



Stockholms
universitet

Analys av priser på små bostadsrätter samt villor i Uppsala

Dragos Raileanu

Kandidatuppsats 2013:3
Matematisk statistik
Juni 2013

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Analys av priser på små bostadsrätter samt villor i Uppsala

Dragos Raileanu*

Juni 2013

Sammanfattning

Den här uppsatsen har till syfte att hjälpa Fastighetsbyrån i Uppsala att bättre förstå hur bostadsmarknaden för små bostadsrätter samt villor har påverkats sedan januari 2010 i Uppsala kommun. På grund av preferenserna från Fastighetsbyrån delar vi in uppsatsen i två delar där vi analyserar bostadsmarknaden för små bostadsrätter separat och bostadsmarknaden för villor separat. Vi kommer att ta hjälp av metoder inom Regressionsanalys för att på bästa sätt få en klarare bild av vad som kan ha påverkat försäljningen av vår önskade del av bostadsmarknaden. Vi kommer att ha en del variabler att jobba med och våra analyser riktar sig mot att undersöka vilka variabler som påverkat priserna samt vilken kombination av dessa som ger den mest användbara modellen. En del av variablerna kommer vi att kunna exkludera helt från vår fortsatta analys samtidigt som vi kommer att lägga till en variabel, område, för att göra det möjligt att förklara hur området påverkar priserna.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.
E-post: dracus6@hotmail.com. Handledare: Mikael Petersson.

Abstract

This essay is designed to help real estate agency Fastighetsbyrå in Uppsala to better understand how the housing market for small condominiums and villas have been affected since January 2010 in Uppsala. Because of preferences from Fastighetsbyrå we divide the paper into two parts where we analyze the housing market for small condominiums separately from the housing market for villas. We will be using methods in Regression analysis to get the better and clearer picture of what may have affected the sales of our desired portion of the housing market. We will have some variables to work with and our analyzes aimed at examining in which variables that has affected the prices and in which combination of these that provide the most useful model. Some of the variables we will be able to exclude completely from our further analysis as well as we will add a variable, "area", to make it possible to explain if the situation of the house has any meaning, since this field is something that Fastighetsbyrå are very interested in.

Förord

Detta arbete är ett arbete som resulterar i en kandidatuppsats och är på 15 högskolepoäng vid institutionen Matematisk Statistik på Stockholms Universitet.

Jag vill tacka min handledare, doktorand Mikael Petersson, på avdelningen Matematik. Han har alltid ställt upp när det behövs och hans rådgivning har hjälpt mig att göra denna uppsats. Vidare vill jag tacka Magnus Jopevi, fastighetsmäklare och franchisetagare, samt Elna Nilsson, fastighetsmäklare, från Fastighetsbyrån i Uppsala för möjligheten att skriva detta arbete för dem. Tack även till Per-Arne Sandegren, Analys- & IT- chef på Mäklarstatistik, för att jag har fått ta del av deras statistik.

Till sist vill jag tacka mina nära och kära för att de har stöttat mig igenom hela arbetet, bra som dåliga dagar.

Innehåll

1	Introduktion	1
1.1	Inledning	1
1.2	Syfte & metod	1
2	Bakgrundsfakta	2
2.1	Uppsala kommun	2
2.2	Fastighetsbyrån	2
2.3	Datamaterial	2
3	Metoder	5
3.1	Regression	5
3.2	Enkel linjär regression	5
3.3	Multipel linjär regression	5
3.4	Signifikant förklaringsvariabel	6
3.5	Stegvis regression	6
3.6	Förklaringsgrad	6
4	Resultat	7
4.1	Bostadsrätter	7
4.1.1	Inledning	7
4.1.2	Dataundersökning & plottar	7
4.1.3	Regressionsanalys bostäder	16
4.2	Villor	19
4.2.1	Inledning	19
4.2.2	Dataundersökningar & plottar	19
4.2.3	Regressionsanalys villor	27
5	Residualanalys	30
6	Slutsatser	33
7	Diskussion	34
	Referenser	

1 Introduktion

1.1 Inledning

Bostadsmarknaden har länge varit ett hett ämne och många kopplar ett bostadsköp med en bra investering. Detta har lett till otaligt många forskningar och analyser. Frågor som "Vad är värdet på min lägenhet?" och "Vad kostar en lägenhet?" dyker ofta upp och rapporter och index finns att hitta på nästan vilken sida som helst på internet som behandlar statistik. Det var när jag planerade att flytta till Uppsala som jag började ställa liknande frågor och det visade sig att Fastighetsbyrån också var intresserade av nästan samma frågor. Vi tänker oss ett spel med tre rutor där vi i dagsläget befinner oss på ruta två och vill ta oss till ruta tre. Enligt gammal hederligt vis kan vi göra detta på tre olika sätt:

- Vi försöker ta oss till ruta tre med hjälp av de kunskaper vi har nu, tidigare erfarenheter är irrelevanta.
- Vi analyserar det vi lärde oss på ruta ett för att ta oss till ruta två och med de kunskaperna försöker vi ta oss vidare till ruta tre.
- Vi analyserar ingenting utan vi försöker direkt ta oss till ruta tre och anser att allt var och är en ren slump.

Fastighetsbyrån har valt att gå på punkt två och det är det som vi ska hjälpa de med. De är fullt medvetna om att boarean är något som påverkar priset på bostaden väldigt mycket men vad mer kan påverka priset?

1.2 Syfte och metod

Som vi vet så påverkas försäljningspriset av en rad olika faktorer och syftet med denna uppsats är att undersöka hur endast en del utvalda av dessa har fått det slutliga priset att hamna på det den hamnade på. Samtidigt görs den här uppsatsen för att jag ska bredda mina kunskaper inom regressionsanalys.

Det vi kommer att titta på är förklaringsgraden, signifikansen av enskilda variabler samt antalet variabler i modellen. Detta gör vi för att hitta den modell som bäst beskriver det vi söker. För detta kan man använda sig av bland annat programmet R men jag valde att genom hela arbetet använda mig av programmet SAS.

Vi har fått datamaterial från Mäklarstatistik, Avanza och riksbankens hemsida. Totalt har vi 1999 observationer på små bostadsrätter och 1776 observationer på villor. Perioden är mellan januari 2010 och september 2012. Området som vi ska analysera är Uppsala kommun.

2 Bakgrundsfakta

2.1 Uppsala kommun

Uppsala kommun är en kommun i Uppsala län. Centralorten är Uppsala (i framtiden kallad Uppsala tätort). Kommunen är Sveriges fjärde folkrikaste kommun med ca 200 000 invånare och centralorten är landets fjärde största tätort. Uppsala kommun består av Järlåsa, Vänge, Bällinge, Lövstalöt, Björklinge, Skyttorp, Vattholma, Storvreta, Gåvsta, Gunsta, Länna, Almunge, Knutby samt Uppsala tätort. Av dessa ca 200 000 invånare så finns det ca 40 000 studenter som studerar på Uppsala universitet och där de allra flesta bor i Uppsala tätort.

2.2 Fastighetsbyrån

Fastighetsbyrån grundades 1966 och är ett helägt dotterbolag till Swedbank. Mäklarfirman har ca 1400 anställda på ca 250 kontor runt om i landet och i delar av Spanien. Fastighetsbyrån fungerar som en franchise- kedja och varje kontor har en eller flera franchisetagare. År 2011 förmedlade Fastighetsbyrån ca 36 000 bostäder till ett försäljningsvärde av 50 miljarder kronor och är marknadsledare på den svenska bostadsmarknaden, både i omsättning och i antal affärer.

2.3 Datamaterialet

Datamaterialet har vi fått av Per- Arne Sandegren, analys- & IT-chef på Svensk Mäklarstatistik AB. Den innefattar de flesta försäljningar av små bostadsrätter i Uppsala tätort samt villor i Uppsala kommun under perioden januari 2010 till oktober 2012. Vi har även fått OMXs30 av Avanza. Vi kommer till en början att ta med så många variabler som möjligt för att sedan exkludera de som visar sig inte vara signifikanta i senare undersökningar. Materialet innehöll variablerna:

År

Året då bostaden såldes

Månad

Månaden då bostaden såldes

Lan

Länet som bostaden befinner sig i

Kommun

Kommunen som bostaden befinner sig i

Kontraktsdatum

Det exakta datumet då kontraktet blev skriven

Pris

Det exakta beloppet som bostaden såldes för

Boyta

Bostadens beboliga yta

Manavgift

Månadsavgiften

Rum

Antalet rum som bostaden är uppdelad i

Kboy

Antalet kronor per kvadratmeter

SCB_KB

Ett normerat K/T-tal då vi nu har ett inflöde med både 2012 och 2009-års taxeringsvärden. K/T-talet som statistiskt begrepp används av både Mäklarstatistik och SCB och är alltså köpesumman dividerat med taxeringsvärdet.

