett utvärderingsmått för att kunna göra en kvalitetsbedömning av nuvarande TYKO-gruppering.

2.2 Deltagare

Michael Franzén, PCA/MIH, projektledare och –deltagare; f d produktansvarig för TYKOPAK.

Helen Marklund, PCA/LEDN, projektdeltagare; f d produktansvarig för TYKOPAK

Christian Skarman, BV/PI, projektdeltagare; demograf och sammankallande för KSP:s prognosgrupp. Har tidigare arbetat med regionala befolkningsprognoser vid Region Skåne.

Kajsa Swensson, BV/BE, projektbeställare; enhetschef på BV/BE

Anders Karlsson, BV/BE, projektdeltagare; produktansvarig för TYKOPAK

Alexandra Håkansson, BV/BE, projektdeltagare; arbetar med regionala befolkningsprognoser

Under projektets gång har Martin Verhage och Magnus Walestad på RM/BF tillfrågats angående fastighetsvariabler.

Martin Ribe, PCA/LEDN, rådfrågades om klustringsmetodiken.

KSP:s prognosgrupp har fungerat som extern expert- och användargrupp vad det gäller användandet av TYKO i regionala befolkningsprognoser.

3 Kunder och kundönskemål

BV/BE har beställt genomgången till en eventuell implementering av en ny TYKO-gruppering. Kostnadsuppskattningen och tidsplanen för implementering finns i avsnitt 5.11

Förslag till TYKO-grupperingar och dokumenterat klassificeringsprogram har levererats till BV/BE.

4 Projektarbete och erfarenheter

Projektet har haft sex arbetsmöten och ett par möten med projektbeställaren. Den 13 april träffade projektgruppen KSP:s prognosgrupp.

4.1 Vad som var positivt

Vi har nu en större uppfattning om arbetet med att ta fram en sådan här gruppering innebär. Vi har nu även program som helt eller delvis kan användas för att göra grupperingar i framtiden. Vi har också större möjligheter att utvärdera den nuvarande TYKO-koden mer regelbundet.

4.2 Vad som kan förbättras

SAS-programmeringen av klustringsprogrammet (Chi2-testningen) tog mycket mer tid än befarat. Grupperingen var komplex och tog även den längre tid än befarat. Detta har lett till att tids- och resursplanen har ändrats under projektets gång.

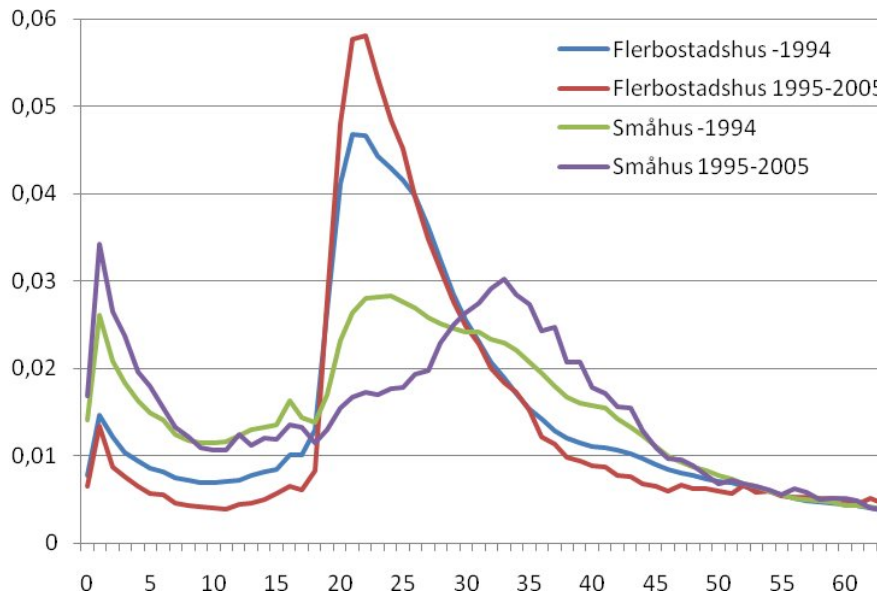
5 Resultat

Projektets resultat presenteras nedan punktvis.

5.1 Det tillkomna beståndet och beståndet 1998

Vid den senaste grupperingen användes data innehållande bostadsbeståndet t o m 1994. I diagram 1 jämförs inflyttningsfördelningen för detta bestånd med fastigheter tillkomna 1995-2005, uppdelat på flerbostadshus och småhus.

