

Regressionsanalys av bostäder i Ekerö kommun

Ellen Karlsson

Kandidatuppsats 2016:10
Matematisk statistik
Juni 2016

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Regressionsanalys av bostäder i Ekerö kommun

Ellen Karlsson*

Juni 2016

Sammanfattning

Syftet med den här uppsatsen är att analysera vilken eller vilka faktorer som påverkar slutpriset av huspriser i Ekerö kommun under 2015. Metoden är, med hjälp av SAS, att göra en regressionsanalys. För att få fram den mest lämpad modell kommer stepwise regression att användas. Grundmodellen består av 15 förklarande variabler. Genom att studera residualplottarna för samtliga variabler logaritmeras några variabler samt responsvariabeln Slutpris. När stepwise regression genomförts återstår sex förklarande variabler i slutmodellen. Under januari och februari 2016 såldes åtta hus i Ekerö kommun. Genom att använda den slutgiltiga modell som fås fram genom stepwise regression, predikteras slutpriser och prediktionsintervall för dessa åtta observationer. Med en konfidensgrad på 95

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.
E-post: elka8049@student.su.se. Handledare: Gudrun Brattström och Jan-Olov Persson.

Innehåll

Sammanfattning.....	0
Inledning.....	3
Syfte och metod.....	3
Teori	3
Regression	3
Enkel linjär regression.....	3
Multipel linjär regression	4
Parameterskattning	4
Stepwise regression	5
Förklaringsgrad	5
Hypotestest	6
VIF-faktor.....	7
Datamaterial	7
Undersökning av datamaterialet.....	8
Genomförande	14
Korrelationsanalys.....	15
Regressionsanalys	17
Prediktionsanalys.....	21
Normalfördelningsantagandet	23
Resultat.....	24
Diskussion	25
Appendix	27
Referenser.....	33

Inledning

Hushållen blir fler och många drömmer om ett hus på landet, men vill ändå ha närheten till storstaden. Ekerö är en växande kommun med nästan 27 000 invånare i slutet av 2015, som erbjuder närhet till natur och vatten men samtidigt närheten till Stockholms innerstad. De flesta av bostäderna i Ekerö kommun består av villor, rad- och kedjehus. Det finns även bostadsrätter, i tätorterna Tappström och Stenhamra.

Syfte och metod

Syftet med den här uppsatsen är att analysera vilken eller vilka faktorer som påverkar slutpriset av huspriser i Ekerö kommun under 2015. Eftersom det finns flera olika faktorer som påverkar slutpriset, är syftet i denna uppsats att ta fram en modell som bäst förklarar varför slutpriser blev som det blev.

Datamaterialet består av 121 hus som såldes 2015. Materialet togs fram manuellt från hemnet.se samt eniro.se. Om ett sålt hus som saknar någon förklarande variabel på hemnet.se, kommer huset att exkluderas från datamaterialet. Metoden är, med hjälp av SAS, att göra en regressionsanalys. För att få fram den mest lämpade modellen kommer stepwise regression att användas.

Under januari och februari 2016 har det sålts 8 hus som har med alla förklarande variabler enligt Hemnet.se i Ekerö kommun. Med hjälp av den slutgiltiga modell som fås fram genom stepwise regression, beräknas både predikterade värdena samt prediktionsintervallet för var och en av dessa åtta observationer.

Teori

Regression

Enligt Rolf Sundbergs kompendium Lineära Statistiska Modeller är regressionsmetoden är en metod som handlar om att undersöka sambandet mellan en responsvariabel och en eller flera förklarande variabler. Under arbetets gång används regression som tillämpar minstakvadratmetoden.

Enkel linjär regression

Modellen för enkel linjär regression definieras som:

$$Y_i = \alpha + \beta x_i + \varepsilon_i \quad , \text{där } i = 1, \dots, N$$

Där Y_i är responsvariabeln, som i denna uppsats är slutpris.

x_i är de förklarande variablerna.

α och β är parametrar och där ε_i är oberoende normalfördelade med väntevärde 0 och varians $\hat{\sigma}^2$.

Modellens väntevärdessfunktion är:

$$\mu_i = \alpha + \beta x_i$$

Väntevärdet $= E(Y_i)$ beror på givna storheter x_i .

Multipel linjär regression

Multipel linjär regression görs då responsvariabeln Y_i beror på två eller flera förklarande variabler. Modellen definieras som:

$$Y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_m x_{mi} + \varepsilon_i, \text{ där } i = 1, \dots, N \quad (1)$$

$Y_i, i = 1, \dots, N$, är en uppsättning av N oberoende stokastiska variabler. Y_i är responsvariabeln, som i detta fall är slutpriset.

$x_{ji} \quad j = 1, \dots, m, i = 1, \dots, N$

ε_i är oberoende normalfördelad med väntevärde 0 och varians σ^2 , som kallas även för försöksfelet.

β_j är parametrar, där $j = 0, \dots, m$ som består av $k = m + 1$ okända parametrar.

Modellens väntevärdessfunktion är:

$$\mu = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m$$

Modellen (1) kan skrivas på matrisform:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{11} & \dots & x_{m1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1N} & \dots & x_{mN} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}, \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

Parameterskattning

MK-metoden är en metod som används för att skatta parametrarna i en modell.

För att hitta skattningarna för parametrarna innebär minsta kvadratmetoden att minimera avståndet mellan observationerna och deras väntevärde. Det betyder att vi ska hitta $\boldsymbol{\beta}$ för som ska minimera:

$$\sum_{i=0}^N (y_i - \sum_{j=0}^m x_{ji} \beta_j)^2, \text{ där } x_{0i} = 1 \text{ för alla } i \quad (2)$$

(2) kan skrivas om som:

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 \quad (3)$$

Att minimera (3) betyder att minimera $\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|$, vars parametervektor är:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (4)$$

$\hat{\boldsymbol{\beta}}$ ser vi som en stokastisk variabel och som beror på $\mathbf{Y}$. Skriver (4) som:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

Vi har att $\hat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1})$

Vi har att $\widehat{\sigma^2}$ är skattningen av σ^2 :

$$\widehat{\sigma^2} = \frac{1}{N-k} \|\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\|^2$$

Prediktionsintervall

Prediktionsintervall är ett intervall med bestämd sannolikhet innehåller det utfall som ska predikteras. Prediktionsgränser som innehåller kommande Y-värden för $x = x_{10}, \dots, x_{m0}$ är:

$$\hat{\beta}_0 + \hat{\beta}_1 x_{10} + \dots + \hat{\beta}_m x_{m0} \pm \hat{\sigma} \sqrt{1 + \frac{1}{N} + \dot{x}^T \mathbf{W} \dot{x}}, \text{ där } \dot{x} \text{ är en uppsättning nya värden och där}$$

$$\mathbf{W} = (\mathbf{U}^T \mathbf{U})^{-1}, \text{ där}$$

$$\mathbf{U} = \begin{bmatrix} x_{11} - \bar{x}_{1\cdot} & x_{21} - \bar{x}_{2\cdot} & \dots & x_{m1} - \bar{x}_{m\cdot} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1N} - \bar{x}_{1\cdot} & x_{2N} - \bar{x}_{2\cdot} & \dots & x_{mN} - \bar{x}_{m\cdot} \end{bmatrix}$$

Stepwise regression

Stepwise regression går ut på att man vid varje steg kontrollerar för varje variabel som införts i modellen ifall de ger signifikant utslag vid test av hypotesen, att den motsvarande parametern är noll. De variabler som inte ger ett signifikant utslag elimineras bort.

Förklaringsgrad

Förklaringsgrad definieras som den andel av den totala variationen som modellen förklarar. Förklaringsgraden det vanligaste anpassningsmättet i samband med linjära modeller. Den betecknas R^2

$$R^2 = \frac{KVS(regression)}{KVS(total)} = 1 - \frac{KVS(residual)}{KVS(total)}$$

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

$$KVS(residual) = \sum (y_i - \hat{y}_i)^2, \text{ där } \hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \dots + \hat{\beta}_m x_{mi}$$

$$KVS(regression) = \sum (\hat{y}_i - \bar{\hat{y}})^2, \text{ där } \bar{\hat{y}} \text{ är medelvärdet av } \hat{y}$$

Om man lägger till en x-variabel kommer förklaringsgraden alltid att öka. För att studera ifall det lönar sig att lägga till ytterligare en variabel är det bättre att studera ifall $\hat{\sigma}^2$ minskar. Om den minskar kan man tolka det som att det är mindre grad slump kvar i vår modell. Detta mått kallas för Adjusted R^2 , och betecknas R_{adj}^2

$$R_{adj}^2 = 1 - \hat{\sigma}^2 / \hat{\sigma}_0^2$$

Där $\hat{\sigma}_0^2$ är:

$$\hat{\sigma}_0^2 = \frac{KVS(\text{totalt})}{N-1}$$

Hypotestest

Hypotestesten i denna studie kommer att undersökas om parametrarna i modellen är signifikant skild från noll eller inte. Det betyder alltså om motsvarande variabel har en inverkan på responsvariabeln eller inte. Det kan uttryckas:

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0 \quad (5)$$

Där $j = 1, \dots, m$. För att testa hypotesen (5) används t-test.

Kan genom testvariabelmetoden testa ifall nollhypotesen kan förkastas eller inte.

Nollhypotesen förkastas om resultatet är signifikant. Vi utnyttjar följande och testa alternativhypotesen H_1 :

$$T = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma} \sqrt{((\mathbf{X}^T \mathbf{X})^{-1})_{jj}}} \sim t(N - k) \quad (6)$$

Vi har att $\hat{\beta}_j \sim N(\beta_j, \sigma((\mathbf{X}^T \mathbf{X})^{-1})_{jj})$, där $((\mathbf{X}^T \mathbf{X})^{-1})_{jj}$ är det j-te diagonalelementet i matrisen $(\mathbf{X}^T \mathbf{X})^{-1}$.

Vi låter $\beta_j = 0$. Med hjälp av (6) kan hypotesen (5) testas. I den här rapporten väljer vi signifikansnivån $\alpha = 0,05$. Vi förkastar H_0 om $|T| > t_{0,025}(N - k)$

VIF-faktor

VIF står för Variance Inflation Factor. Enligt Rolf Sundbergs kompendium Lineära Statistiska Modeller uttrycker VIF-faktorn hur mycket större varians β_j har i närvaro av de andra regressionstermerna, än vad den skulle haft om x_j hade varit ensam. VIF-faktorn beräknas som:

$$VIF = \frac{1}{1 - R_j^2}$$

Där R_j^2 står för den förklaringsgrad som anger hur mycket av variationen i x_j som förklaras av de andra x-variablerna.