Taxeringsvarde

Det värde på en fastighet som ligger till grund för fastighetsskatt och tidigare också förmögenhetsskatt. Taxeringsvärdet fastställs vid fastighetstaxering och ska motsvara 75 % av fastighetens marknadsvärde, med viss eftersläpning.

Adress

Den exakta adressen var bostaden är belägen

Taxar

Taxeringsåret

Utöver dessa variabler så är Fastighetsbyrån intresserade av hur försäljningspriset påverkas av börsen, reporäntan samt området där bostaden är belägen, så vi lägger till ytterligare tre variabler:

OMX

Börsvärdet, på OMXs30, dagen då försäljningen av bostaden gjordes

Reporanta

Riksbankens nivå på styrräntan vid försäljningen av bostaden

Område

Området som bostaden befinner sig i. Denna variabel kommer att beskrivas mer utförligt senare

3 Metoder & viktiga begrepp

3.1 Regression

Syftet med metoden är att hitta en funktion som på bästa sätt passar de observerade data. Under hela arbetet använder vi oss av regression som tillämpar minsta kvadrat metoden. Regressionsmodeller kännetecknas enligt Rolf Sundbergs kompendium Tillämpad Matematisk Statistik av att en mätstorhet under slumpmässig osäkerhet beror genom ett linjärt funktionssamband av en eller flera precis kända variabler.

3.2 Enkel linjär regression

En enkel linjär regression definieras enligt följande:

$$Y_i = \alpha + \beta * x_i + \epsilon_i$$

Där Y_i är responsvariabeln, x_i den förklarande variabeln, α och β är parametrar och ϵ_i de slumpmässiga variationerna. Vi har även att i är tal mellan 1 och N där N är antalet gjorda mätningar. För att få en enkel linjär regression så betraktar vi ϵ_i som oberoende och normalfördelad med väntevärde 0 och varians σ^2 . Modellen har väntevärdesfunktionen:

$$\mu = \alpha + \beta * x$$

3.3 Multipel linjär regression

En multipel regression gör man då man misstänker att responsvariabeln Y_i beror på två eller fler förklaringsvariabler. Den här formen av linjär regression definieras som:

$$Y_i = \alpha + \beta_1 * x_{1i} + \beta_2 * x_{2i} + \dots + \beta_n * x_{ni} + \epsilon_i$$

Om vi betraktar ϵ_i som oberoende och normalfördelad med väntevärde 0 och varians σ^2 och $i=1, \dots, N$ där N är antalet gjorda mätningar. Modellen har väntevärdesfunktionen:

$$\mu = \alpha + \beta_1 * x_1 + \dots + \beta_n * x_n$$

3.4 Signifikant förklaringsvariabel

En nollhypotes förkastas om det observerade utfallet inträffar i mindre än 5 procent av fallen givet nollhypotesen. En förklaringsvariabel kallas signifikant om den vid test av hypotesen att motsvarande parameter är noll ger signifikant utslag.

3.5 Stegvis regression

Enligt Rolf Sundbergs Lineära Statistiska Modeller är "Stepwise regression" en något mer avancerad version av "forward selection". Det den gör är att efter varje steg så kontrolleras för var och en av de tidigare införda variablerna ifall de ger signifikanta utslag vid test av hypotesen att motsvarande parametrar är noll och eventuella variabler som inte längre ger signifikanta utslag elimineras ur modellen. Det kan vara så att ett par variabler tillsammans beskriver data väldigt bra men var för sig beskriver de mindre än en tredje variabel som blir överflödigt när de andra paren är med.

3.6 Förklaringsgrad

Förklaringsgraden anges i mått av procent och är den del av variationen i Y som förklaras med hjälp av X . Förklaringsgraden benämns som R^2 och är troligen det vanligaste anpassningsmålet i samband med linjära modeller, vanligast enligt Rolf Sundbergs kompendium Tillämpad Matematisk Statistik. Den definieras som följande:

$$R^2 = KVS_{\text{modell}} / KVS_{\text{total}} = 1 - KVS_{\text{residual}} / KVS_{\text{total}}$$

Där vi har att

$$KVS_{\text{total}} = \sum (y_i - \bar{y})^2$$

$$KVS_{\text{modell}} = \sum (\hat{y}_i - \check{y})^2 \quad , \text{ där } \check{y} \text{ är medelvärdet av } \hat{y}$$

$$KVS_{\text{residual}} = \sum (y_i - \hat{y})^2$$

4 Resultat

4.1 Bostadsrätter

4.1.1 Inledning

I Uppsala finns det närmare 40 000 studenter så det är inte förvånande att Fastighetsbyrån intresserar sig för denna grupp. De flesta studerar på Uppsala universitet och bor så nära skolan som möjligt. Med detta i åtanke så begränsar vi oss till endast Uppsala tätort. Vi har fått ta del av data för 1999 sålda bostadsrätter i Uppsala tätort. För att få en klarare bild av våra data så börjar vi med att studera plottar för att se ifall det finns någon korrelation mellan våra variabler. Ifall fler variabler samvarierar i materialet så försöker vi reducera antalet till en enda. Här näst finner vi resultaten och resonemangen kring plottarna.

4.1.2 Dataundersökning & plottar

Vi börjar med att titta på våra variabler. För bostadsrätterna har vi samtliga variabler nämnda ovan utom "Taxeringsvarde".

För tillfället har vi tre variabler som i princip säger oss precis samma sak, "År", "Kontraktsdatum" och "Månad". Eftersom vi inte behöver veta det exakta datumet väljer vi att istället för tre separata variabler skapa en enda variabel som vi döper till "Tid" som representerar månaderna. Denna variabel går mellan 1 och 33 där talet 1 motsvarar januari 2010 och talet 33 motsvarar september 2012. Nu kan de andra tre variablerna uteslutas ur vårt material.

Vidare tittar vi på variablerna "Kommun" och "Lan". Dessa två variabler säger oss samma sak och bidrar inte med något eftersom samtliga observationer befinner sig i Uppsala tätort, så dessa variabler kan vi också utesluta.

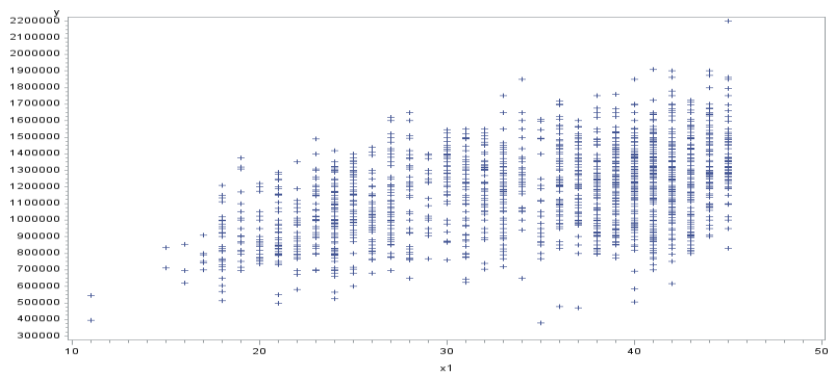
Vidare kan vi konstatera att "SCB_KB" inte är av ett intresse för bostadsrätter då vi har exkluderat variabeln "Taxeringsvarde". Utan taxeringsvärdet säger denna variabel ingenting alls så den tar vi bort från vårt material.

Under variabeln "Adress" kan vi hitta samtliga gatunamn och gatunummer på de sålda bostadsrätterna. Eftersom Fastighetsbyrån nu riktar in sig på studenterna i Uppsala är de även intresserade av att se hur avståndet från bostad till skolan kan påverka försäljningspriset.

Det vi gör är att vi skapar variabeln "Område" där vi delar in Uppsala tätort i tre områden, område A, B och C. Ifall bostaden befinner sig inom en radie på 300 meter från en universitetslokal eller byggnad (tentamenslokaler räknas inte hit) så säger vi att bostaden befinner sig i område A. Ifall den befinner sig utanför 300 meter men inom en radie av en kilometer från en universitetslokal eller byggnad så befinner den sig i område B. Alla bostäder som befinner sig utanför en kilometers radie klassar vi sedan som område C.

Under hela arbetet kommer vi att använda variabeln "Pris" som responsvariabel. Detta gör att ingen av förklaringsvariablerna bör vara en direkt funktion av priset. Variabeln "Kboy" bör således inte vara en förklaringsvariabel och vi tar då även bort denna variabel ur vårt material

Vi kan nu börja studera lite plottar och vi väljer att börja plotta samtliga förklaringsvariabler mot "Pris".