Diagram 1. Inflyttningsfördelning efter ålder för flerbostadshus och småhus byggda t o m 1994 och 1995-2005, andelar av den totala inflyttningen



I diagrammet kan man se att åldersfördelningen för de som flyttar in i det nyare beståndet skiljer sig åt från åldersfördelningen för de som flyttar till det äldre beståndet. Byggnationen efter 1995 verkar ha en äldre befolkning än den äldre byggnationen. Till flerbostadshusen verkar det också flytta en något större andel familjehushåll. Det finns alltså tecken på att gruppen fastigheter byggda 1995-2005 skulle kunna utgöra en egen grupp i TYKO-indelningen, vilket motiverar en nygruppering.

5.2 Möte med KSP:s prognosgrupp

Projektgruppen träffade KSP:s prognosgrupp på deras möte den 13 april. Det som togs upp till diskussion var utvärderingen/valideringen av den nya grupperingen kontra den gamla. Det föreslogs att man kan testköra prognoser

på områden på Nyko-nivå 5 med unika TYKO-koder och jämföra resultat mellan olika grupperingar. Områdena behöver vara av en viss storlek men KSP ville även ha mindre områden, då det är för små områden som TYKO är mest användbart. Då kan man dela in områdena i grupper, t ex över 1 000 och under 1 000 personer.

5.3 Jämförande och utvärdering av grupperingar

Utvärderingen av modellerna har två olika ansatser. Den första är att utvärdera hur väl fördelningarna för de utvalda grupperna är uppdelade i signifikant skilda enheter. Detta kan göras med att jämföra med de ingående gruppernas Chi2-värden från tester med närliggande grupper.

Den andra ansatsen är hur väl grupperna ger bra prognosresultat. Detta kan göras genom att jämföra prognoser gjorda av olika grupperingar ur olika aspekter. Detta beskrivs utförligt i bilaga 1. Denna ansats har inte använts då det ansetts vara ett för omfattande arbete som inte ryms inom projektet.

5.4 Variabler

Variablerna är hämtade från Fastighetstaxeringsregistret och Registret över totalbefolkningen.

5.4.1 Fastighetsvariabler

Nedan följande en kort beskrivning på de fastighetsvariabler ansågs vara intressanta för klusteranalysen:

Jurform – Beskriver juridisk form med kod.

Jurformgrp – Gruppering av jurform.

Typkod – Anger bostadstyp såsom småhus och flerbostadshus.

Typbebygg – Beskriver typ av bebyggelse såsom hyreshus, kedjehus, radhus eller friliggande byggnad.

Tätortsstorlek – Tätortens storlek där fastigheten ligger.

Värde - Taxeringsenhetens totala taxeringsvärde.

Nivåfaktor – Beskriver områdets attraktivitet där fastigheten ligger.

ArNyVarde – Nybyggnadsår/år för senaste ombyggnation eller renovering.

Bostadsyta – Byggnadens bostadsarea. Finns indelad i total bostadsyta för småhus och total bostadsyta för hyreshus.

Antal rum – Variabel från lägenhetsregistret. Anger antal rum i lägenheten. Denna variabel är ännu inte användbar.

Jufrom + typkod – Likt i nuvarande TYKO-indelning slogs juridisk form och bostadstyp ihop för att bilda typerna bostadsrätt i flerbostadshus resp småhus, hyresrätt i flerbostadshus resp. småhus och äganderätt i småhus.

Vissa fastigheter har flera byggnader. Eftersom matchning mot RTB måste väljs ett värde för dessa variabler som får representera fastigheten. För bostadsyta väljs den byggnad med störst bostadsyta ut. För byggåret (arnyvarde) används det högsta värdet.

Det bestämdes att projektet skulle koncentrera sig på att använda de variabler som ingår i nuvarande gruppering (jurform + typkod, arnyvarde, tätortstorlek och nivåfaktor). Jämförelsen mellan den nuvarande grupperingen och de nya blir därmed enklare. Därtill har övriga föreslagna variabler inte varit tillräckligt intressanta eller beskriver samma saker som de redan ingående variablerna. En tredje anledning är att en eventuell implementering av en ny gruppering blir mindre kostsam och omfattande. Det räcker i princip med att ändra gränserna nominellt i nuvarande SQL-program som används i produktionen av TYKOPAK.