Om $VIF = 1$ så är $R_j^2 = 0$, vilket betyder att x_j är ortogonal mot de övriga x-variablerna. Hur hög VIF-faktorn bör vara är svårt att avgöra, men bör sättas en gräns mellan 5 och 10. Om VIF-faktor är högre än den gränsen så är den förklarande variabeln starkt kollinjär. I denna uppsats sätts gränsen vid 5. Om en variabel har en stor VIF-faktor kan en lösning vara att exkludera variabeln från modellen.

Datamaterial

Datamaterialet är framtaget från hemnet.se och innefattar 121 observationer som såldes under 2015. Dessa togs manuellt fram från hemnet.se. Det fanns vissa villor och radhus/kedjehus på hemnet.se som sålts under denna period där det fattades någon eller några av de förklarande variablerna. De har således exkluderats bort. Endast villor och kedjehus/radhus som sålts under 2015 kommer att tas med i denna studie. Det innebär att slutpriset på bostadsrätter, fritidshus samt gårdar för samma period inte tas med i datamaterialet.

Slutpris agerar som responsvariabel, som beskriver det belopp som huset sålts för. Variablerna skolavstånd, vattenavstånd samt busshållplatsavstånd fanns inte med som kategori på hemnet.se, utan togs fram manuellt på eniro.se. För att ta reda på vatten- skol- och busshållplatsavståndet användes funktionerna vägbeskrivning samt avståndsuträkning från en punkt till en annan. När skol- och busshållplatsavståndet räknades ut användes vägbeskrivningen, och för avståndet till vattnet räknades fågelvägen ut.

Materialet består av följande variabler:

Boarea

Husets beboliga area i m^2 .

Biarea

Exempelvis garage, soprum, pannrum i m^2 .

Tomt

Husets tomtarea i m^2 .

Byggår

Husets byggår i årtal.

Antal rum

Antalet rum som husets har.

Bostadstyp

Beskriver ifall det sålda huset antingen är en villa, vinterbonat fritidshus eller radhus/kedjehus.

Område

Beskriver i vilket område det sålda huset befinner sig i. Delas in i fyra områden: Adelsö, Munsö, Ekerö och Färingsö.

Månad

Vilken månad under år 2015 som huset köptes.

Vatten

Avståndet från huset till Mälaren.

Skola

Avståndet från huset till närmaste skola.

Busshållplats

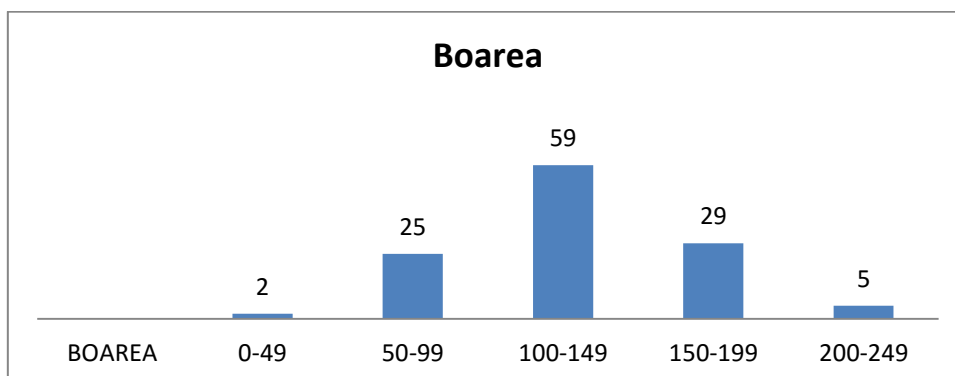
Avståndet från huset till närmaste busshållplats.

Undersökning av datamaterialet

Varje förklarande variabel undersöks nu var för sig. Vid vissa förklarande variabler har en uppdelning av observationerna gjorts, för att få en överblick hur materialet ser ut. Denna uppdelning kommer inte vara med när regressionsanalysen genomförs.

BOAREA

Till att börja med, delas alla observationer in i fem grupper, (0-49 kvm, 50-99 kvm, 100-149 kvm, 150-199 kvm och 200-249 kvm), för att få en överblick hur materialet ser ut.

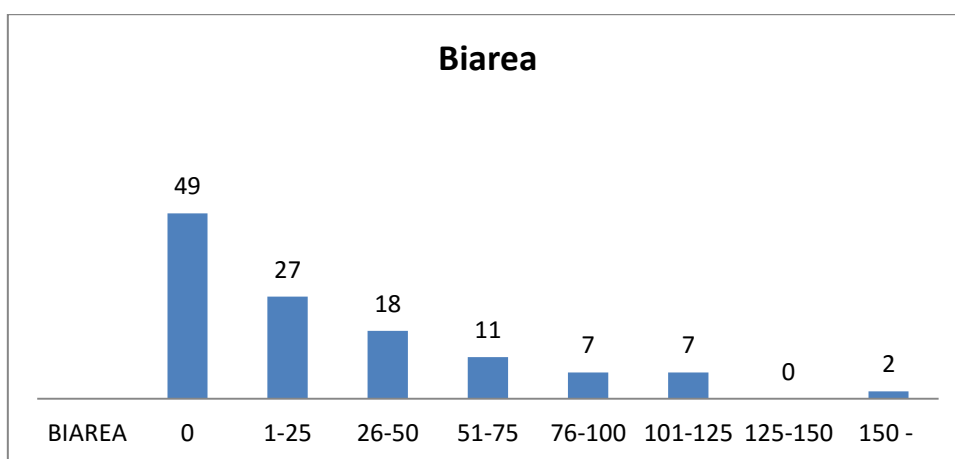


Figur 1. Uppdelning av boarean för alla sålda hus i Ekerö kommun år 2015.

De flesta sålda hus har en boarea mellan 100- 149 kvm. Sedan finns det två hus som endast hade boarea mellan 0-49 samt fem hus med en boarea på 200-249 kvm. Medelboarean ligger på 126 kvm.

BIAREA

Biarean delas in i olika grupper (0 kvm, 1-25 kvm, 26-50 kvm och så vidare till sist gruppen med en biarea över 150 kvm), för att få en överblick hur över biarean.

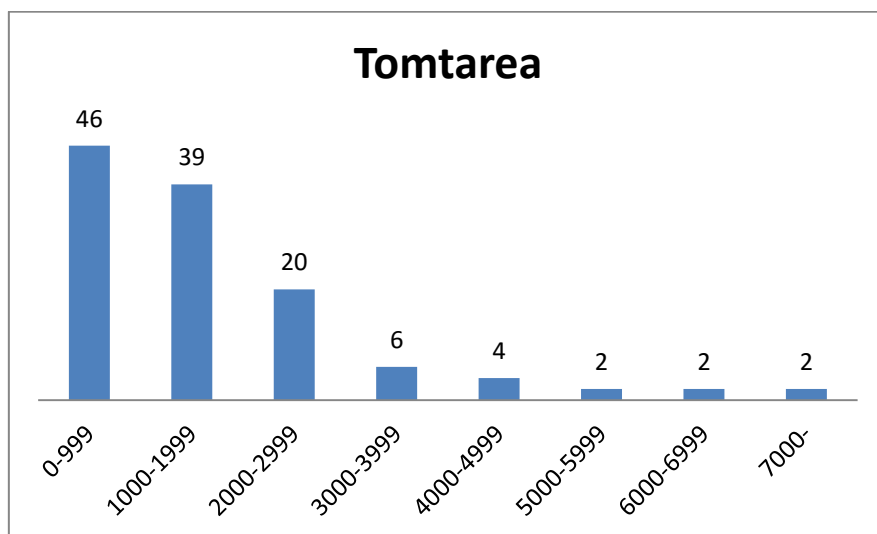


Figur 2. Uppdelningen över biarea för alla sålda hus i Ekerö kommun i 2015.

49 sålda hus har en biarea på 0 kvm. Ju större biarea, desto färre hus har sålts. Det finns några få hus som har en biarea större än 150 kvm. Det finns två avvikande observationer med en biarea på 200 kvm respektive 464 kvm. Medelbiarean ligger på 32 kvm.

TOMTAREA

Observationerna för tomtarean delas in i 8 tomtgrupper; 0 – 999 kvm, 1000 – 1999 kvm, 2000 – 2999 kvm och så vidare.

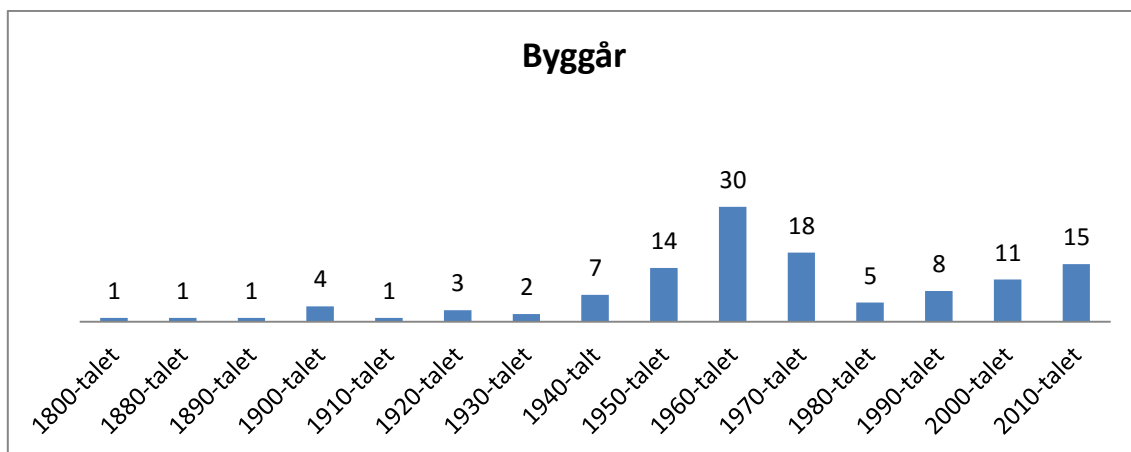


Figur 3 Uppdelning av tomtarea för alla sålda hus i Ekerö kommun år 2015.

De flesta sålda hus en tomtarea mellan 0 - 999 kvm men att det är väldigt få hus som har en tomtarea större än 3000. Det finns två utstickande observationer med en tomtarea på 28 060 kvm och 13 650 kvm. Medeltomtarea är 1 895 kvm.