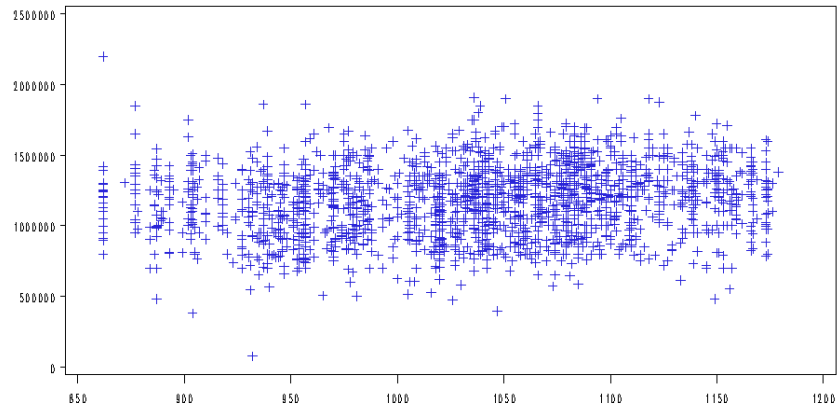


Pris mot boarean

Variabel	Frihetsgrader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	696606	22296	31.24	<.0001
Boarea	1	13850	626.41	22.11	<.0001

Vi ser ett ganska tydligt samband. P- värdet är väldigt lågt vilket bekräftar våra misstankar. Vi ser att boarean har en positiv inverkan på priset.

Detta är föga förvånande och sambandet mellan dessa två variabler visste Fastighetsbyrån redan om. Vi lägger inte så mycket mer tid på det uppenbara utan vi går vidare till resterande förklaringsvariabler.

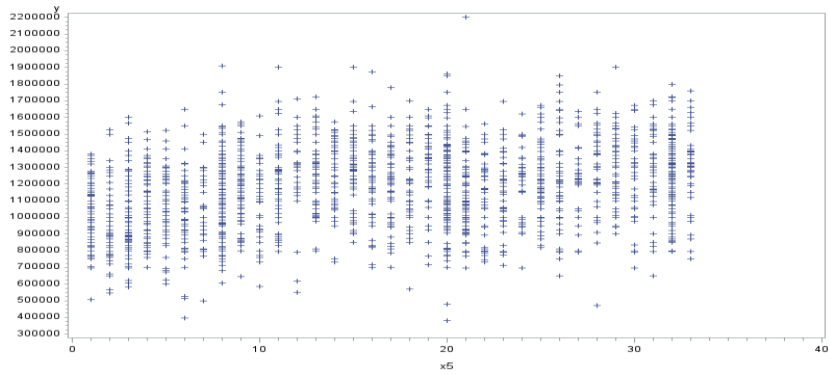


Priset mot OMX

Vi kan här inte dra någon slutsats om att det finns ett samband mellan antalet sålda bostadsrätter och värdet på OMX och har lite svårt att dra en direkt slutsats om huruvida priset förändras eller inte. En skattning av variabeln ger följande.

Variabel	Frihets grader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	773198	67638	11.43	<.0001
Omx	1	390	65	5.99	<.0001

Ett positivt samband mellan omx och priset. Vi får åter igen ett mycket lågt P- värde.



Priset mot Tid

Ännu en gång kan vi inte dra några slutsatser utifrån denna plott.

Vi ser på tabellen på nästa sida att bostadspriserna ser ut att ha ökat med tiden under den tidsperiod vi studerar. Vi gör därför en tabell där vi undersöker denna förklaringsvariabel lite närmare. Vi skattar även tiden och ser att den har en positiv inverkan på priset.

Variabel	Frihets grader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	1053187	10621	99.16	<.0001
Tid	1	7329	544	13.47	<.0001

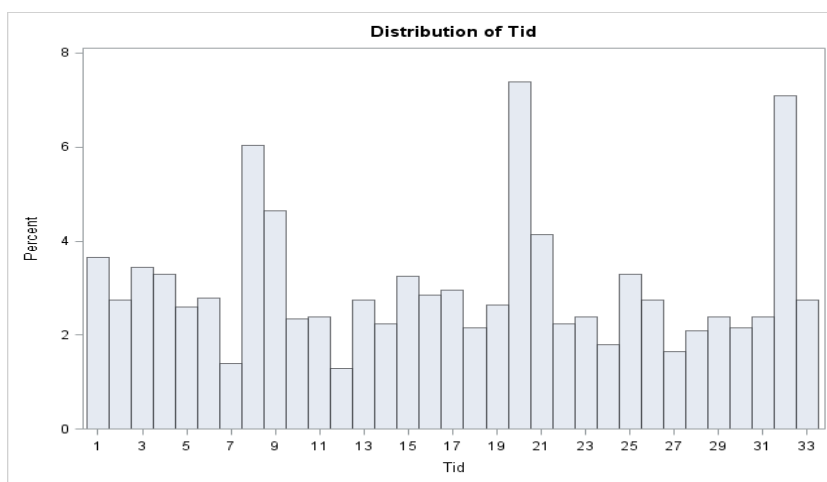
<u>Tid</u>	<u>Antal sålda bostäder</u>	<u>Lägsta priset</u>	<u>Högsta priset</u>	<u>Pris medelvärde</u>
Januari-2010	73	505000	1380000	1047740
Februari	55	545000	1525000	965818
Mars	69	580000	1600000	1005870
April	66	700000	1515000	1086212
Maj	52	600000	1520000	1051455
Juni	56	395000	1650000	1010893
Juli	28	500000	1500000	1050735
Augusti	121	605000	1910000	1131416
September	93	645000	1575000	1163011
Oktober	47	585000	1610000	1087766
November	48	795000	1900000	1193542
December	27	480000	1710000	1236185
Januari-2011	55	800000	1725000	1246090
Februari	45	730000	1575000	1250410
Mars	65	850000	1900000	1281077
April	57	700000	1875000	1219912
Maj	59	700000	1780000	1211271
Juni	43	570000	1700000	1216209
Juli	53	715000	1650000	1315000
Augusti	148	380000	1860000	1199713
September	84	695000	2200000	1129881
Oktober	45	730000	1560000	1104667
November	48	710000	1695000	1162500
December	36	695000	1620000	1188264
Januari-2012	65	820000	1670000	1228623
Februari	55	650000	1850000	1250546
Mars	33	790000	1635000	1224849
April	42	470000	1750000	1289488
Maj	48	900000	1900000	1277847
Juni	43	695000	1670000	1258140
Juli	48	650000	1700000	1291667
Augusti	142	795000	1800000	1253504
September	55	750000	1760000	1289090

De viktigaste kolumnerna ovan är naturligtvis antalet sålda bostäder och medelvärdet på priset. Medelvärdet ökar nästan stadigt och det vet vi beror på att bostadsmarknaden har blivit mycket mer intressant.

Som vi sa innan så har fler och fler kopplat ett köp av en bostad med en bra framtida investering.

Detta har lett till att efterfrågan på bostädsrätter har ökat och i sin tur har man kunnat pressa upp priserna. Vi tittar närmare på kolumnen antalet sålda bostäder. Vi ser att här finns det också ett mönster. Augusti månad och september månad är helt dominerande. Detta beror troligen på att de flesta programmen på Uppsala universitet börjar i slutet på augusti och början på september. Att september också är en populär månad för bostadsköp kan bero på att det andra antagningsbeskedet kommer i augusti så man väntar med att köpa bostad till september.

Vi tar och för ett histogram mellan antalet sålda bostadsrätter och månaden



Antalet sålda bostadsrätter mot månaderna

Vi ser ett klart samband mellan dessa två variabler. Men som vi nämnde ovan så säljs det väldigt många fler lägenheter under augusti och september än någon annan period under året. Så det kan vara intressant att se ifall priset också ändras och inte bara antalet. Det vi gör är att införa dummyvariabler. För månaderna augusti och september, varje år, får den värdet 1 och för alla andra skriver vi 0. Vi får då följande tabell för DummyTid

Variabel	Frihets grader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	1170835	6690	175	<.0001
DummyTid	1	20069	11815	1.7	0.0895

Vi ser att variabeln är signifikant på signifikansnivå 10% men inte på signifikantsnivå 5%. Dessa två månader det totala priset positivt.

När vi kommer till området så ger det oss ingenting alls att studera plotten. Det vi istället gör är att dela in områdena i område A, B samt C. Vi skapar sen två dummyvariabler där området B och C får vara 1 och området A får vara 0 samt området C får vara 1 och områdena A och B 0. Dessa variabler döper vi till DummyB och DummyC. Anledningen till varför vi väljer att inte ha en dummy för området A beror på att vi valt området A som ett område mellan noll och 300 meter från en universitetsbyggnad. Det som intresserar oss är att se om och i så fall hur mycket det skiljer sig från att ha en bostadsrätt längre ut från detta område. Vi misstänker att ju längre ifrån en universitetsbyggnad som bostaden befinner sig ju mer sjunker priset. Vi väljer att ha en dummyvariabel för både område B och C ihop för att på så sätt få ut "skillnad i att bo utanför område A (alltså B och C tillsammans)". Vår andra dummy är en simpel dummy som berättar skillnaden i att bo endast i området längst ut, alltså C. Vi börjar med att se hur medelvärdet är fördelade i respektive område.