Kvaliteten i variablerna har bedömts med hjälp av Magnus Walestad och Martin Verhage på RM/BF. Mer ingående analys och kommentarer finns i bilaga 2.

5.4.2 Demografiska variabler

De demografiska variablerna som använts i analysen är *inflyttningsfördelning* och *utflyttningsrisker*. Användandet av åldersspecifik fruktsamhet diskuterades men lades åt sidan då man inte använt denna variabel i föregående analys och att tiden inte räckte till. Utflyttningsriskerna beräknades med åldersspecifika *medelfolkmängder*. Beskrivning av dessa mått finns i *Beskrivning av statistiken: BE0101 Befolkningsstatistik*

Kvaliteten i dessa mått/variabler anses vara god då de används frekvent inom befolkningsstatistiken och ingår i RTB.

5.5 Datamaterialet

Det datamaterial som ligger till grund för analyserna är i stora drag de data som används i TYKOPAK 2008. På så sätt är det lättare att jämföra denna gruppering som använts i produktion med de nya grupperingarna. Datamaterialet bortser från relativt nybyggda fastigheter för att få stabilitet i undersökningsdata. Detta innebär att enbart fastigheter som varit bebodda de senaste fyra åren finns med i analysen.

5.6 Klustringsmetodik och modell

Data för de demografiska variablerna matchades med fastighetsdata och aggregerades för varje värde på de fyra utvalda fastighetsvariablerna

(jurform/typkod, arnyvarde, tätortstorlek och nivåfaktor) för att skapa fördelningar som kan användas i testningen.

5.6.1 Klusteranalys

En inledande studie med klusteranalys på tillräckligt stora grupper efter en förgruppering med Chi2-tester gav inga kluster som enhetligt kunde analyseras. Klustren innehöll grupper med vitt skilda värden för fastighetsvariablerna. Därför övergavs denna metodik till förmån för Chi2-testning och manuell gruppering via studier av diagram. Omfattande manuell klustring utfördes även vid framtagandet av den TYKO-klassificering som används idag.

5.6.2 Chi2-testning

Chi2-testningen testar huruvida två närliggande grupper är lämpliga att slås ihop eller kan anses utgöra två enskilda grupper. Med närliggande grupper avses grupper med fastighetsvärden som ligger nära varandra i det fyrdimensionella rummet. Hypotestestet innebär att man testar:

H0: Det finns ingen skillnad mellan grupperna. Dessa tillhör samma fördelning och bör slås ihop.

H1: Det finns skillnad mellan grupperna. Dessa tillhör olika fördelningar och bör inte slås ihop.

H0 förkastas om testet visar höga F-värden och p-värdet är lägre än signifikansnivån.

Testningen har gjorts med proceduren *proc freq* i SAS.

Inledande test gav ett stort antal grupperingar varpå signifikansnivån sattes till 1 ppm samt att gruppen inte fick understiga 10 000 observationer. Därtill utfördes testningen på åldersgrupperna 1-7 och 16-44 år vilka påverkar flyttningsmönstret mest. Därefter utfördes en *manuell gruppering* av signifikanta grupper via visuell bedömning av diagram över grupperna. Om två närliggande grupper anses ha samma fördelning slås dessa ihop. Därefter görs en ny Chi2-testning av de kvarvarande grupperna för att fastslå signifikanta skillnader.

Testningen har i första hand utförts på inflyttningsfördelningarna. Därefter har utflyttningsriskerna studerats för de utvalda grupperna och man har då funnit att de varit godtagbara även för utflyttningsriskerna.

5.7 Kvalitet i nuvarande gruppering

Kvaliteten i nuvarande gruppering har bedömts efter hur väl den avgränsar de ingående grupperna (33st) med hjälp av de kriterier (Chi2-värden, antal observationer) som använts i testningen av nya grupper.