BYGGÅR

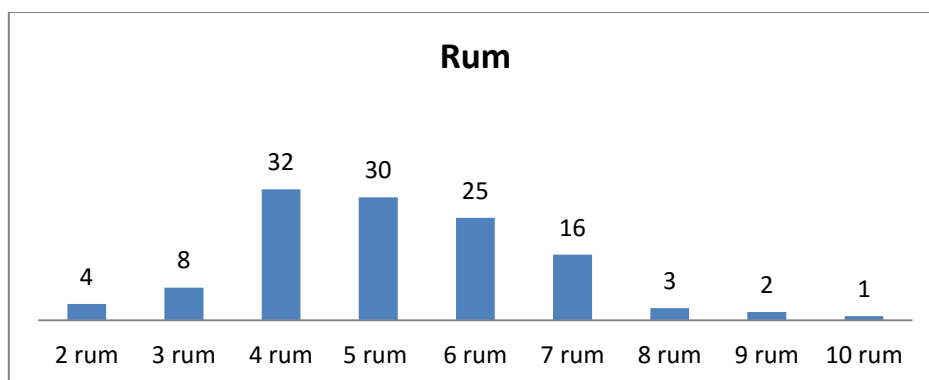
Det äldsta huset byggdes 1800 medan det nyaste byggs 2016.



Figur 4 Uppdelning över byggår för varje hus som sålts i Ekerö kommun år 2015.

De flesta hus som såldes byggdes under 1960-talet. I figur 4 kan vi studera att antalet hus som är byggdes innan 1950-talet är få i jämförelse med senare årtionden.

ANTAL RUM



Figur 5 Uppdelning av antal rum för alla sålda hus i Ekerö kommun år 2015.

Den största gruppen av antal rum som sålts har antingen 4, 5 eller 6 rum. Det få sålda hus som ha 8, 9 och 10 rum. Medelvärde för antal rum är 5 rum.

BOSTADSTYP

Bostadstyp	Antal
Villa	27
Friliggande villa	57
Vinterbonat fritidshus	11
Radhus/Kedjehus	26

Tabell 1. Antal sålda villor, friliggande villa, vinterbonad fritidshus och radhus/kedjehus.

Det har sålt flest friliggande villor under perioden 2015 i Ekerö kommun och minst antal vinterbonade fritidshus.

Skillnaden mellan villa och friliggande villa är enligt hemnet.se olika begrepp som mäklaren använder sig av.

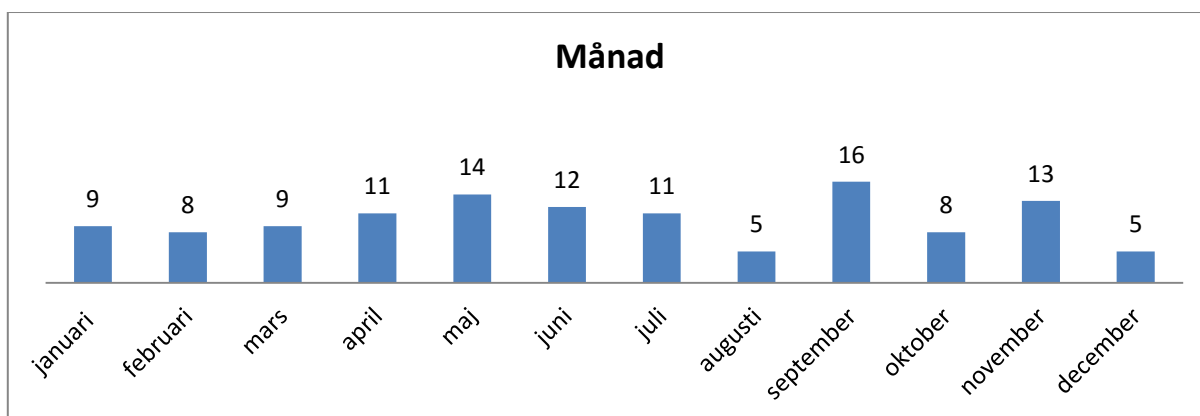
OMRÅDE

Område	Antal
Ekerö	60
Färingsö	45
Munsö	12
Adelsö	4

Tabell 2. Antal sålda hus i områdena Ekerö, Färingsö, Munsö och Adelsö.

De flesta hus som såldes tillhörde området Ekerö och sedan Färingsö. Endast 4 hus såldes under samma period på Adelsö.

MÅNAD

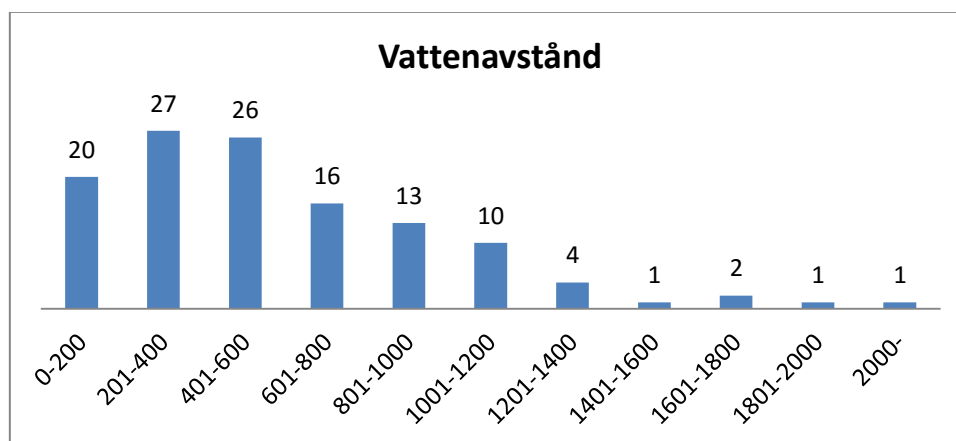


Figur 6 Månadsuppdelning för alla sålda hus i Ekerö kommun år 2015.

Enligt figur 6 ovan var det under september 2015 som flest antal hus såldes i Ekerö kommun. En ökning av antalet sålda hus sker från februari till maj, för att sedan avta under sommarmånaderna. Det har sålts lägst antal hus under augusti och december med 5 sålda hus var. För de övriga månaderna såldes det mellan 8 till 12 stycken hus per månad.

VATTENAVSTÅND

Eftersom vattenavståndet inte fanns med som variabel på hemnet.se, mättes avståndet till Mälaren med hjälp av eniro.se. Samtliga observationer delas in efter avståndet till Mälaren.

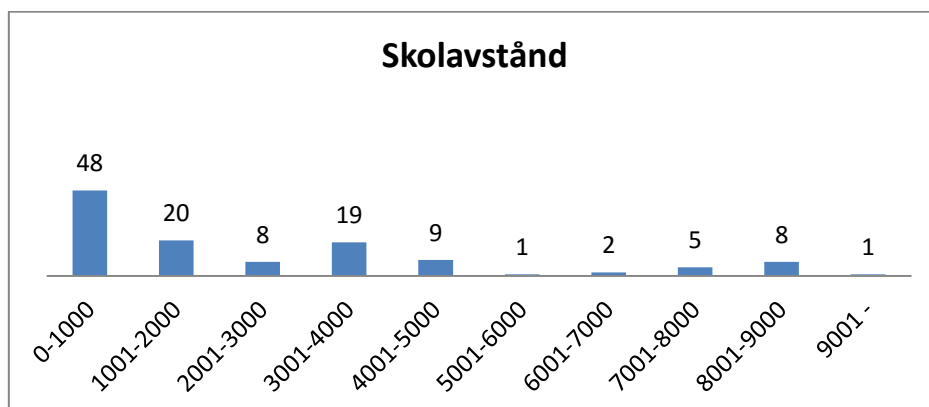


Figur 7. Vattenavståndet för alla sålda hus i Ekerö kommun år 2015.

Det som sticker ut i diagrammet ovan är att ju längre bort från Mälaren, desto färre hus har sålts. Det har sålts flest antal hus med 201-400 meter och 401-600 meter till Mälaren. Medelavståndet till Mälaren är 594 meter.

SKOLAVSTÅND

Avståndet till närmaste grundskola för samtliga observationer delas in i 10 grupper, 0-1000 meter, 1001-2000 meter och så vidare fram till 9001 och längre.

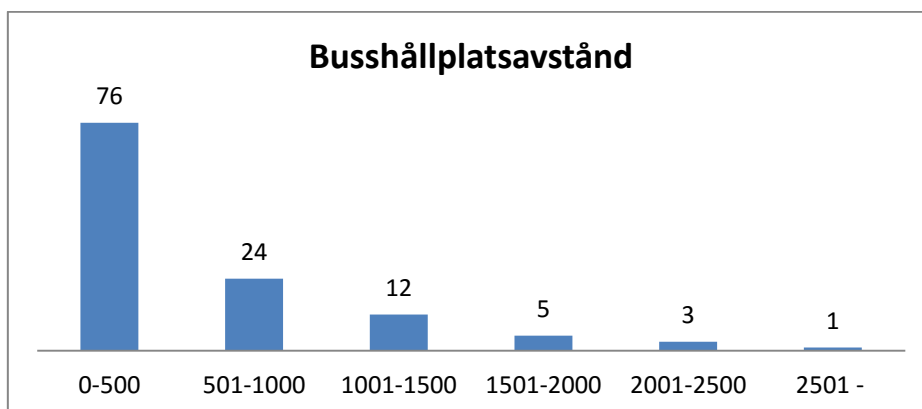


Figur 8. Skolavståndet för alla sålda hus i Ekerö kommun år 2015.

Genom att studera diagrammet ovan kan det konstateras att de flesta hus som såldes har ett skolavstånd mindre än 1 000 meter. Det har sålts 19 och 20 hus med ett skolavstånd på 1001-2000 meter respektive 3001 – 4000 meter. I de resterande grupperna är antalet sålda hus få. Det längsta avståndet till närmaste skola är 9100 meter. Medelavståndet är 2614 meter.

BUSSAVSTÅND

Det är även intressant ifall avståndet till bussen påverkar slutpriset. Observationerna delas in efter avståndet till närmaste busshållplats. Indelningen är 0-500 meter, 501-1000 meter och så vidare fram till sista indelningen som är 2501 meter och längre.



Figur 9. Busshållplatsavstånd för samtliga hus i Ekerö kommun som sålts 2015.

De flesta hus har ett avstånd till närmaste busshållplats på mindre än 500 meter. Det finns ett hus som avviker som hade ett avstånd på över 2500 meter. Medelavståndet ligger på 588 meter

Genomförande

Innan regressionsanalysen genomförs studeras först vilka förklarande variabler som kommer att användas i vår regressionsanalys. Dessa variabler tas med:

Slutpris	Responsvariabel
Boarea	Kontinuerlig variabel
Biarea	Kontinuerlig variabel
Tomtarea	Kontinuerlig variabel
Byggår	Kontinuerlig variabel
Rum	Kontinuerlig variabel
Bostadstyp	Kategorisk variabel
Område	Kategorisk variabel
Månad	Kontinuerlig variabel
Vattenavstånd	Kontinuerlig variabel
Skolavstånd	Kontinuerlig variabel
Busshållplatsavstånd	Kontinuerlig variabel

Tabell 3: De förklarande variablerna.

Det är totalt 11 förklarande variabler, varav två är kategoriska variabler (område och bostadstyp).