<u>Område</u>	<u>Antal</u>	<u>Pris medelvärde</u>
A	301	1262028
B	1012	1228968
C	686	1063979

Som vi kan se så skiljer sig lägenhetspriserna lite mellan området A och B och ännu mera mellan A och C eller B och C. Vi skattar nu våra dummyvariabler och får följande

Variabel	Frihets grader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	1266221	14078	89.94	<.0001
DummyB	1	-104662	15271	-6.85	<.0001

Skattade DummyB

Variabel	Frihets grader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	1236268	6422	192.48	<.0001
DummyC	1	-172181	10973	-15.69	<.0001

Skattade DummyC

Vi ser att vi har en negativ skattning vilket är det vi misstänkte. Det är inte helt säkert att det finns ett direkt orsakssamband här. Nu råkar det vara så att de flesta skolorna är väldigt centrala och som vi vet så är priserna i innerstaden högre än de i förorterna. Så ju längre ut från en universitetsbyggnad man köper sin bostadsrätt, desto längre ut från centrum kommer man och därför blir priset lägre.

Det skulle kunna förklara det negativa sambandet mellan priset och området. Men vi har säkerställt att ju längre ifrån en universitetsbyggnad man bor desto mindre betalar man för sin bostadsrätt.

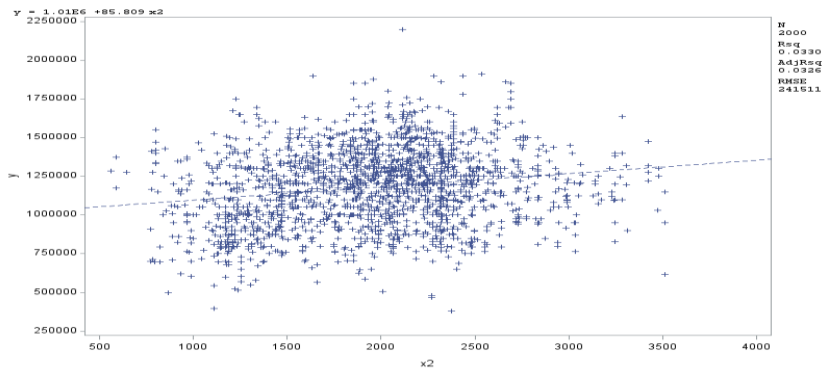
Vi hade kunnat vända på denna analys och haft en dummyA och dummyB. Alla observationer i område A och B hade fått vara 1 och de i område C 0 och kallat denna för dummyB samtidigt som dummyA hade fått vara som vår nuvarande dummyC. Då hade vi fått positiva skattningar, det vill säga att det blir dyrare att bo närmare en universitetsbyggnad.

När vi nu tittar på Reporäntan så väljer vi även här att inte ta med plotten utan vi väljer att direkt titta på skattningen av variabeln för att dra en slutsats.

Variabel	Frihetsgrader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	1053536	11640	90.51	<.0001
Reporänta	1	25170	2097.97085	12	<.0001

Ännu en gång har vi ett väldigt lågt P- värde och vi konstaterar att den har en positiv inverkan på priset.

Vi avslutar nu denna del med att se hur månadsavgiften påverkar priset. Denna variabel var inte av större vikt för Fastighetsbyrån då det anses som en självklarhet att månadsavgiften ska påverka priset. Även om denna variabel redan är självklar så väljer vi att ta med den av två anledningar: vi har data för denna variabel samt att vi kan analysera den för att bekräfta det man redan misstänker.



Priset mot månadsavgiften

Vi skattar variabeln och ser att även den ger en liten men positiv effekt på priset.

Variabel	Frihets grader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	1010496	21003	48.11	<.0001
Månadsavgift	1	85	10	8.26	<.0001

Redan här kan vi misstänka att månadsavgiften i kombination med boytan kommer att ha en negativ inverkan på priset. Vi misstänker alltså att för två bostadsrätter i samma storlek så förväntar vi oss att bostaden med lägre månadsavgift kommer att vara dyrare.

4.1.3 Regressionsanalys bostadsrätter

För att få en så bra bild som möjligt av datamaterialet går vi vidare med att undersöka hur bra vi skulle kunna förklara variationen i data med hjälp av de variabler som vi har kvar, nämligen variablerna:

Boyta
OMX
DummyC
DummyB
Reporänta
Tid
Månadsavgift

Vi börjar med att göra en multipel regression på priset med de övriga variablerna som förklaringsvariabler. ANOVA tabellen ger oss då

Källa	Frihetsgrader	Kvs	F- värde	Pr > F	R ²
Modell	7	6.345149E13	316.4	<.0001	0.5265
Error	1992	5.706893E13			
Corrected Total	1999	1.205204E14			

Vi har en relativt låg förklaringsgrad och även efter att vi har utfört en stepwise, forward samt backward regression så har vi samtliga variabler med på 5 % signifikantsnivån. Vi såg tidigare att de flesta variabler för sig gav en positiv inverkan på priset. Vi får följande tabell när vi skattar alla variabler

Variabel	Skattad parameter
Intercept	233183
Boyta	23135
Månadsavgift	-151
Reporänta	6082
Tid	7452
DummyC	-175460
Omx	408
DummyB	-96612

Priset påverkas negativt av månadsavgiften samt områdena B och C. Att bo längre ut blir alltså billigare. Förklaringsgraden är oförrändrad men variablerna är olika viktiga. Vi börjar med att se signifikansen för samtliga variabler,

Variabel	Pr > F
Boyta	<.0001
DummyC	<.0001
Tid	<.0001
Månadsavgift	<.0001
DummyB	<.0001
Omx	<.0001
Reporänta	0.0119

Samtliga variabler är signifikanta på 5 % nivån men är det den här modellen som vi verkligen söker?

Vi tar och ser hur mycket varje variabel bidrar med till den totala förklaringsgraden. Vi får då följande tabell

Variabel	R ²
Boyta	0.1951
DummyC	0.1398
Tid	0.1033
Månadsavgift	0.0495
DummyB	0.0194
Omx	0.0180
Reporänta	0.0015

Resultat efter att vi använt stepwise regression

Som vi kan se så är variabeln Reporänta signifikant på 5 % nivån men den bidrar knappt med någonting till förklaringsgraden. Eftersom vi vill ha en modell med så få variabler som möjligt så väljer vi att ta bort denna variabel. Sen kan det vara svårt att bestämma sig för ifall man även ska ta bort variabeln omx och dummyB. Dessa bidrar med lite mindre än 2 % var vilket inte är mycket. Är det så att man anser att förklaringsgraden är väldigt låg så räcker det med att endast bortse från reporäntan. I vårt fall anser vi att strax under 50 % är tillräckligt högt och vi föredrar att ha tre färre variabler på bekostnad av 4 % i förklaringsgrad. Så vår slutliga modell blir

Källa	Frihetsgrader	Kvs	F- värde	Pr > F	R ²
Modell	4	5.8749E13	474.5	<.0001	0.4872
Error	1995	6.1845E13			
Corrected Total	1999	1.20520E14			

där vår totala förklaringsgrad uppnår strax över 48 %. De skattade variablerna blir

Variabel	Frihetsgrader	Parameter skattning	t- värde	Pr > t
Intercept	1	607866	30.4	<.0001
Boyta	1	22157	32.14	<.0001
Månadsavgift	1	-142.13	-13.93	<.0001
Tid	1	8670.25	21.24	<.0001
DummyC	1	196534	8332.6	<.0001

Skattade parameter i den slutliga modellen

Vår låga förklaringsgrad kan bero på många saker men en av dem är att vi endast har riktat in oss på bostadsrätter som får vara max 45 kvm. Hade vi tagit med samtliga bostadsrätter skulle förklaringsgraden vara högre.

4.2 Villor

4.2.1 Inledning

Som vi nämnde tidigare består Uppsala kommun av Järlåsa, Vänge, Bällinge, Lövsalöt, Björklinge, Skyttorp, Vattholma, Storvreta, Gåvsta, Gunsta, Länna, Almunge, Knutby samt Uppsala tätort. När det gällde villorna så hade inte Fastighetsbyrån några restriktioner till vilka villor vi specifikt skulle titta på utan vi kommer ta med samtliga villor som såldes under perioden 1 januari 2010 till 30 september 2012 och utföra samma tester som vi utförde på bostadsrätterna. Eftersom villor inte riktigt är en students förstahandsval så har vi inga begränsningar på hur stor boytan får vara och inte heller om avståndet från en universitetsbyggnad påverkar vårt pris.