Tabell 1. Nuvarande gruppering. Test för inflyttningsfördelningen

Tyko	p-värde under H0	Total inflyttning
FBR128	0,376	18 221
FBR138	0,000	26 402
FBR148	0,000	40 456
FBR228	0,000	37 400
FBR238	0,000	33 648
FBR248	0,000	61 249
FBR328	0,000	127 707
FBR329	0,000	13 308
FBR338	0,000	34 216
FBR339	0,000	691
FBR348	0,000	99 612
FBR349	0,000	1 656
FHR128	0,000	125 684
FHR138	0,000	84 079
FHR148	0,000	259 408
FHR228	0,000	171 811
FHR238	0,000	95 774
FHR248	0,000	299 372
FHR328	0,000	210 324
FHR329	0,000	8 737
FHR338	0,000	112 862
FHR339	0,000	1 508
FHR348	0,000	298 585
FHR349	0,000	2 738
SBR000	0,000	67 172
SER801	0,000	153 968
SER802	0,015	44 408
SER803	0,014	2 268
SER804	0,000	232
SER901	0,000	342 721
SER902	0,000	253 214
SER903	0,000	44 163
SER904	0,000	8 072

Kvaliteten anses vara relativt god med undantag från ett par grupperingar som inte visade sig vara statistiskt signifikanta eller tillräckligt stora.

5.8 Nya Grupperingar

Nedan beskrivs två nya grupperingsförslag som är framtagna enligt den metod som beskrivits i avsnitt 5.5.2

Den indelning av *jurform* + *typkod* som används i analysen är framtagen tillsammans med RM/BF för småhusen (och flerbostadshusen) och skiljer sig från den som används i nuvarande gruppering. Se avsnitt 5.9.

I dessa grupperingar har värdena de fem kombinerade värdena (Hyresrätt i flerbostadshus, Hyresrätt i småhus, Bostadsrätt i flerbostadshus, bostadsrätt i småhus, Äganderätt i småhus) för *jurform* och *typkod* analyserats var för sig. Dessa ses alltså som huvudvariabel i indelningen.

5.8.1 Flytande gränser

Inför testning med denna gruppering är data sorterat efter bostadstyp, tätortsstorlek, byggår och nivåfaktor. Grupperingen tar inte hänsyn till gränser för de enskilda variablerna. Det innebär i vissa fall att fastigheter från 2005 kan slås ihop med fastigheter byggda 1900.

Tabell 2. Gruppering med flytande gränser. Test för inflyttningsfördelningen

Gruppering	Total inflyttning	Gruppering	Total inflyttning
FBR1_1970_65	60 417	FHR3_1929_65	15 994
FBR1_2005_6	32 005	FHR3_1967_675	425 989
FBR2_1965_575	77 690	FHR3_1968_675	31 231
FBR3_1900_875	65 225	FHR3_1969_46	11 240
FBR3_1931_875	65 354	FHR3_1969_725	20 173
FBR3_1945_825	56 447	FHR3_1970_48	17 781
FBR3_1949_8	15 004	FHR3_1970_75	14 139
FBR3_1954_75	14 623	FHR3_1971_5	11 006
FBR3_1965_875	43 902	FHR3_1971_675	16 478
FBR3_1968_875	14 356	FHR3_1971_875	10 943
FBR3_1972_55	15 975	FHR3_1972_55	12 628
FBR3_1991_9	43 091	FHR3_1972_75	10 758
> FBR3_1991_9	37 204	FHR3_1974_875	39 872
FHR1_1929_38	29 942	FHR3_1981_625	28 604
FHR1_1965_36	221 099	FHR3_1987_75	30 488
FHR1_1979_35	174 273	FHR3_1996_775	40 928
FHR1_1991_575	89 756	FHR3_2004_8	15 366
FHR2_1900_65	39 778	> FHR3_2004_8	9 206
FHR2_1965_48	294 173	SBR	71 927
FHR2_1966_875	35 886	SER1_1909_9	231 166
FHR2_1967_65	22 646	SER1_1999_95	335 102
FHR2_1968_5	12 364	SER2_1909_95	15 733
FHR2_1971_575	72 864	SER2_1967_95	62 713
FHR2_1972_38	12 151	SER3_2007_16	192 927
FHR2_1972_65	12 141	SHR3_1909_11	67 104
FHR2_1973_6	16 982	> SHR3_1909_11	6 988
FHR2_1992_525	109 746		
FHR2_1997_575	25 642		
FHR2_2003_575	15 854		

Samtliga grupper har p-värden nära 0.

Nackdelen med denna gruppering är att den är svårförstådd och inte så pedagogisk då den innehåller grupper med t ex vitt skilda byggnadsår. Däremot ger de en statistiskt sätt starkt åtskilda grupper. Grupperingen ger dock ett stort antal grupper, 55 st.