Område: Variabeln Område delas in i tre variabler; Ekerö, Färingsö och AdelsöMunsö. Anledningen till ett gemensamt område för Adelsö och Munsö är för att områdena ligger nära varandra och för att de båda har en gemensam grundskola samt mataffär. Adelsö är mestadels sommarstugeområde, men människorna som permanent bor på Adelsö måste åka till Munsö för att handla mat, eller gå i skolan.

Ekerö är referenskategori.

Dummyvariabel	Förklaring
Färingsö	1 om husen köptes i område Färingsö, resten 0
AdelsöMunsö	1 om husen köptes i område Adelsö och Munsö, resten 0

Tabell 4. Dummyvariablerna Färingsö och AdelsöMunsö.

Bostadstyp: Som tidigare nämnts studeras endast villor, friliggande villa, vinterbonat fritidshus samt radhus/kedjehus. Eftersom villa och friliggande villa betyder samma sak, slås dessa ihop till en variabel, Villa.

Villa agerar som referensvariabel.

Dummyvariabel	Förklaring
Vinterbonat	1 om husen tillhör kategorik Vinterbonat fritidshus, resten 0.
RadKedjehus	1 om husen är radhus/kedjehus, resten 0.

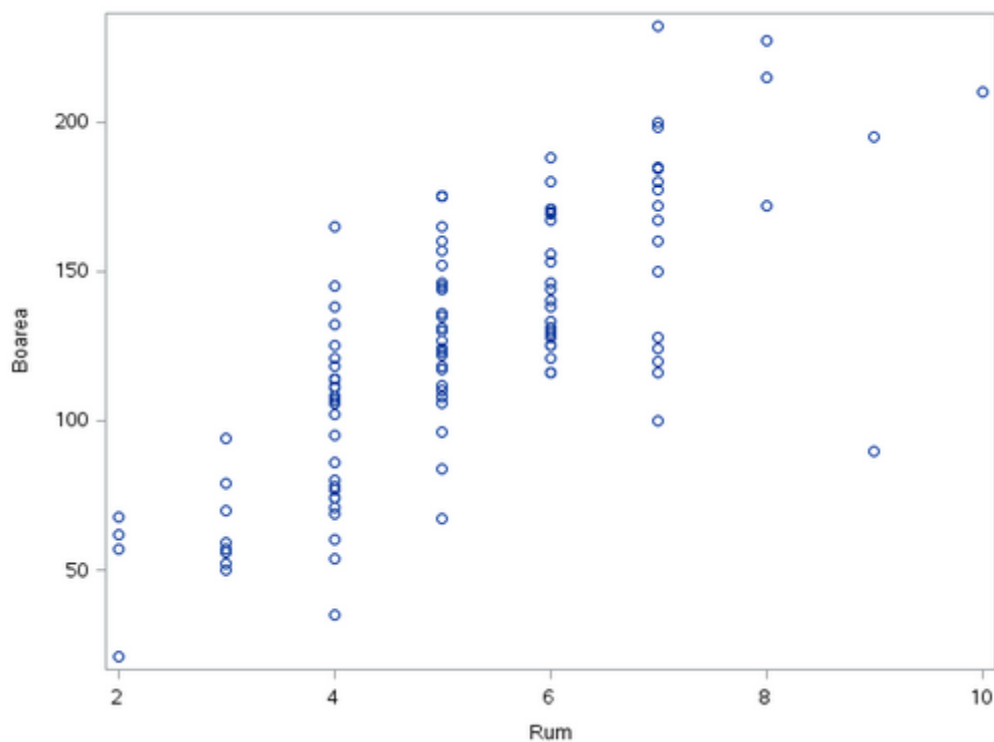
Tabell 5. Dummyvariablerna Vinterbonat och RadKedjehus.

När alla förklarande variabler är färdigstuderade, studeras nu korrelationen mellan de förklarande variablerna att studeras.

Korrelationsanalys

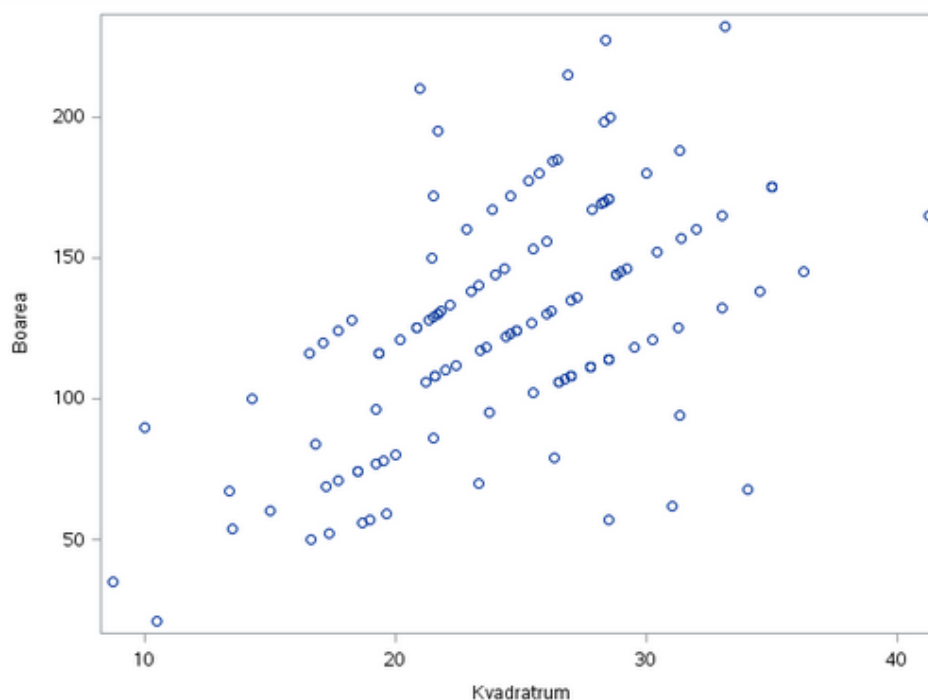
Anledningen till att studera korrelationen mellan de förklarande variablerna, är för att studera om några variabler beror på varandra. Det kan exempelvis vara att boarean korrelerar med antalet rum, ju större boyta desto fler rum. Ifall det finns en korrelation mellan två variabler, exempelvis boyta och antalet rum, kan ena variabeln uteslutas. En annan lösning är att dividera boytan med antalet rum och införa kvoten som en ny variabel.

Genom att plotta boarean mot antalet rum mot varandra visas en korrelation. Se figur 10 nedan.



Figur 10 Plott mellan de förklarande variablerna Boarea och Rum

Eftersom det finns en korrelation mellan variablerna Boarea och Rum, skapas en ny variabel genom att dividera Boarea med Rum. Efter införande av den nya variabel, kvadratmeter per rum (Kvadratrums), kan variabeln rum tas bort. Nedan visas plotten mellan Boarea och Kvadratrums. Vi kan se att korrelationen har minskat.



Figur 11 Plott mellan Boarea och Kvadraturum

Det är intressant att se ifall det finns någon korrelation mellan de övriga förklarande variablerna. I figur 16 i appendix ser vi korrelationen mellan samtliga förklarande variabler. Som visas finns ingen korrelation mellan någon förklarande variabel.

Nedan studeras VIF-faktorn (höga kolumnen) för samtliga förklarande variabler.

Variabel	VIF-faktor
Boarea	1,857
Kvadraturum	1,926
Tomt	1,419
Byggår	1,228
Vinterbonad	2,347
RadKedjehus	1,753
AdelsöMunsö	2,784
Färingsö	1,830
Månad	1,313
Vatten	1,484
Skola	2,522
Busshållplats	1,681
Biarea	1,201

Tabell 6. VIF-faktor för förklarade variablerna.

Som visas ovan, är VIF-faktorn för de förklarande variablerna inte större än 3. Kan konstatera att ingen av de förklarande variablerna är approximativt linjärt oberoende av varandra.

Regressionsanalys

En enkel linjär regression genomförs av Slutpris på varje förklarande variabel. Det finns en misstanke att boarea är den variabel som påverkar slutpriset mest. Vi får följande förklaringsgrad vid enkel regression av följande förklarande variabler.

Responsvariabel	Förklarande variabel	Förklaringsgrad
Slutpris	Boarea	0,3106
Slutpris	Biarea	0,0069
Slutpris	Tomtarea	0,0039
Slutpris	Byggår	0,0395
Slutpris	Kvadratrum	0,1585
Slutpris	Vattenavstånd	0,0831
Slutpris	Skolavstånd	0,0541
Slutpris	Busshållplatsavstånd	0,0038
Slutpris	Månad	0,0134

Tabell 7. Förklaringsgrad för de förklarande variablerna och responsvariabeln Slutpris vid enkel regression.

Boarea är den förklarande variabel som har högst förklaringsgrad på 31,06 %. Tomtarea och Vattenavstånd är de förklarande variabler som har lägst förklaringsgrad. Att förklaringsgraden för boarea är högst kan bero på att Boarea är den förklarande variabel som mest påverkar slutpriset, medan busshållplatsavstånd och tomtarea förklarar slutpriset minst.

För att få en så bra modell som möjligt kommer residualplottarna för varje förklarande variabel att studeras. Ifall en residualplott visar på något samband eller mönster, kan det vara en lösning att transformera den förklarande variabeln. Detta gäller även för responsvariabeln.

Multipel regression

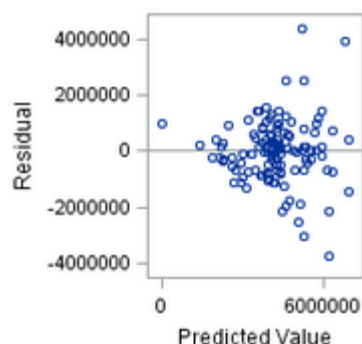
En multipel regression genomförs på följande modell, som vi kallar för modell 1:

$$\begin{aligned} \text{Slutpris} = & \beta_0 + \beta_1 \text{Boarea} + \beta_2 \text{Biarea} + \beta_3 \text{Tomt} + \beta_4 \text{Byggår} + \beta_5 \text{Kvadratrum} \\ & + \beta_6 \text{Vinterbonad} + \beta_7 \text{RadKedjehus} + \beta_8 \text{Färingsö} + \beta_9 \text{AdelsöMunsö} \\ & + \beta_{10} \text{Månad} + \beta_{11} \text{Vatten} + \beta_{12} \text{Skola} + \beta_{13} \text{Busshållplats} \end{aligned}$$

Modell	R^2	R^2_{adj}
Modell 1	52,54 %	46,77 %

Tabell 8. R^2 och R^2_{adj} av modell 1 vid multipel regression

Modell 1 har en låg förklaringsgrad på 52,54 %.



Figur 12. Residualplot mot predikterade värden för responsvariabeln Slutpris.

