



Stockholms
universitet

Biltäthet i Stockholms läns kommuner - en statistisk analys av kommunala skillnader

Fatemeh Engqvist

Kandidatuppsats 2017:23
Matematisk statistik
Juni 2017

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Biltäthet i Stockholms läns kommuner - en statistisk analys av kommunala skillnader

Fatemeh Engqvist*

Juni 2017

Sammanfattning

Biltätheten i Stockholms län är en av de lägsta i Sverige men biltätheten varierar mycket mellan länets kommuner. Denna uppsats skrivs med avsikt att statistiskt belysa och förklara den kommunala variation som förekommer i Stockholms läns fordonsstatistik. Syftet är att hitta faktorer som förklarar variationen beträffande biltäthet med avseende på antal fysiskt ägda personbilar i trafik per 1000 invånare. För att åstadkomma detta utformas en linjär regressionsmodell med biltäthet som responsvariabel. Observationerna sträcker sig under tidsperioden 2006-2015. Förklarande variabler som inkluderas i arbetet är de socioekonomiska, demografiska och geografiska faktorer som kan påverka biltäthet. För att avgöra vilken modell som beskriver data bäst samt för att reducera variabler används backward elimination och forward selection procedurer samt AIC kriterium.

Resultatet visar att bland de alla faktorer har vägvstånd till Stockholms centrum (Sergels torg) störst inverkan på biltäthet. Faktorerna medelinkomst och antal förvärvsarbetande utpendlare per 1000 invånare har stark positiv påverkan på biltäthet. Dessutom har åldersmässigt fördelade variabler olika påverkan på biltäthet. Det visade sig att åldersgrupp 3 (35-44 år), åldersgrupp 4 (45-54 år) och åldersgrupp 6 (65+) har signifikant och positivt effekt på biltäthet. Däremot har befolkningstäthet, utveckling i tid (år) och bensinpris negativ effekt på biltäthet.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.
E-post: fatemeh.engqvist@gmail.com. Handledare: Tom Britton och Benjamin Allévius.

Biltäthet i Stockholms läns kommuner - en statistisk analys av kommunala skillnader

Fatemeh Engqvist

5 juni 2017

Abstract

Stockholm has been one of the counties with the lowest private car density in Sweden, but the private car density varies considerably between municipalities in this county. The purpose of this paper is to investigate the variation in private car density between the municipalities in the Stockholm County. To achieve this a linear regression model is used with private car per 1000 inhabitants as the response variable. Explanatory variables that are included in this study are among the socio economic, demographic and geographic factors that can explain the private car density. Observations are in the 10-year time period from 2006 to 2015. To find a simple model that describes the data best stepwise regression methods (backward elimination and forward selection) and AIC criterion are used.

The result shows that the distance to Stockholm centre has the biggest effect on private car density. Moreover, the medium income and the number of commuters have a large positive effect as well. The model also shows that different age groups have different effect on the private car density. Age group 3 (35-44 year-old), age group 4 (45-54 year-old) and age group 6 (65+) have significant positive effect on private car density. In contrary, population density, time (year) and gas price all have negative effect on private car density.

Sammanfattning

Biltätheten i Stockholms län är en av de lägsta i Sverige men biltätheten varierar mycket mellan länets kommuner. Denna uppsats skrivs med avsikt att statistiskt belysa och förklara den kommunala variation som förekommer i Stockholms läns fordonsstatistik. Syftet är att hitta faktorer som förklarar variationen beträffande biltäthet med avseende på antal fysiskt ägda personbilar i trafik per 1000 invånare. För att åstadkomma detta utformas en linjär regressionsmodell med biltäthet som responsvariabel. Observationerna sträcker sig under tidsperioden 2006-2015. Förklarande variabler som inkluderas i arbetet är de socioekonomiska, demografiska och geografiska faktorer som kan påverka biltäthet. För att avgöra vilken modell som beskriver data bäst samt

för att reducera variabler används backward elimination och forward selection procedurer samt AIC kriterium.

Resultatet visar att bland de alla faktorer har vägvstånd till Stockholms centrum (Sergels torg) störst inverkan på biltäthet. Faktorerna medelinkomst och antal förvärvsarbetande utpendlare per 1000 invånare har stark positiv påverkan på biltäthet. Dessutom har åldersmässigt fördelade variabler olika påverkan på biltäthet. Det visade sig att åldersgrupp 3 (35-44 år), åldersgrupp 4 (45-54 år) och åldersgrupp 6 (65+) har signifikant och positivt effekt på biltäthet. Däremot har befolkningstäthet, utveckling i tid (år) och bensinpris negativ effekt på biltäthet.