4.2.2 Dataundersökningar och plottar

Vi börjar än en gång med att titta på våra variabler. För villorna har vi nu samtliga variabler. Vi har då även med "Taxeringsvärde" vilket vi saknade när vi såg på bostadsrätter.

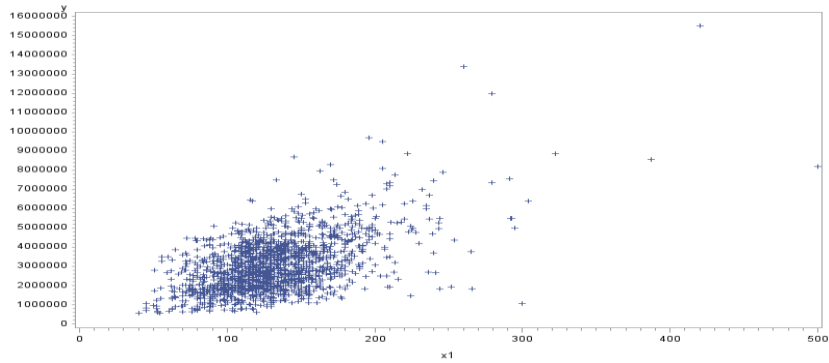
Även här har vi tre variabler som säger oss nästan samma sak, "Mån", "År" och "Kontraktsdatum". Dessa tre variabler gör vi då om och bildar den nya variabeln "Tid" som endast representerar månaderna. Denna går mellan 1 och 33 där talet 1 motsvarar januari 2010 och talet 33 motsvarar september 2012. Nu kan de andra tre variablerna än en gång uteslutas ur vårt material.

Variablerna "Län" och "Kommun" kan vi också ta bort från vårt datamaterial eftersom samtliga observationer ligger i samma län och samma kommun.

Variabeln "Adress" ger oss information om den exakta gatan och numret på bostaden. Denna variabel är, precis som på bostadsrätter, viktig för Fastighetsbyrån. Frågan som vi har fått av Fastighetsbyrån är ungefär som på bostadsrätter. Men, som vi nämnde ovan, så kommer vi inte rikta in oss på universitetsbyggnader, utan här kommer vi ha den mer "vanliga" mittpunkten som centrum, nämligen Stora torget. Stora torget är Uppsala tätorts T- Centralen och det vi gör är att dela in hela Uppsala kommun i 3 områden. Alla villor inom en radie av en kilometer ifrån Stora torget kommer att tillhöra område A. Alla villor som befinner sig utanför denna radie men fortfarande befinner sig i Uppsala tätort kommer att tillhöra område B. De resterande villorna som befinner sig utanför Uppsala tätort tillhör således område C.

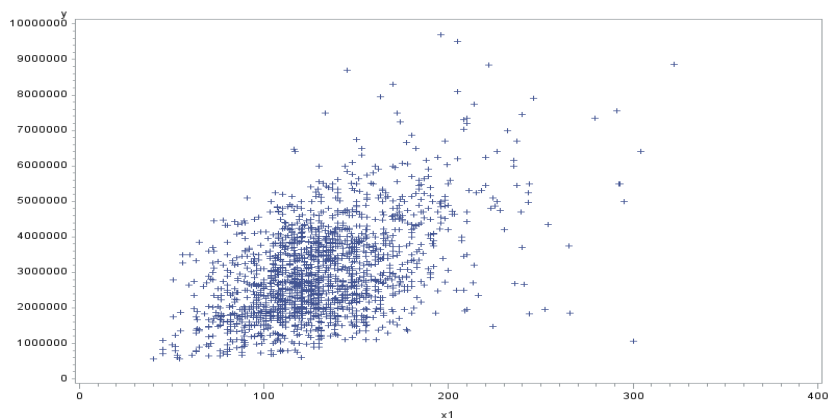
Även här kommer vi att använda variabeln "Pris" som responsvariabel så detta gör att ingen av förklaringsvariablerna bör vara en direkt funktion av priset. Variablerna "Kboy" och "SCB_KB" bör således inte vara en förklaringsvariabel eftersom "SCB_KB" är ett K/T tal, det vill säga den totala köpesumman (eller priset) delat med taxeringsvärdet och "Kboy" är kronor per kvadratmeter.

Vi är nu redo att studera plottar och vi börjar med att plotta varje förklaringsvariabel för sig mot responsvariabeln "Pris".



Pris mot boarea

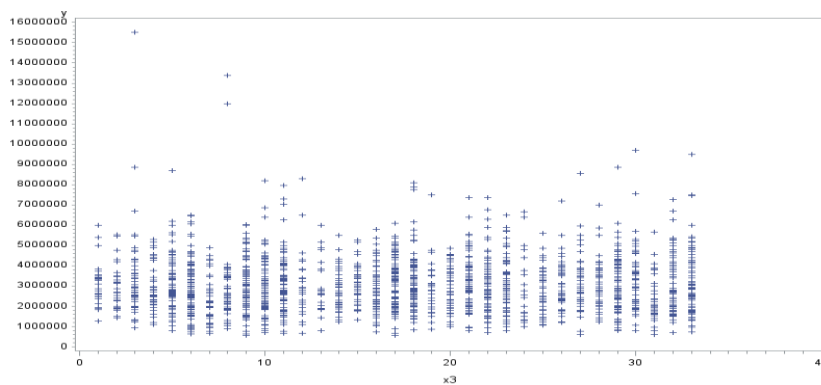
Observationerna tycks samla ihop sig till en boll runt 75 till 175 kvadratmeter. Vi ser även att vi har några observationer som avviker kraftigt från de andra. Vad detta beror på kan man spekulera i men eftersom vi har så många observationer är det ingen större förlust att ta bort ett fåtal av våra observationer. Det vi gör är att vi säger till SAS att ta bort samtliga observationer som har en boarea över 300 kvadratmeter eller ett försäljningspris som överstiger 10 000 000. Detta innebär att vi tar bort 6 observationer och vi får en ny plott som ser ut på följande vis



Plotten ser ut att stämma överens med våra misstankar om att boytan och priset skulle vara korrelerade.

Variabel	Frihetsgrader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	448332	99792	4.49	<.0001
Boarean	1	19758	728.35758	27.13	<.0001

Vi kan här se att det finns ett samband mellan boarean och priset. Detta är, som för bostadsrätter, ingen nyhet så vi går vidare utan att lägga allt för mycket tid på denna variabel.



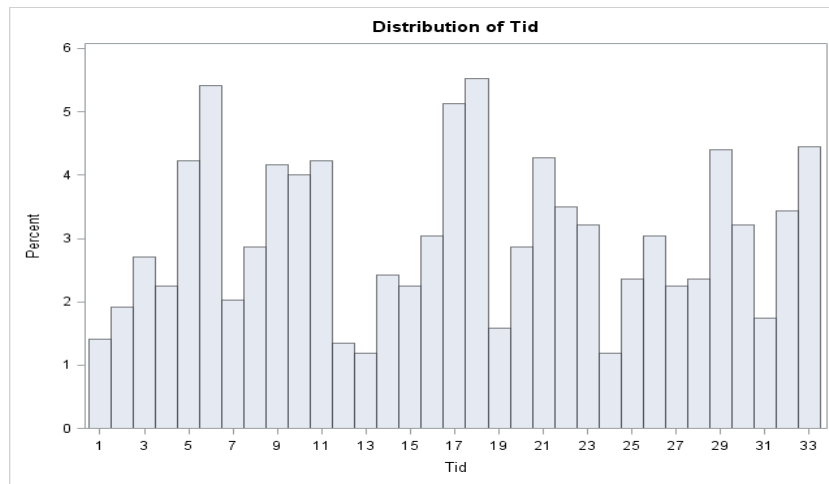
Pris mot Tid

Vi kan inte se något mönster endast av att studera plotten. Det vi även misstänker är att antalet köpta villor också följer ett mönster. Vi skattar först tiden och gör sen en liknande tabell som vi gjorde för bostadsrätterna.