I figur 12 van kan vi studera residualerna mot de predikterade värdena. Det tyder att residualplotten inte är jämnt utspridda. I figur 19 i appendix kan samtliga plottar av residualerna mot de förklarande variablerna studeras när vi genomförde multipel regression av modell 1. När hela datamaterialet undersöktes kunde avvikande observationer upptäckas. Dessa avvikande observationer kan urskiljas i de enskilda residualplottarna.

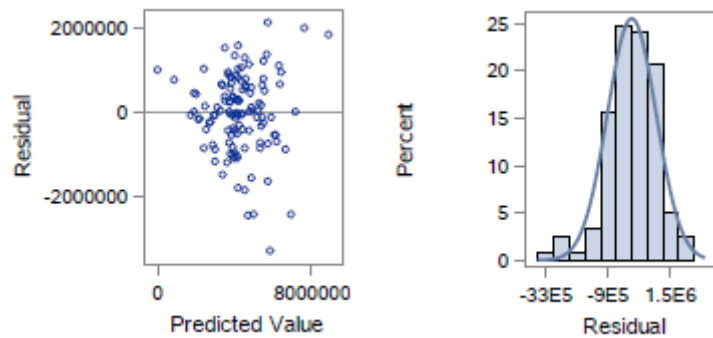
Målet är att få en så spridd residualplott som möjligt. Genom att studera residualplottarna för varje enskild variabel kan urskiljas att Boarea och Kvadratrums residualplottar inte är homogena, det vill säga jämt spridd. Inte heller variabeln Vattenstånd är helt spridd. Det är svårt att tyda residualplotten för Tomt på grund av de två avvikande observationerna. Dock misstänks att tomtarean påverka priset, ju större tomt, desto högre pris. Det gäller även för variabeln Boarea. Även vattenavståndet kan ha en påverkan på slutpriset, ju närmare vatten, desto högre pris. Därför logaritmeras Boarea, Tomt, Vatten, Busshållplats men även Kvadratrums på grund av att residualplotten inte är homogen. De övriga variablerna görs ingen förändring.

Den nya modellen, som vi kallar för modell 2, att undersöka är:

$$\begin{aligned} \text{Slutpris} = & \beta_0 + \beta_1 \text{LogBoarea} + \beta_2 \text{Biarea} + \beta_3 \text{LogTomt} + \beta_4 \text{Byggår} + \beta_5 \text{LogKvadratrums} \\ & + \beta_6 \text{Vinterbonad} + \beta_7 \text{RadKedjehus} + \beta_8 \text{Färingsö} + \beta_9 \text{AdelsöMunsö} \\ & + \beta_{10} \text{Månad} + \beta_{11} \text{LogVatten} + \beta_{12} \text{Skola} + \beta_{13} \text{LogBusshållplats} \end{aligned}$$

Modell	R^2	R^2_{adj}
Modell 2	66,76 %	62,72 %

Tabell 9. R^2 och R^2_{adj} av modell 2 vid multipel regression



Figur 13. Residualplott mot predikterade värden och histogram för modell 2.

Genom att transformera de fyra variablerna ökar både R^2 och R^2_{adj} , vilket tyder att denna modell är mer lämplig.

I figur 20 i appendix kan samtliga plottar av residualerna mot de förklarande variablerna studeras. De fyra variablerna som transformerades har en mer utspridd residualplott i jämförelse med innan de transformerades. Däremot kan visas i figur 13 ovan att residualplotten mot de predikterade värdena för hela modellen inte är jämt utspridd.

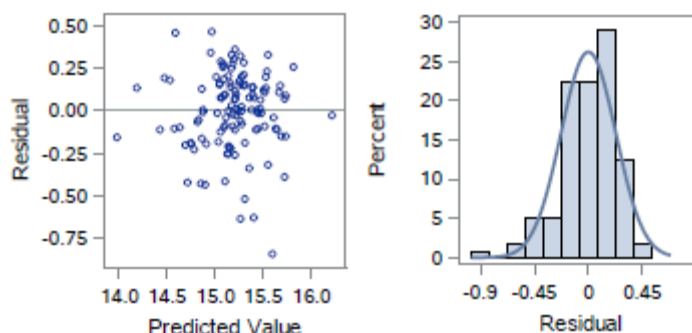
För att få så bra modell som möjligt vill vi att residualplotten för hela modellen också ska vara homogen. För att förbättra modellen och få en mer homogen residualplott, är en lösning att transformera responsvariabeln Slutpris.

Vi väljer att transformera responsvariabeln Slutpris och genomför en multipel regression på följande modell, som vi döper till Modell 3:

$$\begin{aligned} \text{LogSlutpris} = & \beta_0 + \beta_1 \text{LogBoarea} + \beta_2 \text{Biarea} + \beta_3 \text{LogTomt} + \beta_4 \text{Byggår} \\ & + \beta_5 \text{LogKvadratrums} + \beta_6 \text{Vinterbonad} + \beta_7 \text{RadKedjehus} + \beta_8 \text{Färingsö} \\ & + \beta_9 \text{AdelsöMunsö} + \beta_{10} \text{Månad} + \beta_{11} \text{LogVatten} + \beta_{12} \text{Skola} \\ & + \beta_{13} \text{LogBusshållplats} \end{aligned}$$

Modell	R^2	R^2_{adj}
Modell 3	68,18%	64,31 %

Tabell 10. R^2 och R^2_{adj} av modell 3 vid multipel regression



Figur 14. Residualplott mot de predikterade värdena och histogram

Residualplotten mot de predikterade värdena är mer spridd, se figur 14 ovan. I figur 21 i Appendix kan samtliga plottar av residualerna mot de förklarande variablerna studeras.

Frågan är nu ifall det finns något samspel mellan de förklarande variablerna. Om det antas att Boarean är den förklarande variabel som påverkar Slutpriset mest, och därför kan det finnas något samspel mellan Boarea och någon eller några av de andra förklarande variablerna.

För att studera samspel läggs en ytterligare variabel till i modellen. Om det antas exempelvis samspel mellan Boarea och Tomtarea, införs en ny variabel Boarea*Tomtarea i modellen. Efter att ha studerat ifall det finns något samspel mellan Boarea och de förklarande variablerna, kan konstateras att förklaringsgraden ökar maximalt med 0.04 procentenheter. R^2_{adj} ökade inte heller markant, och det konstateras därför att det inte finns tillräckligt stort samspel mellan Boarea och de övriga förklarande variablerna.

Stepwise regression

En stepwise regression genomförs av modell 3 för att få fram den slutgiltiga modellen som bäst beskriver slutpriset på fastigheter i Ekerö kommun.

Residualplottarna för samtliga variabler för denna modell är homogena, se figur 22 i Appendix.

$$R^2 = 66,8 \%, \quad R^2_{adj} = 65,05 \%, \quad \hat{\sigma}^2 = 0,0574, \quad \hat{\sigma} = 0,24$$

$$\text{LogSlutpris} = \beta_0 + \beta_1 \text{LogBoarea} + \beta_2 \text{LogTomt} + \beta_3 \text{AdelsoMunsö} + \beta_4 \text{Färingsö} + \beta_5 \text{LogVatten} + \beta_6 \text{Skola} + \epsilon \quad (7)$$

	Parameterskattning	Standardfel	p-värde
Intercept	12,85764	0,34042	< 0,0001
LogBoarea	0,52137	0,05982	< 0,0001
LogTomt	0,09296	0,02402	0,0002
AdelsöMunsö	– 0,33847	0,09375	0,0005
Färingsö	– 0,14810	0,09375	0,0083
LogVatten	– 0,10508	0,01731	<0,0001
Skola	– 0,00002878	0,00001183	0,0165

Tabell 11. Parameterskattning, Standardfel och p-värde för slutgiltiga modellen efter stepwise regression

Prediktionsanalys

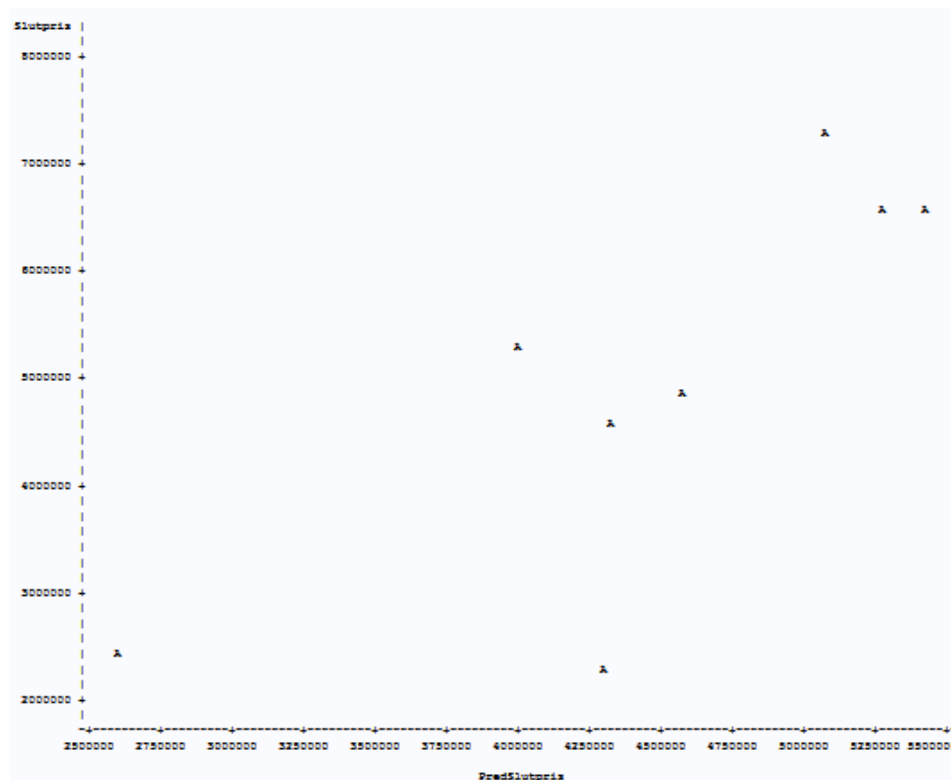
Under januari och februari 2016 såldes det 8 hus i Ekerö kommun. Först tas all data fram för dessa åtta observationer. Detta görs på samma sätt som för de observationerna under år 2015 och används av samma förklarande variabler som tidigare. Genom att använda (7) som är den slutgiltiga modellen som fås fram genom stepwise regression, predikteras dessa åtta värden med hjälp av modellen.