Acknowledgment

Jag vill börja med att rikta ett stort tack till mina två handledare Tom Britton och Benjamin Allévius som bidragit med råd och stort engagemang under arbetets gång. Ett stort tack går också till Anders Engqvist för hans stöd genom hela utbildningen.

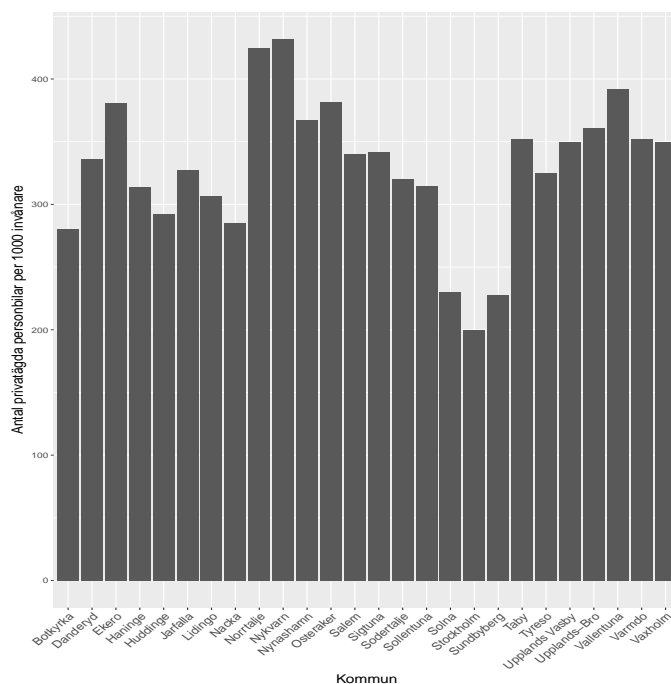
Detta är ett examensarbete motsvarande 15 högskolepoäng i matematisk-statistik. Arbetet är skrivet vid matematiska institutionen på Stockholms Universitet.

Innehåll

| | | |
|----------|---|-----------|
| 1 | Inledning | 5 |
| 2 | Teori: Multipel linjär regression | 7 |
| 2.1 | Modellen | 7 |
| 2.2 | Modell antaganden | 7 |
| 2.3 | Minsta-kvadratmetoden och parameterskattningar | 8 |
| 2.4 | Hypotesprövning och p -värde | 9 |
| 2.5 | Förklaringsgrad R^2 | 9 |
| 2.6 | Multikollinearitet | 10 |
| 2.7 | Val av modell | 11 |
| 2.7.1 | Stegvis variabelselektion: Backward elimination | 11 |
| 2.7.2 | Stegvis variabelselektion: Forward selection | 11 |
| 2.7.3 | Akaikes Information Criterion som en jämförelsemått | 11 |
| 3 | Datamaterial | 12 |
| 3.1 | Val av faktorer | 12 |
| 3.2 | Bearbetning av data | 13 |
| 4 | Analys av data och modellval | 15 |
| 4.1 | Multikollinearitet | 18 |
| 4.2 | Modellval | 21 |
| 4.2.1 | Modell baserad på Backward elimination | 21 |
| 4.2.2 | Modell baserad på Forward selection | 22 |
| 4.2.3 | Modell baserad på AIC kriterium | 22 |
| 4.2.4 | Vald modell | 22 |
| 5 | Modellvalidering, resultat och diskussion | 23 |
| 5.1 | Modellvalidering | 23 |
| 5.2 | Resultat | 26 |
| 5.3 | Diskussion | 27 |
| 6 | Bilaga | 29 |
| 7 | Referenser | 31 |

1 Inledning

Fordonsstatistik visar att Stockholms län har varit det minst biltäta län i Sverige under tioårsperioden 2006-2015 med i genomsnitt 271 antal privatägda bilar per 1000 invånare. Men biltätheten i länet varierar stort bland de 26 kommunerna. Till exempel år 2015 var Nykvarn kommun den biltätaste kommunen med 432 privatägda bilar per 1000 invånare medan motsvarande siffra för Stockholms kommun var 200 (SCB, 2017).



Figur 1.1: Antal privatägda personbilar per 1000 invånare i Stockholms läns kommuner år 2015.