Variabel	Frihetsgrader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	1054121	10561	99.81	<.0001
Tid	1	7325	540	13.54	<.0001

<u>Tid</u>	<u>Antal</u>	<u>Lägsta priset</u>	<u>Högsta priset</u>	<u>Pris medelvärde</u>
Januari-2010	25	1270000	6000000	3128400
Februari	34	1450000	5525000	2791177
Mars	49	920000	15500000	3495633
April	40	1100000	5300000	3052500
Maj	75	800000	8700000	3086333
Juni	96	650000	6500000	2987370
Juli	36	670000	4900000	2448333
Augusti	53	895000	13400000	2881509
September	74	570000	6030000	3106216
Oktober	72	695000	8200000	3020833
November	75	675000	7950000	3109933
December	24	675000	8300000	3101875
Januari-2011	21	800000	5985000	3020048
Februari	43	1235000	5500000	2892791
Mars	40	1330000	5250000	3096875
April	54	730000	5800000	2906482
Maj	91	565000	6080000	3023352
Juni	98	830000	8100000	3039194
Juli	28	875000	7500000	2841964
Augusti	51	995000	4850000	2837451
September	76	800000	7350000	3294737
Oktober	63	695000	7350000	3202631
November	56	795000	6500000	3036518
December	21	1000000	6660000	2852619
Januari-2012	42	1060000	5600000	2823929
Februari	54	1205000	7200000	2931111
Mars	41	600000	8550000	3170102
April	42	1230000	7000000	3060714
Maj	79	795000	8870000	3115253
Juni	56	810000	9700000	3385536
Juli	32	600000	5650000	2425391
Augusti	61	715000	7250000	3274057
September	79	740000	9500000	3187405

En snabb titt på denna tabell gör att vi kan misstänka att priset påverkas beroende på vilken månad man väljer att köpa. Men vi ser även ett mönster mellan antalet sålda villor och vilken månad man väljer att köpa villan. Vi gör även ett histogram för att lättare se ifall att detta kan stämma.



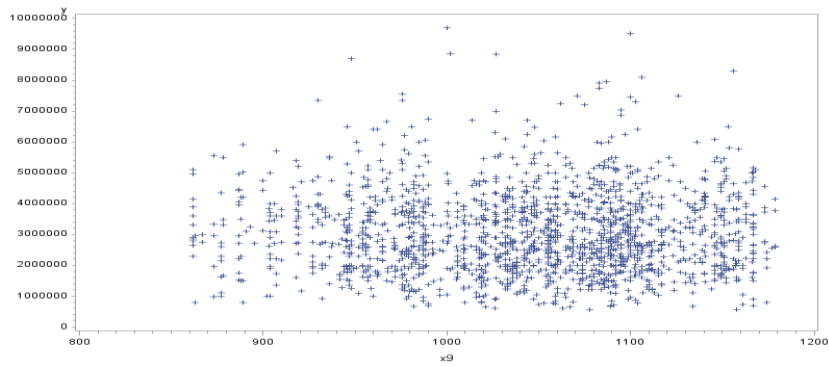
Nu blir det lättare att se ett samband mellan månaden och antalet sålda villor. Vi ser att man helst köper en villa på våren eller hösten. Vad detta beror på kan vara olika anledningar men vi vet, av vad Fastighetsbyrån berättat för oss angående villor, att personer som väljer att sälja sin villa gör gärna det när man kan ta en fin bild på sin villa för att lägga upp den i annonsen. Det ska gärna vara fint väder, grönt gräs och kanske blommor. Varför man då inte säljer särskilt mycket över sommaren kan bero på till exempel att man är bortrest och väljer att köpa eller sälja sin villa när man kommer hem igen.

Vi gör även en liknande undersökning som vi gjorde för bostadsrätterna där vi skapar en dummy variabel för perioderna maj-juni och september- oktober. En skattning för vår dummyvariabel ger oss följande

Variabel	Frihets grader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	2984456	46033	64.83	<.0001
DummyTid	1	130715	66303	1.97	0.0488

Skattning av dummyTid

Som vi kan se så har vi en positiv inverkan på priset av både tiden och vår dummyvariabel. Vi spekulerar inte vidare på detta utan väljer att gå vidare till att se om börsen, OMX, kan påverka priset.



Pris mot OMX

Det finns ingen anledning att misstänka ett direkt samband mellan OMX och priset genom att endast studera denna plott och vi får följande tabell när vi skattar variabeln

Variabel	Frihets grader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	3172992	479281	6.62	<.0001
DummyTid	1	-120	458	-0.26	0.7929

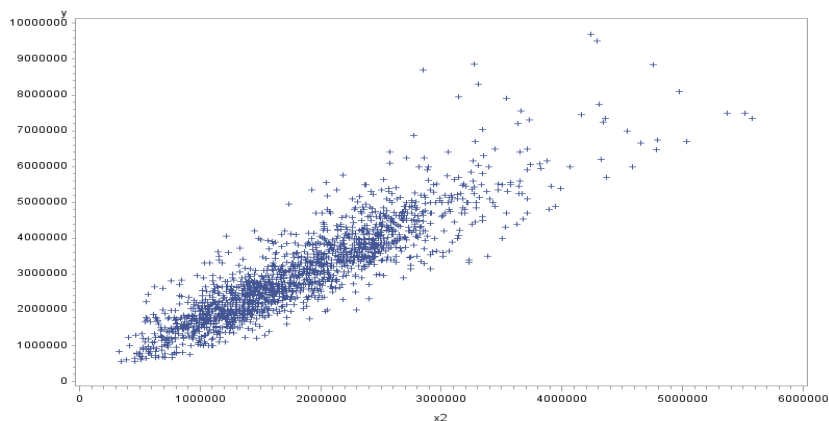
Vi har ett väldigt stort P- värde och ett negativt samband mellan priset och omx.

En plott av reporäntan mot priset skulle inte säga oss någonting så vi väljer skatta variabeln för att sen gå vidare till mer intressanta variabler.

Variabel	Frihets grader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	3041722	74454	40.85	<.0001
DummyTid	1	1155	13412	0.09	0.9314

Det räcker med att säga att vi har ett stort värde på P.

Vi tittar nu på en intressant variabel, taxeringsvärdet. Det är lätt att tro att denna variabel inte ska vara med ty taxeringsvärdet är en fastighetstaxering på 75 % av fastighetens marknadsvärde. Vi sa tidigare att vi inte tog med variabler som berodde på priset eftersom den var vår responsvariabel. Men priset som vi har är det priset som villan blev såld för och inte marknadsvärdet. Därför är denna variabel intressant att undersöka.



Vi ser ett kraftigt samband mellan taxeringsvärdet och priset. Bilden är tydlig och vi är redan här övertygade om att taxeringsvärdet är korrelerad med priset.

Variabel	Frihets grader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	165702	36652	4.52	<.0001
Taxeringsvärde	1	1.56482	0.01825	85.76	<.0001

Ett väldigt lågt P- värde vilket inte är så konstigt. Taxeringsvärdet är trots allt en bestämd procent, 75 %, av fastighetens marknadsvärde, med viss eftersläpning.

Vi tittar nu på vår sista förklaringsvariabel, område, och konstaterar ganska snabbt att vi inte kan säga särskilt mycket om den endast genom att studera plotten. Vi kommer här att göra som vi gjorde för bostadsrätterna. Vi inför dummyvariablerna dummyC samt dummyB där den första är 1 för området C och 0 för områdena A samt B och för den senare är områdena B och C 1 och området A 0. Vi använder här samma tankesätt som för bostadsrätterna. Området A är ett område mellan 0 och 1000 meter från centrum och vi vill se om och i så fall hur priserna då påverkas ju längre ifrån området A man kommer. DummyB kan då tolkas som "skillnaden i att bo utanför området A" och dummyC kan tolkas som "skillnaden i att bo i område C.

Variabel	Frihets grader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	5347200	189998	28.14	<.0001
DummyB	1	-2366203	192724	-12.28	<.0001

Parameterskattning dummyB

Variabel	Frihets grader	Skattad parameter	Standard Error	T Värde	Pr > t
Intercept	1	4008972	38510	104.10	<.0001
DummyC	1	-1755370	52033	-33.74	<.0001

Parameterskattning dummyC

Som vi misstänkte så har vi negativa skattningar. Att bo längre ut från centrala Uppsala tätort kommer att resultera i att man får betala ett lägre pris. Och precis som för bostadsrätterna hade vi kunnat vända på våra dummyvariabler och ha en dummyvariabel som heter dummyA och en som heter dummyB där områdena A och B skulle få vara 1 och C 0 och där dummyA skulle få vara som vår nuvarande dummyC. Vi hade då fått positiva skattningar och vi skulle då tolka det som att det blev dyrare ju närmare Stora torget man kom.