För de åtta observationerna fås följande predikterade värden:

Observation	Predikterat Slutpris, (i kr)	Predikterat LogSlutpris	Verkligt Slutpris, (kr)	Verkligt LogSlutpris	Begärt Pris, (i kr)
1	4 310 665	15,277	2 245 000	14,624	5 445 000
2	5 263 335	15,476	6 570 000	15,698	5 875 000
3	2 592 309	14,768	2 400 000	14,691	2 375 000
4	5 418 720	15,505	6 600 000	15,703	6 195 000
5	3 996 418	15,201	5 310 000	15,485	4 900 000
6	4 579 183	15,337	4 790 000	15,382	4 195 000
7	4 324 920	15,280	4 600 000	15,342	4 700 000
8	5 074 647	15,440	7 260 000	15,798	7 250 000

Tabell 12. Prediktionsvärde och verkliga värdet för observationerna under 2016.

De predikterade värdena skiljer sig en hel del från de verkliga värdena. 2 stycken predikterade värdena är över de verkliga värdena. Däremot visas att första observationen hade ett begärt pris mycket högre än det slutgiltiga priset.



Figur 15. Plott av Slutpris på Predikterat Slutpris

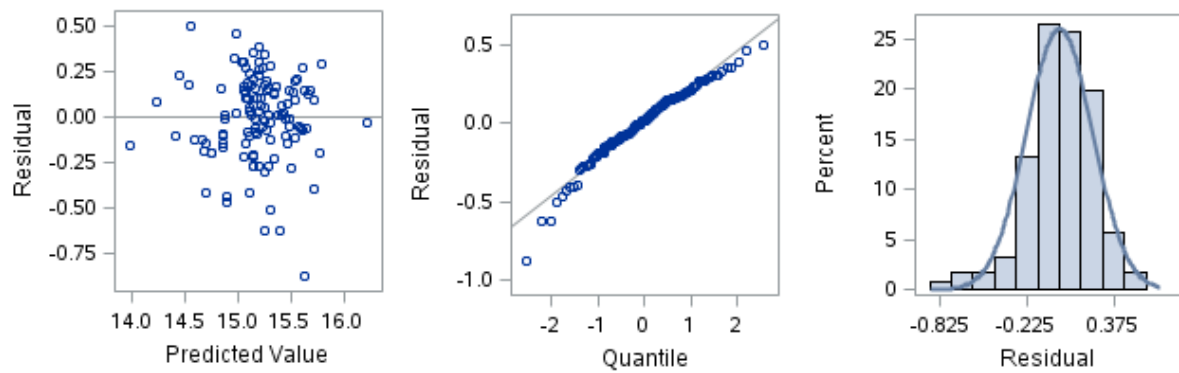
För att undersöka om de predikterade värdena är rimliga, görs ett prediktionsintervall för samtliga observationer. Vi låter konfidensgraden vara på 95 %.

Observation	Prediktionsintervall	Prediktionsintervall
1	(14.81,15.75)	(2 679 630, 6 888 207)
2	(15.01,15.95)	(3 287 899, 8 425 907)
3	(14.28,15.26)	(1 590 229, 4 225 849)
4	(15.04,15.97)	(3 390 166, 8 661 086)
5	(14.73,15.67)	(2 500 649, 6 386 885)
6	(14.87,15.81)	(2 864 710, 7 319 737)
7	(14.81,15.75)	(2 703 040, 6 919 961)
8	(14.96,15.92)	(3 138 650, 8 204 815)

Tabell 13. Prediktionsintervall för observationerna under 2016.

Som vi kan se i tabellerna ovan har första observationen ett slutpris utanför intervallet. Prediktionsintervallen för samtliga observationer är väldigt vida.

Normalfördelningsantagandet



Figur 16. Residualplott, normalfördelningsplott och histogram för slutgiltiga modellen.

Resultat

Den slutgiltiga modellen är:

$$\begin{aligned} \text{LogSlutpris} = & \beta_0 + \beta_1 \text{LogBoarea} + \beta_2 \text{LogTomt} + \beta_3 \text{AdelsöMunsö} + \beta_4 \text{Färingsö} \\ & + \beta_5 \text{LogVatten} + \beta_6 \text{Skola} + \epsilon \end{aligned}$$

Där parameterskattningarna är:

	Parameterskattning
Intercept	12,85764
LogBoarea	0,52137
LogTomt	0,09296
AdelsöMunsö	− 0,33847
Färingsö	− 0,14810
LogVatten	− 0,10508
Skola	− 0,00002878

Tabell 14. Parameterskattning av slutgiltiga modellen efter stepwise regression.

Det predikterade priset samt prediktionsintervall för de åtta husköp under januari till februari 2016 blev:

Observation	Predikterat Slutpris, (kr)	Predikterat LogSlutpris	Prediktionsintervall för LogSlutpris	Prediktionsintervall för Slutpris (kr)
1	4 310 665	15,277	(14.81,15.75)	(2 679 630, 6 888 207)
2	5 263 335	15,476	(15.01,15.95)	(3 287 899, 8 425 907)
3	2 592 309	14,768	(14.28,15.26)	(1 590 229, 4 225 849)
4	5 418 720	15,505	(15.04,15.97)	(3 390 166, 8 661 086)
5	3 996 418	15,201	(14.73,15.67)	(2 500 649, 6 386 885)
6	4 579 183	15,337	(14.87,15.81)	(2 864 710, 7 319 737)
7	4 324 920	15,280	(14.81,15.75)	(2 703 040, 6 919 961)
8	5 074 647	15,440	(14.96,15.92)	(3 138 650, 8 204 815)

Tabell 15. Predikterat värde och prediktionsintervall för observationer under 2016.

Diskussion

Syftet med den här uppsatsen var att hitta den modell som bäst beskriver vilka förklarande variabler som påverkar slutpriset för en villa eller rad-/kedjehus Ekerö kommun under 2015. För att ta reda på den bäst lämpade modellen, valdes först att studera datamaterialet för samtliga förklarande variabel. Därefter genomfördes en enkel linjär regression för samtliga förklarande variabler. Boarea hade den högsta förklaringsgrad. Sedan genomfördes en multipel regression av Slutpris på samtliga förklarande variabler och en transformering gjordes på de variabler som inte hade en homogen residualplott. Förklaringsgraden jämfördes och ökade efter transformeringen. Den slutgiltiga modellen som fås fram genom stepwise regression innehåller sex förklarande variabler.

Att LogBoarea, LogTomt samt LogVatten är signifikanta är ganska givet. Människor som söker efter en villa i Ekerö kommun är säkerligen intresserade av närhet till vattnet och har ett intresse av en viss storlek på både tomt och boarea. Parameterskattningen för LogVatten är negativ, vilket kan tolkas som att den påverkar priset negativt. Ju längre bort från vattnet, desto lägre pris.

Områdena AdelsöMunsö och Färingsö är signifikanta med negativa parameterskattningar är inte heller så överraskande. Adelsö och Munsö är två områden som ligger långt bort. Många som vill flytta till Ekerö kommun vill bo på centralare delar. På Adelsö är dessutom ett område med många sommarstugor, och de som köper hus i det området söker efter just sommarstugor och inte permanent boende.

I den slutgiltiga modellen studerades att skolavståndet är signifikant med negativ parameter. Många som bestämmer sig för att flytta ut till Ekerö kommun är barnfamiljer. Att bo nära skolan är något viktigt. Ju längre bort, desto lägre pris.

Månad är inte signifikant. Genom att studera stapeldiagrammet för månadshusköpen, konstaterades att husköpen var som högst under september. Det var även mer husköp under sommarmånaderna än under vintermånaderna. Det kan exempelvis vara mer lockande att köpa en bostad om visningsbilderna är tagna på vår och sommar. Speciellt kan det tänkas att en grönskande trädgård är mer lockande än en trist trädgård under höst- och vintermånaderna. Dock kan anledningen till att Månad inte är signifikant vara för att det fanns för få observationer och studerade endast under perioden 2015.

Slutgiltiga modellen har en förklaringsgrad på 66,8 %. Den är relativt låg vilket betyder att det finns fler förklarande variabler som påverkar priset. Det kan vara ifall huset bor nära en större väg, närheten till skog och natur. Sedan finns det andra förklarande variabler som man inte kunnat ta fram i datamaterialet. Hur huset se ut såsom utsida men även insidan är väldigt viktigt vid ett husköp

Det kan även finnas samspel mellan de förklarande variablerna i datamaterialet. Vi studerade ifall samspel mellan boarea och vissa förklarande variabler (skolavstånd, busshållplatsavstånd, tomtarea och månad), men inget verka tyda på ett samspel. Förklaringsgraden ökade som mest med 0,04 procentenheter. Det är svårt att studera samspel

mellan alla förklarande variabler, men det finns säkerligen någon typ av samspel mellan förklarande variablerna. Detta kan påverka förklaringsgraden

Studerar det predikterade värdena för de hus som såldes i början av år 2016, kan ses att det skiljer sig en hel del från de verkliga värdena. I och med att det inte finns information hur husen ser ut både invändigt och utvändigt, kan detta orsaka stora skillnader mellan predikterade och verkliga värdet. Om vi haft en högre förklaringsgrad, kanske det predikterade värdet hade varit närmare det verkliga värdet. Genom att studera prediktionsintervallet, kan vi konstatera att alla förutom ett av det verkliga värdet befinner sig i intervallet.

Appendix

The CORR Procedure

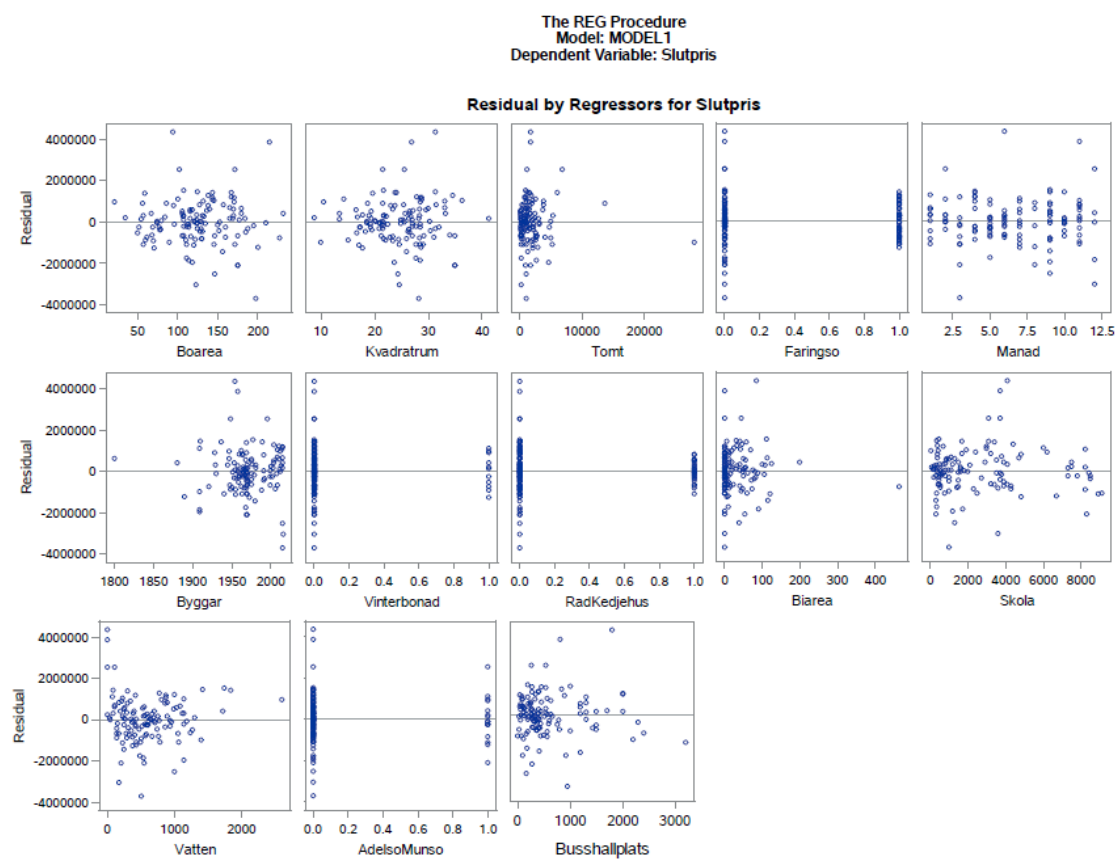
Pearson Correlation Coefficients
Prob > |r| under H0: Rho=0
Number of Observations