5.8.2 Uppdelade gränser

I denna gruppering har de grupper som föll ut i den flytande grupperingen används för att bedöma vilka gränser för de olika variablerna som verkar lämplig. Frekvent förekommande värden som åtskiljer grupper används som tänkbara markörer för grupper. Därefter plockas dessa fram varvid den nya indelningen undersöks.

Tabell 3. Gruppering med uppdelade gränser. Test för inflyttfördelningen.

Total		Lathund:		
Gruppering	inflyttning			
FBR111	124 788	FBR		
FBR112	4 678	tätort	år	nivåfaktor
FBR121	103 599	1- 2	-1965	-8,7
FBR122	2 084	3	1966-	>8.7
FBR211	101 709	FHR		
FBR212	94 686	tätort	år	nivåfaktor
FBR221	78 766	1	-1965	-6,5
FBR222	30 983	2	>1965	>6.5
FHR111	245 666	3		
FHR112	11 669	SBR		
FHR121	272 928			
FHR122	23 877	SER		
FHR211	205 356	tätort	år	nivåfaktor
FHR212	108 709	1	-	-9,5
FHR221	245 023	2		9,5
FHR222	83 095	3		
FHR311	62 792	SHR		
FHR312	340 736	tätort	år	nivåfaktor
FHR321	170 501	1-2	-	-
FHR322	177 769	3		
SBR000	71 927			
SER101	521 356			
SER102	56 499			
SER201	103 732			
SER202	35 187			
SER301	30 422			
SER302	90 445			
SHR100	66 947			
SHR200	7 145			

Fördelen med denna gruppering är att den är pedagogisk och innehåller inte flytande grupper. Däremot behöver den inte vara optimal då den teoretiskt sätt kan innehålla grupper som egentligen inte hör ihop (även om de

statistiskt sätt är starkt åtskilda). Samtliga grupper har p-värden nära 0. Denna tes stärks av att om man väljer något annorlunda värden än de frekvent förekommande får man ändå starka signifikanser.

Det gjordes även en gruppering på detta sätt men med samma princip som för nuvarande TYKO-indelning, vilken man kan kalla *sammanslagna gränser*. Det innebär att samma gränser användes för samtliga fem grupper i huvudvariabeln. Denna indelning visade sämre signifikanser och hade fler grupper sammanlagt och publiceras därför inte här.

5.9 Förändring av nuvarande Tyko-indelning

De ej signifikanta grupperna (i avsnitt 5.6) borde kunna slås ihop

En förändring av gränser för nuvarande gruppering kan ses som en ny indelning.

Dessa gränser används för att dela in Typkod och jurformgrp i nuvarande indelning

<i>Namn</i>	<i>Typkod</i>	<i>Jurformgrp</i>
<i>Småhus, bostadsrätt</i>	<i>100-299</i>	<i>7</i>
<i>Småhus, äganderätt</i>	<i>100-299</i>	<i><>7</i>
<i>Flerbostadshus, bostadsrätt</i>	<i>300-399</i>	<i>7</i>
<i>Flerbostadshus, hyresrätt</i>	<i>300-399</i>	<i><>7</i>

Följande gruppering föreslogs istället:

	<i>Typkod</i>	<i>Jurformgrp</i>
<i>Småhus, bostadsrätt</i>	<i>200-299</i>	<i>7</i>
<i>Småhus, äganderätt</i>	<i>200-299</i>	<i>(3), 4, 5, 9</i>
<i>Småhus, hyresrätter</i>	<i>200-299</i>	<i>1, 2, 6, 8</i>
<i>Flerbostadshus, bostadsrätt</i>	<i>300-399</i>	<i>7</i>
<i>Flerbostadshus, hyresrätt</i>	<i>300-399</i>	<i><>7</i>

Se även bilaga 2.

För småhusen bör de med typkod 100-199 undersökas närmare (lantbruksfastigheter) där det förmodligen finns hyresrätter bland äganderätterna. Tillsvidare kan man dock använda den nuvarande indelningen ovan för typkod 100-299.

I nuläget finns inga äganderätter för flerbostadshusen men befaras komma med i FTR10 med typkod >500. Dessa bör beaktas vid implementeringen.

5.10 Kvalitet

Kvaliteten i nuvarande indelning av juridisk form och typkod kan förbättras genom att använda de gränser om beskrivs i avsnitt 5.8.