4.2.3 Regressionsanalys villor

Vi har nu följande variabler som vi ska undersöka:

Boyta

OMX

DummyB

DummyC

Reporänta

Tid

Taxeringsvärde

Vi börjar med att göra en multipel regression på priset med de övriga variablerna som förklaringsvariabler. ANOVA tabellen ger oss då

Source	Frihetsgrader	Kvsum	F-värde	Pr > F	R ²
Modell	7	2.85748E15	1159.61	<.0001	0.8208
Error	1772	6.23836E14			
Corrected Total	1779	3.48131E15			

Vi ser att en väldigt stor del av variationen kan förklaras av dessa variabler, hela 82.08 %. Vi vill ha en modell där vi har så få förklaringsvariabler som möjligt. Vi utför då en stepwise, backward samt forward regression för att se vilka av variablerna som vi kan exkludera utan att vår förklaringsgrad sjunker allt för mycket. Precis som tidigare väljer vi att ställa kraven på 0,05 i programvaran SAS. Den nya ANOVA tabellen blir således:

Source	Frihetsgrader	Kvsum	F-värde	Pr > F	R ²
Modell	5	2.85724E15	2031.64	<.0001	0.8207
Error	1774	6.24076E14			
Corrected Total	1779	3.48131E15			

Stepwise Regression valde då att ta bort variablerna Omx och Reporänta (Backward och Forward gav samma resultat). Även om 2 variabler har sorterats bort så förklarar de resterande mer än 80 procent av variationen. Värt att poängtera är att endast taxeringsvärdet står för de flesta procenten av variationen. Om vi tittar och rangordnar hur mycket av förklaringsgraden varje variabel av de 5 representerar får vi följande tabell,

Variabel	R ²	Pr > F
Taxeringsvärde	0.8053	<.0001
DummyC	0.0046	<.0001
Boyta	0.0079	<.0001
DummyB	0.0023	<.0001
Tid	0.0007	0.0085

Resultat efter att vi använt stepwise regression

Vi ser här att om vi tänker så som vi tänkte oss för bostadsrätterna så kan vi ta bort samtliga variabler utom taxeringsvärdet eftersom vi söker en så förenklad modell som möjligt. Tar vi bort alla variabler utom taxeringsvärdet får vi en förklaringsgrad på över 80 %. Denna variabel är så omfattande att de resterande variablerna framstår som nästan obetydliga. Att taxeringsvärdet ökar eller sjunker beror på, som vi tidigare sa, att den följer marknadspriset för fastigheten. Den kommer således alltid att verka signifikant. Taxeringsvärdet följer marknadspriset för villan men vi vill se vad priset beror på. Vi utför därför en ny regression där vi bortser från taxeringsvärdet. Den nya ANOVA tabellen ser då ut på följande vis

Source	Frihetsgrader	Kvsum	F-värde	Pr > F	R ²
Modell	6	2.21984E15	520	<.0001	0.6376
Error	1773	1.26147E15			
Corrected Total	1779	3.48131E15			

Vi tar sen och utför ytterligare en Stepwise, backward samt forward för att då få följande slutliga tabell

Source	Frihetsgrader	Kvsum	F-värde	Pr > F	R ²
Modell	4	2.21931E15	780	<.0001	0.6375
Error	1771	1.26199E15			
Corrected Total	1779	3.48131E15			

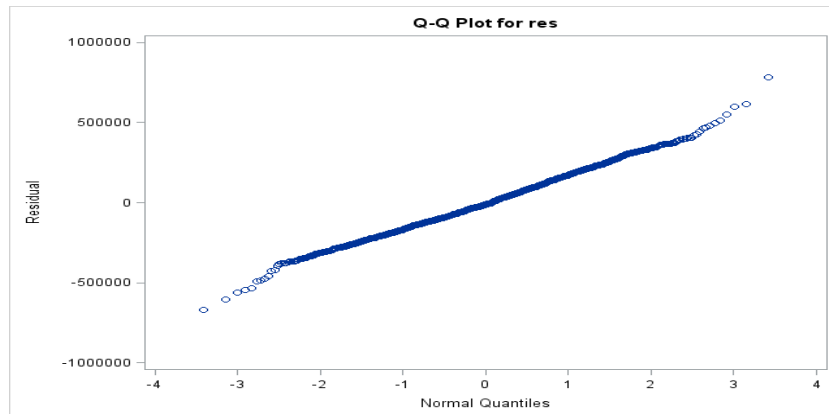
Om vi sen tittar på hur mycket var och en av variablerna bidrar med till förklaringsgraden får vi

Variabel	R ²	Pr > F
DummyC	0.3903	<.0001
Boyta	0.2296	<.0001
DummyB	0.0166	<.0001
Tid	0.0010	0.0249

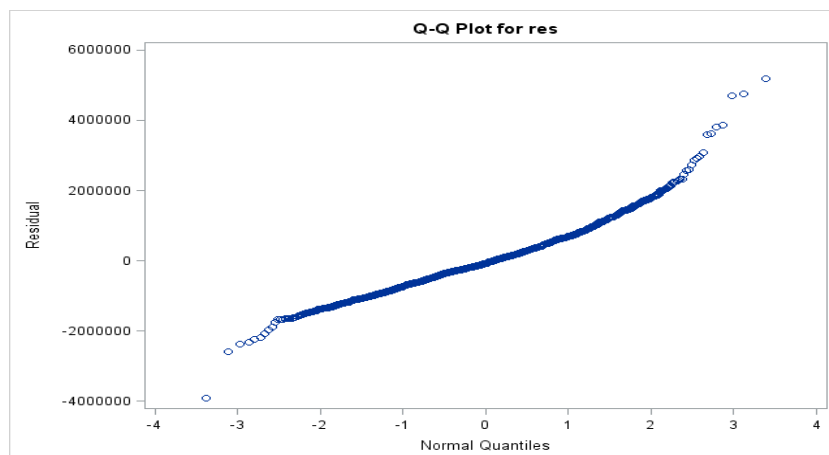
Eftersom vi endast förlorar 0.1% av vår totala förklaringsgrad så väljer vi att ta bort även variabeln tid. Således blir vi kvar med de viktigaste variablerna, områdena B och C och boyta, som tillsammans står för över 60 % av variationen. Vi vet att taxeringsvärdet är viktigt men nu har vi även fått fram andra variabler, som annars tidigare såg relativt betydelselösa ut, att framstå som viktiga. Att bo längre ut från centrala Uppsala tätort kommer att resultera i att man betalar mindre och, som vi också visste från innan, så var boytan också viktig.

5 Residualanalys

Under hela vårt arbete har vi jobbat under antagandet av att våra residualer var normalfördelade med väntevärdet 0 och variansen σ^2 . Vi ska nu avsluta med att se ifall detta har varit ett rimligt antagande. Det vi gör är att plotta residualerna för våra slutliga modeller mot normalfördelnings kvantilen för båda våra data, bostadsrätter och villor. För att vårt antagande ska vara rimligt så ska det vara en så rak linje som möjligt.

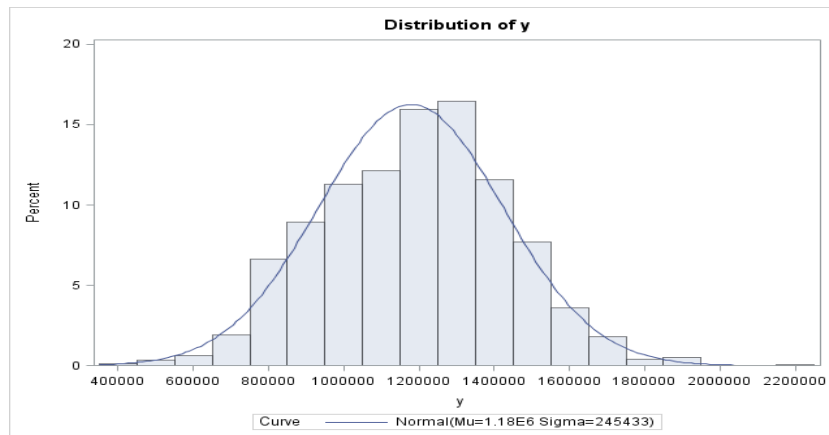


Residualerna i förhållande till normalkvantilen för bostadsrätter

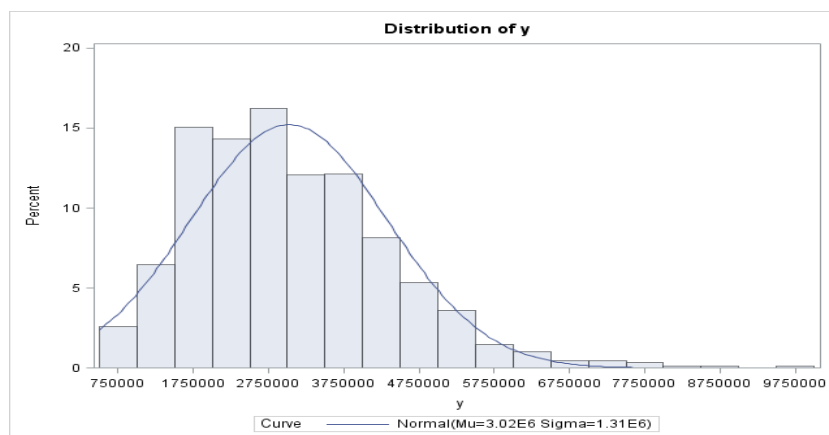


Residualerna i förhållande till normalkvantilen för villor

För bostadsrätterna verkar det som att antagandet av att residualerna är $N(0, \sigma^2)$ men detta tycks inte vara fallet för villorna. Detta beror på att en kvantil- kvantil plott visar kvantilerna av en variabel mot kvantilen av en normal (Gauss) fördelning. Sådana plottar är känsliga för icke normalitet nära ändpunkterna och det är just detta vi ser på plotten för villorna. För att visa detta kan vi ta hjälp av histogram mellan antalet sålda bostäder (i procent) mot priset både för bostadsrätter och för villor.