	Boarea	Tomt	Biarea	Byggar	Villa	Friliggande	Vinterbonad	RadKedjehus	Kvadratrutrum	Faringso	Ekero	Adelsomunso	Manad	Vatten	Skola	Busshallplats
Boarea	1.00000 0.121	-0.06943 0.4492 0.121	0.15496 0.0897 0.121	0.13749 0.1326 0.121	0.14549 0.1113 0.121	0.26855 0.0029 0.121	-0.50649 0.0001 0.121	-0.11934 0.1923 0.121	0.55590 0.0001 0.121	0.13941 0.1272 0.121	0.07446 0.4170 0.121	-0.30882 0.0006 0.121	-0.00012 0.9990 0.121	-0.03469 0.7056 0.121	-0.19908 0.0286 0.121	-0.19893 0.0287 0.121
Tomt	-0.06943 0.4492 0.121	1.00000 0.121	0.10774 0.2395 0.121	-0.29162 0.0012 0.121	0.18676 0.0403 0.121	0.06066 0.5087 0.121	0.03400 0.7112 0.121	-0.28683 0.0014 0.121	-0.19155 0.0353 0.121	0.23923 0.0082 0.121	-0.30400 0.0007 0.121	0.10737 0.2411 0.121	0.16869 0.0644 0.121	0.07084 0.4401 0.121	0.15207 0.0959 0.121	0.02644 0.7734 0.121
Biarea	0.15496 0.0897 0.121	0.10774 0.2395 0.121	1.00000 0.121	-0.22040 0.0151 0.121	0.00176 0.9847 0.121	0.19473 0.0323 0.121	-0.08833 0.3353 0.121	-0.17661 0.0526 0.121	-0.09020 0.3251 0.121	0.05993 0.5138 0.121	-0.01936 0.8331 0.121	-0.05694 0.5350 0.121	-0.07435 0.4177 0.121	0.02932 0.7495 0.121	-0.06467 0.4810 0.121	-0.07695 0.4015 0.121
Byggar	0.13749 0.1326 0.121	-0.29162 0.0012 0.121	-0.22040 0.0151 0.121	1.00000 0.121	-0.07573 0.4090 0.121	0.07845 0.3924 0.121	-0.10425 0.2551 0.121	0.05440 0.5534 0.121	0.26702 0.0031 0.121	-0.03799 0.6791 0.121	0.12864 0.1596 0.121	-0.13568 0.1378 0.121	-0.11547 0.2072 0.121	-0.10799 0.2384 0.121	-0.03486 0.7043 0.121	0.01859 0.8396 0.121
Villa	0.14549 0.1113 0.121	0.18676 0.0403 0.121	0.00176 0.9847 0.121	-0.07573 0.4090 0.121	1.00000 0.122	-0.49923 0.0001 0.122	-0.16782 0.0646 0.122	-0.27744 0.0020 0.122	0.06551 0.4753 0.122	0.00168 0.9854 0.122	0.10748 0.2387 0.122	-0.14863 0.1023 0.122	0.26000 0.0040 0.122	0.00406 0.9648 0.122	-0.01691 0.8540 0.122	-0.09715 0.2891 0.122
Friliggande	0.26855 0.0029 0.121	0.06066 0.5087 0.121	0.0323 0.9847 0.121	0.07845 0.3924 0.121	-0.49923 0.0001 0.122	1.00000 0.122	-0.29479 0.0010 0.122	-0.48734 0.0071 0.122	0.24372 0.0071 0.122	0.20346 0.0246 0.122	-0.19825 0.0286 0.122	0.02553 0.7801 0.122	-0.36420 0.0001 0.122	-0.14187 0.1206 0.122	0.04515 0.6229 0.122	-0.00619 0.9463 0.122
Vinterbonad	-0.50649 0.1923 0.121	0.03400 0.7112 0.121	-0.08833 0.3353 0.121	-0.10425 0.2551 0.121	-0.16782 0.0646 0.122	-0.29479 0.0010 0.122	1.00000 0.122	-0.16383 0.0714 0.122	-0.42387 0.0001 0.121	-0.00340 0.9703 0.122	-0.30968 0.0005 0.122	0.47115 0.0001 0.122	0.05235 0.5685 0.121	0.03872 0.6733 0.121	0.39391 0.0001 0.121	0.48823 0.0001 0.121
RadKedjehus	-0.11934 0.1923 0.121	-0.28683 0.0014 0.121	-0.17661 0.0526 0.121	0.05440 0.5534 0.121	-0.27744 0.0020 0.122	-0.48734 0.0001 0.122	-0.16383 0.0714 0.122	1.00000 0.122	-0.06593 0.4725 0.121	-0.23191 0.0102 0.122	0.36887 0.8568 0.121	-0.20219 0.0255 0.122	0.14241 0.1192 0.121	0.14120 0.1224 0.121	-0.31344 0.0005 0.121	-0.23572 0.0052 0.121
Kvadratrutrum	0.55590 0.0001 0.121	-0.19155 0.0353 0.121	-0.09020 0.3251 0.121	0.26702 0.0031 0.121	0.06551 0.4753 0.121	0.24372 0.0071 0.121	-0.42387 0.0001 0.121	-0.06593 0.4725 0.121	1.00000 0.121	0.21575 0.0175 0.121	0.01658 0.8568 0.121	-0.33230 0.0002 0.121	-0.12198 0.1826 0.121	-0.08272 0.3671 0.121	-0.08161 0.3735 0.121	-0.00134 0.9884 0.121
Faringso	0.13941 0.1272 0.121	0.23923 0.0082 0.121	0.05993 0.5138 0.121	-0.03799 0.6791 0.121	0.00168 0.9854 0.122	0.20346 0.0246 0.122	-0.00340 0.9703 0.122	-0.23191 0.0102 0.122	0.21575 0.0175 0.121	1.00000 0.122	-0.75204 0.0001 0.122	-0.29701 0.0009 0.122	-0.15190 0.0963 0.121	0.21708 0.0168 0.121	0.05566 0.5443 0.121	0.10488 0.2523 0.121
Ekero	0.07446 0.4170 0.121	-0.30400 0.0007 0.121	-0.01936 0.8331 0.121	0.12864 0.1596 0.121	0.10748 0.2387 0.122	-0.19825 0.0286 0.122	-0.30968 0.0005 0.122	0.36887 0.8568 0.121	0.01658 0.8568 0.121	-0.75204 0.0001 0.122	1.00000 0.122	-0.38220 0.0013 0.122	0.15909 0.0813 0.121	-0.20813 0.0220 0.121	-0.44641 0.0001 0.121	-0.23197 0.0105 0.121
Adelsomunso	-0.30882 0.0006 0.121	0.10737 0.2411 0.121	-0.05694 0.5350 0.121	-0.13568 0.1378 0.121	-0.14863 0.1023 0.122	0.02553 0.7801 0.122	0.47115 0.0001 0.122	-0.20219 0.0255 0.122	-0.33230 0.0002 0.121	-0.29701 0.0009 0.122	-0.38220 0.0001 0.122	1.00000 0.122	-0.01808 0.8439 0.121	-0.00253 0.9780 0.121	0.57949 0.0001 0.121	0.19275 0.0342 0.121
Manad	-0.00012 0.9990 0.121	0.16869 0.0644 0.121	-0.07435 0.4177 0.121	-0.11547 0.2072 0.121	0.26000 0.0040 0.121	-0.36420 0.0001 0.121	0.05235 0.5685 0.121	0.14241 0.1192 0.121	-0.12198 0.1826 0.121	-0.15190 0.0963 0.121	0.15909 0.0813 0.121	-0.01808 0.8439 0.121	1.00000 0.121	-0.01692 0.8539 0.121	0.04640 0.6133 0.121	-0.08403 0.3595 0.121
Vatten	-0.03469 0.7056 0.121	0.07084 0.4401 0.121	0.02932 0.7495 0.121	-0.10799 0.2384 0.121	0.00406 0.9648 0.121	-0.14187 0.1206 0.121	0.03872 0.6733 0.121	0.14120 0.1224 0.121	-0.08272 0.3671 0.121	0.21708 0.0168 0.121	-0.20813 0.0220 0.121	-0.00253 0.9780 0.121	-0.01692 0.8539 0.121	1.00000 0.121	-0.31025 0.0005 0.121	-0.20382 0.0249 0.121
Skola	-0.19908 0.0286 0.121	0.15207 0.0959 0.121	-0.06467 0.4810 0.121	-0.03486 0.7043 0.121	-0.01691 0.8540 0.121	0.04515 0.6229 0.121	0.39391 0.0001 0.121	-0.31344 0.0005 0.121	-0.08161 0.3735 0.121	0.05566 0.5443 0.121	-0.44641 0.0001 0.121	0.57949 0.0001 0.121	0.04640 0.6133 0.121	-0.31025 0.0005 0.121	1.00000 0.121	0.44244 0.0001 0.121
Busshallplats	-0.19893 0.0287 0.121	0.02644 0.7734 0.121	-0.07695 0.4015 0.121	0.01859 0.8396 0.121	-0.09715 0.2891 0.121	-0.00619 0.9463 0.121	0.48823 0.0001 0.121	-0.23572 0.0052 0.121	-0.00134 0.9884 0.121	0.10488 0.2523 0.121	-0.23197 0.0105 0.121	0.19275 0.0342 0.121	-0.08403 0.3595 0.121	-0.20382 0.0249 0.121	0.44244 0.0001 0.121	1.00000 0.121