5.11 Klustringsprogram

Det framtagna klustringsprogrammet beskrivs i bilaga 3.

5.12 Tidsplan och kostnad för implementering

Eftersom de grupperingar som är föreslagna innehåller samma variabler som den nuvarande grupperingen består implementeringsarbetet i att ändra gränser för att för att ingå i en specifik TYKO-kod. Inga övriga förändringar i produktionsskripten behöver göras. Arbetet med denna implementering tar uppskattningsvis 25-50 timmar beroende på mängden grupperingsändringar och administrativt arbete kring utvärdering och dokumentation (t. ex. på TYKOPAKs produktsida). Dessa justeringar kan utföras av personal inom arbetsgruppen för TYKOPAK. Implementering bör kunna ske innan produktionen av 2011 års tabeller.

Arbetet blir större om någon av de nya grupperingarna väljs, mindre om förändringen av den nuvarande koden utförs.

6 Fortsättning

Projektet föreslår att man bibehåller nuvarande kod med diverse justeringar i gränserna mellan grupperna och sammanslagning av ett par av dem. Även en mindre utredning av lantbruksfastigheterna samt hänsyn för det nya FTR10 bör tas.

Med de program, metodik och arbetsdokument projektet producerat kan man kontinuerligt kontrollera den gruppering som används. Detta är något man gjort för nivåfaktor inför varje produktion men nu kan man även kolla övriga variabler och göra justeringar i gränserna mellan TYKO-grupperna.

Projektet har inte haft tid att göra en Stegvis toppbotten-analys där man utgår från bostadstypen och dess uppdelningar för att sedan försöka dela upp på tätort, nivåfaktor och byggår. Här skulle man kunna gå på den gällande indelningen och inte använda nivåfaktor för flerbostadshus och inte byggår för småhusen. Vidare borde man undersöka om gruppen SHR (hyresrätter i småhus) ska ingå i grupperingen.

Vid införande av en ny indelning bör någon form av utvärdering enligt ansats två under punkt 5.2 göras av externa användare. Förslagsvis kan KSP:s prognosgrupp utvärdera modellen. Detta ansågs vara en bra idé på mötet den 13 april 2010. Gruppen bör även meddelas vid mindre justeringar av indelningen.

Ett tredje sätt att jämföra grupperingar som inte har utretts i projektet är att studera de enskilda fastigheternas variation i förhållande till gruppens medel där fastigheten ingår. Då får man en bild över hur stor variationen är inom grupperna och hur homogena de är. Klusteranalys utnyttjar objektens variation på detta sätt för att skapa så homogena grupper som möjligt. Det skulle vara intressant att titta närmare på denna typ av variation åtminstone för den indelning som kommer att användas i produktion.

Vidare arbete kan ingå i FMOD för TYKOPAK.

7 Resursinsats

Projektbudgeten har varit 350 arbetstimmar.

8 Godkännande

Avrapporteringen ska godkännas av Kajsa Swenson, BV/BE.

Bilaga 1. Förslag till utvärdering av ny Tyko-indelning

TYKO med observationer från riket är framtaget som ett underlag till kommunerna för att kunna göra mer tillförlitliga delområdesprognoser när det gäller in- och utflyttning, än de kan göra med observationer från delområden. De tänkta användarna är de kommuner som gör delområdesprognoser och de kommuner som har liten befolkning och gör kommunprognos.

För att försöka avgöra om den nya klustringen kommer att ge bättre kommun- och delområdesprognoser behöver vi göra en utvärdering. Den är svår att göra på delområden eftersom de är små områden med få observationer där antalet in- och utflyttare varierar stort från år till år. Den relativa variationen där är större än för områden med större befolkningar, både beroende på att de har mindre befolkning och att fler flyttar kortare än långt.

Utvärderingen är möjlig att göra på delområden men det behövs då ett mycket stort urval där det för varje delområde behövs göras en indelning i prognosområden som överensstämmer med TYKO-klasserna. Det är besvärligt att göra eftersom det hela tiden behöver göras överväganden om vilka områden som är tillräckligt TYKO-rena och samtidigt tillräckligt befolkningsstora för att utgöra ett prognosområde.