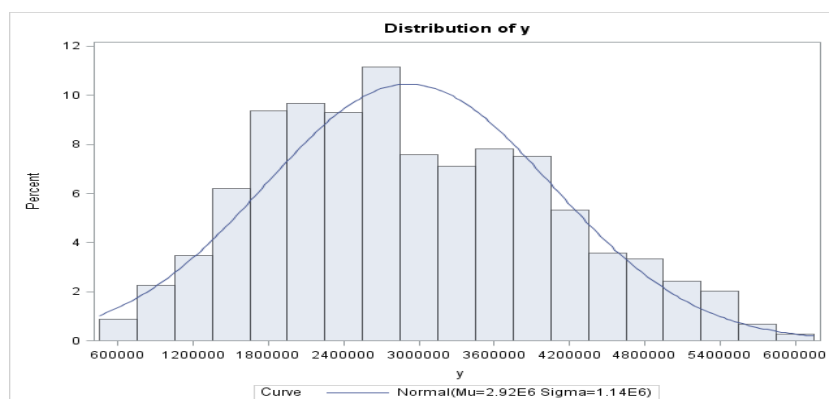


Fördelning av priset för bostadsrätter

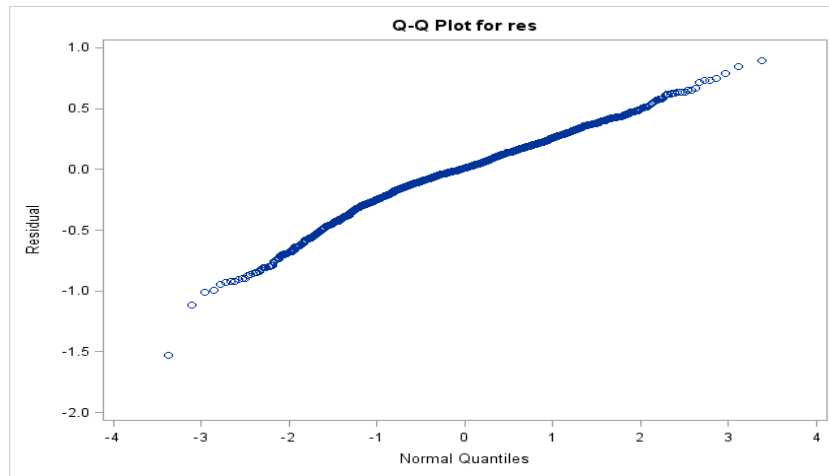


Fördelning av priset för villor

Vi ser att histogrammet är skevt åt vänster vilket var anledningen till vår skeva residualplott. Vi noterar att det inte tycks vara särskilt många observationer som påverkar detta skeva histogram. Med hänsyn till antalet observationer som vi har tar vi oss friheten att be SAS ta bort alla observationer där priset är 6 000 000 kr eller högre. Det handlar då om totalt 46 observationer. Tar vi bort dessa kommer vi då få följande histogram



Detta histogram verkar vara bättre och om vi nu gör ytterligare en residualanalys får vi följande plott



Denna plott tycks nu vara $N(0, \sigma^2)$. Eftersom vi har gjort denna ändring måste vi i våra slutsatser på nästa sida anpassa modellerna så att villorna har 46 färre observationer.

6 Slutsatser

Syftet med uppsatsen var att hitta en modell som bäst förklarade vad som kan ha påverkat det totala försäljningspriset på bostadsrätter och villor i Uppsala kommun. För att bedöma vilken modell som är den mest lämpade i detta syfte så använder vi oss av fakta kring modellens förklaringsgrad och antalet förklarande variabler. Vi vill ha en så hög förklaringsgrad och så få variabler som möjligt. Dock så är inte den modellen som ger högst förklaringsgrad alltid den bästa. Då vi har modeller där vi har färre variabler på bekostnad av en lite lägre förklaringsgrad kanske de är bättre.

Med detta i åtanke är det inte konstigt att vi väljer just modellen

Källa	Frgr	Kvs	F- värde	Pr > F	R ²
Modell	4	5.8749E13	474.5	<.0001	0.4872
Error	1995	6.1845E13			
Corrected Total	1999	1.20520E14			

för bostadsrätter. Även om reporäntan och omx var signifikanta på 5% nivå var för sig så bidrog dessa två variabler endast med dryga 2 % till förklaringsgraden. Därav valet av denna modell som med 4 variabler ger oss en förklaringsgrad på nästan 49 %.

Vi tänker däremot inte likadant då vi ser på villorna. När vi tidigare utförde vår residualanalys tog vi bort 46 observationer. Det påverkar inte vårt val av variabler samt att det påverkar tabellen väldigt lite.

Source	Frihets grader	Kvsum	F-värde	Pr > F	R ²
Modell	5	1.71809E15	1444.89	<.0001	0.8075
Error	1722	4.35333E14			
Corrected Total	1727	2.26172E15			

Detta är inte den slutliga modellen som hjälper oss att förstå varför priset är som den är. Vi utförde en regressionsanalys där vi uteslöt taxeringsvärdet och det är den modellen som kommer att bli vår slutliga modell för villorna.

Source	Frihets grader	Kvsum	F-värde	Pr > F	R ²
Modell	4	1.39754E15	697	<.0001	0.6179
Error	1723	8.64187E14			
Corrected Total	1727	2.26172E15			

7 Diskussion

I vår uppsats valde vi att begränsa oss till att undersöka linjära modeller. Det finns onekligen fler sätt som på ett eller annat sätt kan analysera och förklara vad som påverkat priset. Med detta sagt kan vi inte utesluta att det kan finnas modeller som förklarar bättre vad som påverkat priset.

Att boytan, eller boarean, var så signifikant för bostadsrätter hade vi nästan vetat helt säkert redan innan. Hade vi haft med tomten på villorna hade denna variabel högst troligt varit mer signifikant och förklaringsgraden hade blivit högre. I många fall så är tomten man köper dyrare än huset man bor i så det kan ge en lite missvisande information om man endast ser på den bebodiga ytan.

Det fanns tre variabler som Fastighetsbyrån ville se närmare på ifall de påverkar priset eller inte. Även om omx och reporäntan var signifikanta för bostadsrätterna efter att stepwise regression hade körts så var det endast området som var viktig för både bostadsrätterna och villorna. Det vi kan påpeka här är att det inte fanns lika många observationer i varje område samt att det kanske hade varit en förändring ifall vi hade delat in områdena annorlunda och/eller fler.

Vid en större undersökning finns även möjlighet att utöka antalet variabler. Förslag till några variabler som jag skulle ha kunnat finna intresse för skulle kunna vara storleken på tomten, våning, balkong, skick på lägenhet (gammal eller nyrenoverad), typ av säljare (familj, ensamstående, låg- mellan- eller höginkomsttagare, ung, äldre mm), mäklare (kvinna, man, nyexaminerad, erfaren, ålder), cykelavstånd (till mataffär, centrum, Studenternas IP m.fl) samt media.

Till exempel om vi hade haft en variabel som beskrev medias inverkan så hade den variabeln troligen varit ganska viktig. Av vad Fastighetsbyrån berättade för mig så kunde deras försäljning ibland stå helt stilla eller öka beroende på vad media skrev om bostadsmarknaden. Således skulle det ha varit intressant att ha med media som till exempel en dummy variabel där 1 skulle kunna vara att media uppmärksammat något och 0 att media inte uppmärksammat något angående bostadsmarknaden.

Referenser

- 1 www.fastighetsbyran.se
- 2 Wikipedia, <http://sv.wikipedia.org/wiki/Uppsala> samt <http://sv.wikipedia.org/wiki/Taxeringsv%C3%A4rde>
- 3 www.avanza.se
- 4 Alan Agresti. Categorical Data Analysis, 2nd edition, 2002
- 5 www.maklarstatistik.se, Per Arne Sandegren
- 6 Anna Flodström. "Prediktion av lägenhetspriser i Stockholm – en statistisk undersökning", 2009.
- 7 Sundberg Rolf. Kompendium i Lineära Statistiska Modeller
- 8 Bernard W. Lindgren, Statistical Theory, 4th edition, 1993.