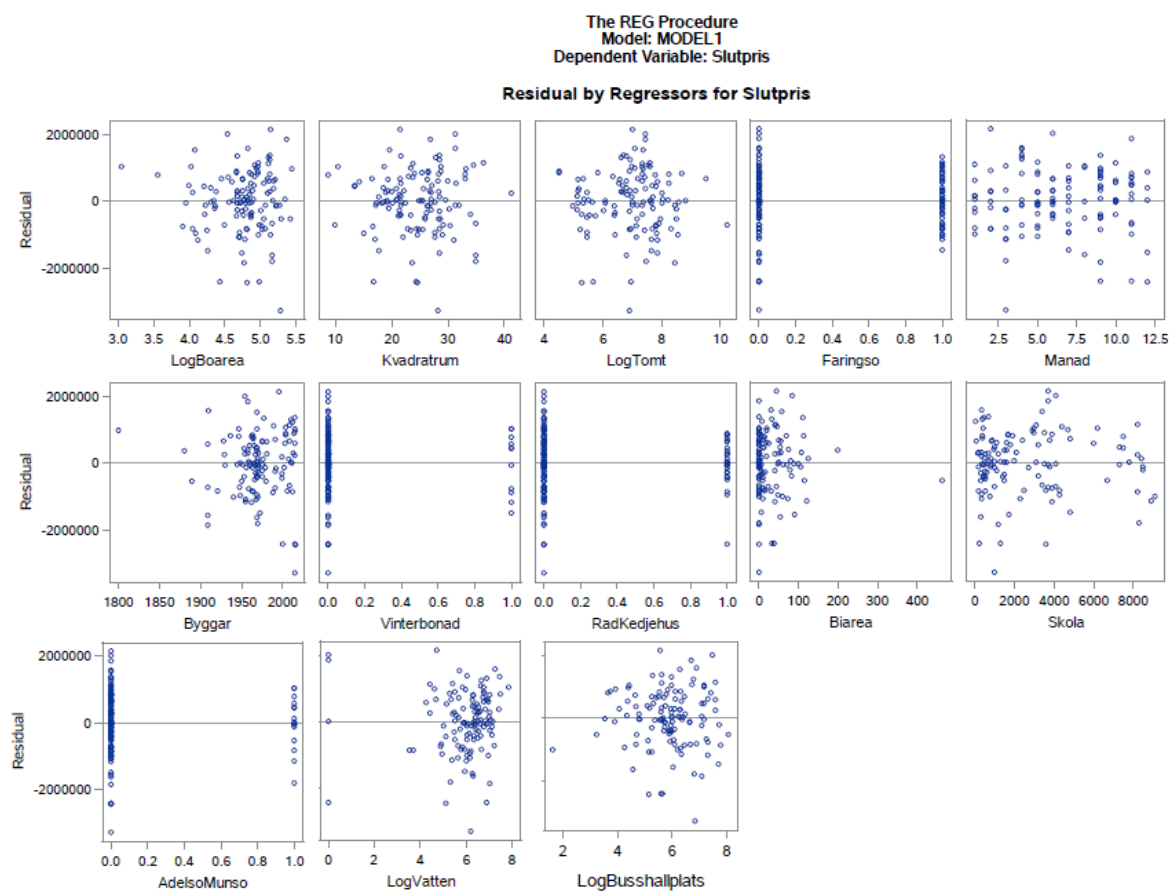
Figur 17. Korrelation mellan samtliga förklarande variabler.

Boarea	LogBoarea	Biarea	Tomt	Bygg	Kvadratrutrum	Omrade	Typ	Manad	Vatten	Skola	Busshallplats
123	4.812184354	12	735	1964	17.571428571	0	1	1	742	312	457
149	5.0039463059	76	1442	1979	16.555555556	0	0	1	527	274	120
75	4.3174881135	0	3600	1939	25	2	0	1	826	2000	290
147	4.9904325868	0	828	1995	29.4	0	0	1	241	83	113
120	4.7874917428	70	710	1967	17.142857143	0	0	2	945	1500	187
107	4.6728288345	9	1168	1949	21.4	0	1	2	294	564	233
156	5.0498560072	0	1390	1996	26	1	0	2	741	1420	1400
172	5.1474944768	0	92	1988	28.666666667	0	0	2	102	1250	592

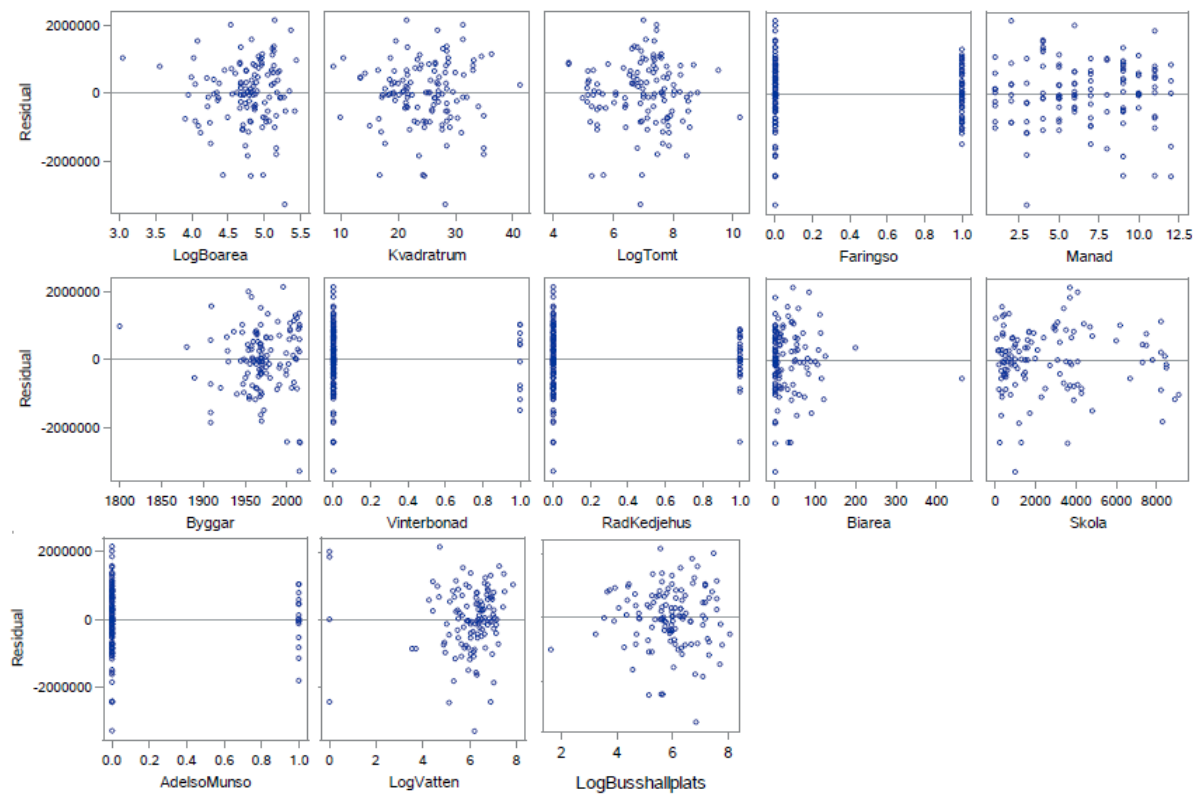
Figur 18. Samtliga observationer för sålda hus under 2016 under januari och februari.



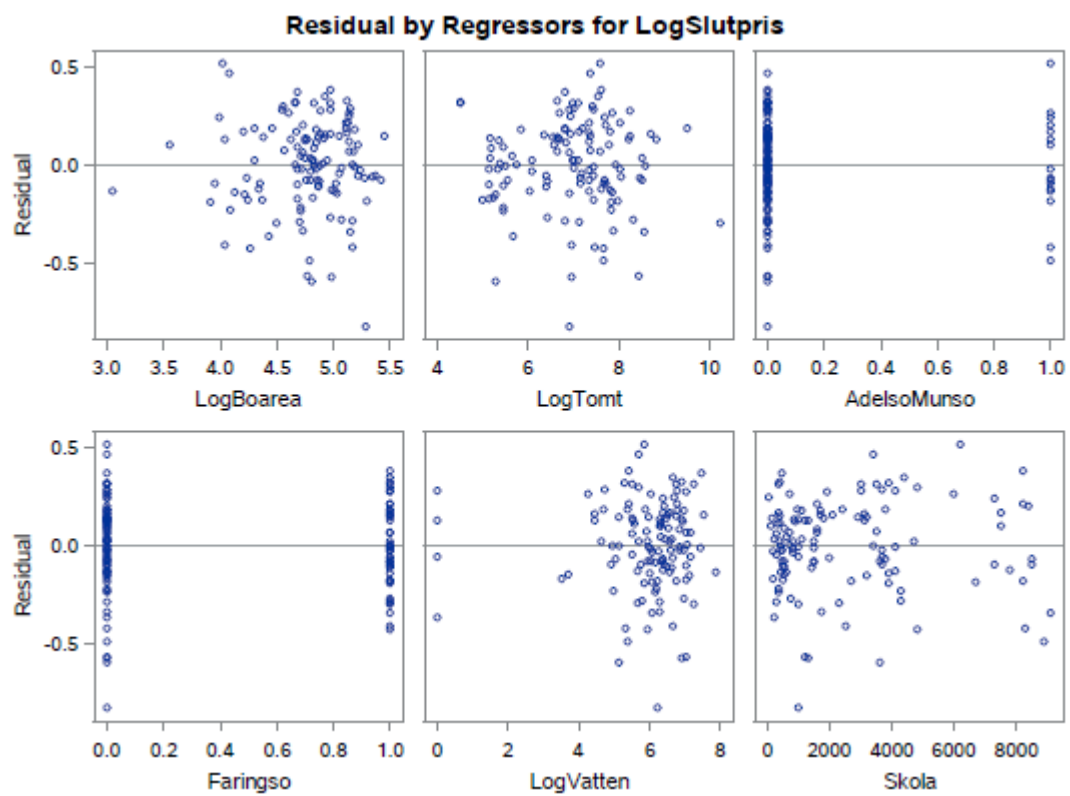
Figur 19. Samtliga residualplottar för modell 1.



Figur 20. Samtliga residualplottar för modell 2.



Figur 21. Samtliga residualplottar för modell med LogSlutpris som responsvariabel



Figur 22. Samtliga residualplottar för den slutgiltiga modellen.

Referenser

1. www.hemnet.se
2. Sundberg Rolf. Kompendium i Lineära Statistiska Modeller
3. Wikipedia, https://sv.wikipedia.org/wiki/Eker%C3%B6_kommun