Ett förslag är att istället göra utvärderingen på små kommuner som delas in i prognosområden för att få arbetet med utvärderingen mera praktiskt hanterbart, eftersom det då blir lättare att bilda tillräckligt TYKO-rena och tillräckligt befolkningsstora prognosområden utan att göra alltför många avväganden. Lika många testobjekt behövs dessutom inte eftersom slumpfaktorn (hos antalet in- och utflyttare) inte är lika stor som den är i delområden.

Utvärdering på små kommuner och deras prognosområden

Använd alla kommuner med upp till 30 000 invånare som testunderlag.

Förfaringssätt:

- Dela upp kommunen i prognosområden m.h.a. TYKO-klasser, vilka inte behöver vara geografiskt sammanhängande. Eventuella för små TYKO-klasser slås ihop med en till karaktären närliggande klass.
- Skriv fram antalet in- och utflyttningar under 2009 (fördelade på kön och ettåriga åldersklasser) till och från kommunen och även till och från områdena m.h.a. Riks-TYKO 2006-2008 enligt gamla indelningen.
- Skriv fram antalet in- och utflyttningar under 2009 (fördelade på kön och ettåriga åldersklasser) till och från kommunen och även till och från områdena m.h.a. Riks-TYKO 2006-2008 enligt nya indelningen.
- Ta fram de sanna värdena för in- och utflyttningar under 2009 (fördelade på kön och ettåriga åldersklasser) till och från kommunen och även till och från områdena.
- De tre åldersfördelade uppsättningarna av in- och utflyttningar (prognos enligt den gamla indelningen, prognos enligt den nya indelningen och observerade värden) ritas in i ett och samma diagram. Dessa kurvor kan nu chi-två-testas parvis för att avgöra om de har mätvärden ur samma fördelning.

Bilda resultat kommunvis, och ge svar på frågorna;

- Den gamla och nya indelningen är; likvärdiga eller skiljer sig åt.
- Den gamla indelningen ger; ”rätt” resultat eller felaktigt resultat.
- Den nya indelningen ger; ”rätt” resultat eller felaktigt resultat.

Sammanställ resultaten för samtliga testkommuner, och ge svar på frågorna;

- Förbättrar TYKO-materialet resultaten i kommun- och delområdesprognoser?
- Finns det en klar bild av om den nya klustringen är bättre än den gamla. Isåfall hur mycket bättre?

Innan vi ser resultaten bör vi ha satt gränser för vad som är bättre. Det får inte bli en subjektiv efterkonstruktion.

Bilaga 2. Kommentarer kring fastighetsvariabler

Sammanfattning av mailkonversation med Martin Verhage och Magnus Walestad.

Fastighetsregistret (FASTPAK)

Fastpak bygger på Fastighetstaxeringsregistret (FTR) (som kommer från Skatteverket där objekten egentligen är taxeringsenheter och värderingsenheter och inte fastigheter) men även Lantmäteriets Fastighetsregister (som bygger på fastigheter och byggnader)

Bostadsyta

För Bostadsytan i Fastpak finns variablerna YtaBostS (total bostadsyta för småhus) respektive YtaBostH (total bostadsyta för hyreshus).

Byggnation

Byggnadsår beskrivs av flera variabler. Variabeln ArBygg är byggnadsår. Dess kvalitet är dock inte så bra och då kan man använda variabeln ArVarde (Värdeår) som är nybyggnadsår korrigerat med år för senaste renovering eller ombyggnad till modernare standard. Variabeln ArNyVarde är en kombination av Arbygg och ArVarde, vilken innehåller i första hand Byggår men om detta saknas innehåller den Värdeåret.

Variabeln ArNyVarde innehåller många observationer med värdet 1929 vilket beror på att det är det lägsta värdeåret som kan förekomma. Många har även byggåret 1909 men det finns ingen förklaring på detta.

Småhus

Småhus (Typkod 200-299) som är skilt från Jurformgrp 7 innehåller både äganderätt och hyresrätt. Majoriteten av dessa antas vara äganderätt. Vi har idag inget bra sätt att ta reda på hur många småhus som faktiskt upplåts med hyresrätt. Man skulle kunna ta reda på hur många småhus som ägs av allmännyttan, dessa är hyresrätter.