Figur 1.1 visar den skillnaden i biltäthet mellan kommunerna för år 2015. De biltätaste kommunerna är Nykvarn, Norrtälje och Vallentuna med 432, 425 samt 392 privatägda bilar per 1000 invånare och de minst biltäta kommunerna är Stockholm, Sundbyberg samt Solna vars motsvarande siffra är 200, 228 och 230. Det föreligger med andra ord stora skillnader mellan länets kommuner.

Syftet med den här uppsatsen är att undersöka och beskriva faktorer som påverkar biltäthet. Syftet leder till frågeställningen: vilka faktorer inverkar på biltätheten på kommunal nivå? För att kunna analysera och förklara den stora skillnaden i biltäthet i länet vill vi att ta reda på och kontrollera sambandet mellan biltäthet och tillämpliga förklarande variabler. Modellen

vi ska använda oss av är multipel linjär regression.

Vi kommer att analysera datamaterial bestående av data från samtliga 26 kommuner i Stockholms län under tioårsperioden 2006-2015. Variabler som används i denna uppsats datamaterial har valts mest utifrån tidigare studier och kan delas in enligt följande:

- Socioekonomiska och demografiska variabler inom respektive kommun:
 - Befolkningsstorlek
 - Befolkningstäthet
 - Medelinkomst
 - Medelålder
 - Åldersmässig fördelning inom respektive kommun
- Geografiska variabler:
 - Arbetspendling som beskriver hur stor andel som pendlar över kommungräns och andel som bor och arbetar i samma kommun
 - Vägavstånd från kommuns centrum till Stockholms centrum (Sergels torg)
- Benispris

I uppsatsen inkluderas inte personbilar som ägs av juridiska personer¹. Anledningen är att det gissningsvis finns olika faktorer som påverkar innehav av personbil för en fysisk person jämfört med juridisk person. Dessutom behövs inte en registrerad personbil ägd av juridisk person i en kommun nödvändigtvis användas i samma kommun. Därför har vi valt att begränsa undersökningen till personbilar som endast ägs av fysiska personer, det vill säga privatägda bilar, för att komma fram till en allmän förklaringsmodell. Så med *biltäthet/fordonstäthet* menas biltäthet/fordonstäthet för privatägda personbilar.

I avsnitt 2 ska vi gå igenom den teori som är relevant för linjära regressionsmodeller. Vidare ska vi i avsnitt 3 beskriva datamaterialet i detalj samt skapa och välja variabler vars påverkan på biltäthet ska analyseras närmare. I avsnitt 4 ska vi ta fram några olika modeller och välja den som är mest lämplig. I avsnitt 5 går vi igenom resultatet och slutmodellen och diskussion och förslag på hur man kan förbättra den här undersökningen.

¹Det finns två typer av ägare: fysiska och juridiska personer. Juridisk person avser företag som bedriver någon form av ekonomisk verksamhet. Ett betydande antal företag är organiserade som personliga företag, som drivs av fysisk person (redovisas som juridiska personer).

2 Teori: Multipel linjär regression

För att belysa förhållandet mellan den beroende variabeln *biltäthet* och ett antal förklarande variabler används en *multipel linjär regressionsmodell*. Med en regressionsmodell syftar man ofta att välja ut de förklarande variabler som maximerar andelen förklarad variation. Avsnittet nedan redogör både de metoder som används i uppsatsen och de specifika modellkriterier och test som används för modellväll.

Följande avsnitt baseras på Rolf Sundbergs kompendium *Lineara Statistiska Modeller* (2016). Andra källor noteras i text om det används.

2.1 Modellen

Den allmänna formen av en multipel linjär regressionsmodell är

$$Y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_m x_{mi} + \varepsilon_i, \quad i = 1, \dots, N. \quad (1)$$

I uttrycket ovan är $\{Y_i, i = 1, \dots, N\}$ en oberoende uppsättning av N stokastiska variabler. Vi betraktar observerade värdena y_i som utfall av de oberoende stokastiska variabeln Y_i med en gemensam okänd varians σ^2 . $\{\beta_j, j = 0, \dots, m\}$ är $k = m + 1$ okända parametrar och $\{X_{ji}, j = 0, \dots, m, i = 1, \dots, N\}$ är värdet på den j :te förklarande variabeln för den i :te observationen. $\{\varepsilon_i, i = 1, \dots, N\}$ är en slumpmässig felterm.