De taxeringsenheter som har JurFormGrp 8 är ett allmännyttigt bostadsföretag i fastighetstaxeringsregistret. Det är inte så lätt att skilja ut hyresrätter för småhus. Gjorde en snabb koll för typkod 200-299 och där kan man se att fysiska personer (JurFormGrp 4) äger flest antal (taxeringsenheter) småhus. Man skulle kunna anta att de som ägs av fysiska personer, dödsbon (JurFormGrp 5) och övriga (JurFormGrp 9) är *äganderätter*. För övriga måste man rensa bort utländska juridiska personer. De som ägs av bostadsrättsföreningar JurFormGrp 7 är *bostadsrätt*. *Hyresrätt* blir då för övriga koder staten (JurFormGrp 1), kommun (JurFormGrp 2), svenskt aktieföretag (JurFormGrp 6) och allmännyttigt bostadsföretag (JurFormGrp 8).

Jurformgrp	Antal Taxeringsenheter
NULL	30
0	11
1	1 586
2	23 236
3	1 053
4	2 188 919
5	20 981
6	32 842
7	6 127
8	4 360
9	48 169

Flerbostadshus

Det finns ett antal hyresrätter i bostadsrättsföreningar, men vi vet inte hur många. De flesta som har JurFormGrp 7 är bostadsrätter. Det finns inga ägarlägenheter för flerbostadshus i FTR 09 men kommer nog några i FTR 10. Dessa kommer att få en egen typkod som börjar på 500.

Bilaga 3. Beskrivning av klustringsprogram

Program att ta fram datamaterialet

Program 1 - *Ta fram datamaterial_1.sql*

1. Hämtar utvalda bostyta, byggper, nivafaktor och typbebygg från byggsma2008 och bygghyres2008
2. Matcha mot Typfast2008 mot Fida2009 med senaste indelningen eftersom Typfast har ett års eftersläpning.
3. Hämta Tätortsstorlek från Tatstorlek2009. Hämta varde från fstgLant och fstgOvriga
4. Matcha uppdaterad tabell från Typfast (steg 2) med utvalda från 1. Övriga variabler hämtas från Typfast.
5. Matcha mot statund2008..TykoFast som innehåller under de senaste fyra åren bebodda fastigheter.

Program 2 – *Ta fram datamaterial_2.sql*

1. Matchar på framskrivna data för födda, döda, flyttningar och folkmängd från Statund med fastighetsfilen från Program 1.

Program 3 – *AggregeraData.sas*

1. Aggregerar upp befolkningsdatat. Tar bort fastigheter med missing values

Program för Chi2-testning

Program 1 – *Chi2testInflytt.sas, Chi2testUtflytt.sas*

1. Ställ in vilka variabler och i vilken ordning de ska ha i den gruppering som ska testas. Typkod_cat som är kombination av jurform och typkod (FBR, FHR osv.) används som huvudvariabel (variabel 1). Datamaterial hämtas och aggregeras till grupper definierade av variablerna
2. Makrot *%chi2test(n)* utför Chi2-testet (för antingen inflyttningsfördelningar eller utflyttningsriskerna) med de i början av programmet inställda förutsättningarna. Dessa är p-värde för sannolikheten att grupperna ha samma fördelning (makrovariabeln *p*) samt minsta antal observationer (*infot* resp. *uftot* och *befot* i utflyttningsvarianten). Om man vill testa vissa åldersgrupper kan vissa plockas bort från testdatasetet. Två grupper åt gången testas mot varandra och slås ihop om de inte visar skillnad $p:H_0 > p$ och slås inte ihop om $p < p$. När sammanslagning inte sker sparas aktuell

- grupp undan i en fil som sedan kan användas som markör för vilka enskilda grupper som fallit ut.
3. Testningen loopas för varje grupp av fastigheter efter (de ingående variabelernas) sorteringsordning.
 4. Resultatfilen sparas, transponeras och exporteras till excelfil *fordelning.xls* eller *utrisker.xls*.

Test av framtagna grupper

Program – 1 TestAvTYKOgrupper.sas, TestAvUtvaldaGrupper.sas

Variant av programmet som utför Chi2-testet finns för nuvarande TYKO-gruppering eller de grupper som valts ut. Programmen fungerar i stora drag på samma sätt som det angivna programmet men använder makrona *%chi2test_tyko* eller *%chi2test_utvaldagrp*. Skillnaden är att i början ska de aktuella grupperna definieras efter utvalda gränser (för variablerna). Programmen testas dock i nuläget bara inflyttningsfördelningar.