Vi kan skriva regressionsmodellen (1) i matrisform

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

där

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{m1} \\ 1 & x_{12} & \cdots & x_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1N} & \cdots & x_{mN} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{pmatrix}.$$

2.2 Modell antaganden

I linjär regression gör vi ett antal antaganden:

1. Linjäritet: Regressionsmodellen är linjär i parametrar och det finns linjärt samband mellan förklarande variabler och responsvariabel.
2. Givet N oberoende (normalfördelade) stokastiska variabler Y_1, \dots, Y_N , där $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, och där $\{\varepsilon_i, i = 1, \dots, N\}$ är normalfördelad slumpmässig felterm med $E(\varepsilon_i) = 0$, det vill säga $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \mathbf{I})$. Att feltermerna har konstant varians är en viktig antagande och kallas homoskedasticitet.
3. Matrisen \mathbf{X} har rang k .
4. Ingen autokorrelation: ε_i och ε_j okorrelerade så att $Cov(\varepsilon_i, \varepsilon_j) = 0$ för alla $i, j; i \neq j$. Detta medför att observationer av den beroende variabeln är också okorrelerade.

2.3 Minsta-kvadratmetoden och parameterskattningar

För att skatta den okända parametervektorn $\boldsymbol{\beta}$ kan minsta-kvadratmetoden användas. Vi betecknar $\hat{\boldsymbol{\beta}}$ som skattning av $\boldsymbol{\beta}$ med minsta-kvadratmetoden. $\hat{\boldsymbol{\beta}}$ minimerar summan av den kvadratiske skillnaden mellan observationerna och deras väntevärden (residualkvadratsumman $\sum_{i=1}^N e_i^2$) som visas i ekvation 2.

$$\sum_{i=1}^N e_i^2 = \sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^m \beta_j x_{ji} \right)^2. \quad (2)$$

Vi kan skriva om ekvation (2) som

$$\sum_{i=1}^N e_i^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2. \quad (3)$$

$\hat{\boldsymbol{\beta}}$ som minimerar ekvationen (3) är

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

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \quad (5)$$

$$Var(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^T \mathbf{X})^{-1} \quad (6)$$

Men σ^2 är okänd, därför skattar vi variansen ur data genom

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^N (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})^2}{N - k} \quad (7)$$

2.4 Hypotesprövning och p -värde

För att undersöka om en (eller flera) förklaringsvariabel X_j har inverkan på *Bilper1000* som responsvariabel testar vi hypotesen om regressionskoefficient β_j är signifikant skild från noll med t -test eller F -test. Under modell antaganden $\hat{\beta}_j \sim N(\beta_j, \sigma^2(\mathbf{X}^T \mathbf{X})^{-1})_{jj}$.

$$H_0 : \beta_j = 0$$

$$H_a : \beta_j \neq 0$$

Teststatistikan är

$$T_j = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma} \sqrt{(\mathbf{X}^T \mathbf{X})^{-1})_{jj}}}$$

vilket under H_0 motsvarar $\frac{\hat{\beta}_j}{\hat{\sigma} \sqrt{(\mathbf{X}^T \mathbf{X})^{-1})_{jj}}}$, och är t -fördelade med $N - k$ frihetsgrader. Vi förkastar H_0 på signifikansnivån α om

$$|T_j| \geq t_\alpha(N - k).$$

Låt t_j det observerade värdet på teststatistika T_j . p -värdet för X_j definieras som sannolikheten att få ett värdet lika med eller mer extremt än observerade värdet t_j då $H_0 : \beta_j = 0$ är sann", det vill säga:

$$p\text{-värde} = P_{H_0}(|T_j| \geq |t_j|).$$

Om vi får stor p -värdet vi kan inte förkasta $H_0 : \beta_j = 0$ och detta betyder att variabeln X_j är ej signifikant. Däremot variabeln X_j är signifikant på signifikansnivån α om p -värdet $\leq \alpha$.

2.5 Förklaringsgrad R^2

Variationen i regressionsmodell delas in i tre termer:

$$\text{Kvs}(\text{totalt}) = \sum_i (Y_j - \bar{Y})^2$$

$$\text{Kvs}(\text{regression}) = \sum_i \left(\sum \hat{\beta}_j (x_{ij} - \bar{x}_j) \right)^2$$

$$\text{Kvs}(\text{residual}) = \text{Kvs}(\text{totalt}) - \text{Kvs}(\text{regression}) = \sum_{i=1}^N \left(y_i - \beta_0 - \sum_{j=1}^m \beta_j x_{ji} \right)^2$$

Förklaringsgraden R^2 är ett anpassningsmått i samband med multipel regressionsmodeller vilket definieras som kvoten mellan regression kvadratsumma och totalt kvadratsumman

$$R^2 = \frac{\text{Kvs}(\text{regression})}{\text{Kvs}(\text{totalt})} = 1 - \frac{\text{Kvs}(\text{residual})}{\text{Kvs}(\text{totalt})}.$$

R^2 är ett mått som visar hur mycket variation i regressionsmodellen kan förklara totala variationen och antar värden mellan 0 och 1 där högre R^2 föredras. Men nackdelen med R^2 är att det alltid ökar när vi tillför fler variabel i modellen. Det kan därför vara bättre att se om $\hat{\sigma}^2$ minskar om man ska tillföra fler variabler. Ett sådant mått kallas Adjusted R^2 , R_{adj}^2 och definieras som

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

där $\hat{\sigma}_0^2 = \text{Kvs}(\text{totalt})/(N - 1)$ är $\hat{\sigma}^2$ -skattningen när man inte har någon x-variabel i modellen (Andersson, P. Tyrcha, J., sidor 29-30).

2.6 Multikollinearitet

Om det finns linjärt kombination mellan förklaringsvariabler kallas det *multikollinearitet* och det kan förekomma när man har många potentiella förklaringsvariabler. VIF är ett mått på multikollinearitet. Det står för "Variance Inflation Factor" och den definieras som

$$VIF_j = \frac{1}{1 - R_j^2},$$

där R_j^2 är förklaringsgraden i en anpassad modell där X_j förklaras av övriga X-variabler. Stora värde för R_j^2 betyder att variationen i variabeln X_j kan förklaras av de andra variabler, detta leder till mindre nämnare i VIF-kvoten och till stora VIF-värdet. Så för att undvika multikollinearitet vill vi utesluta variabler som har höga VIF-värdet. Multikollinearitet kan anses som för stor om värdet av R_j^2 överstiger en viss gräns.

2.7 Val av modell

Det finns olika metoder för att kunna reducera en modell till enklare modell med färre parametrar. Stegvisa variabelselektionsmetoder använder p -värdet och förklaringsgraden R^2 som kriterium för att välja lämpligaste modell medan det finns andra kriterium som till exempel Akaikes Information Criterion (AIC) som kan användas som en jämförelsemått. Nedan beskriver vi tre olika metoder för att välja en modell.

2.7.1 Stegvis variabelselektion: Backward elimination

I denna procedur börjar vi med en multipel regression med samtliga förklarande variabler. Sedan utesluter vi den mest icke-signifikant variabeln (med högst p -värde). Detta är ekvivalent med att eliminera den variabel vars eliminering har mindre sänkning i förklaringsgraden R^2 . Vi fortsätter proceduren tills alla förklarande variabler som kvar är signifikanta.

2.7.2 Stegvis variabelselektion: Forward selection

I denna procedur utgår från en modell utan variabel och adderar en variabel i taget till modellen. Vi börjar med variabel som har högsta korrelation med responsvariabeln. Sedan i varje steg lägger vi en variabel som är mest signifikant i test av om motsvarande parameter är noll (lägst p -värde). Proceduren slutar när det inte finns fler signifikanta variabler att inkludera.

2.7.3 Akaikes Information Criterion som en jämförelsemått

Akaikes Information Criterion (AIC) är ett annat jämförelsemått som kan användas för att bestämma lämpligaste modell för att beskriva data. För givet Modell definieras AIC som:

$$AIC(\text{Modell}) = -2L(\text{Modell}) + 2p(\text{Modell})$$

där $L(\text{Modell})$ är den maximerade log-likelihoodfunktionen av modellens parametrar och $p(\text{Modell})$ är antalet parametrar i modell. Modellen med lägst AIC föredras. AIC är ett mått för att jämföra mellan olika modeller och talar inte om hur bra en modell är i för sig (Fahrmeir, L., m.fl., sidan 148).

3 Datamaterial

3.1 Val av faktorer

För att välja de faktorer som är relevanta till biltäthet har tidigare studier undersökts. En stor andel studier undersöker bilinnehavet på enskild individnivå så som faktorer som ålder, kön och inkomst. Eftersom vi inte har tillgång till data på enskild nivå har vi valt att använda aggregerad data på kommuns nivå.

En studie gjort av Matstoms (2002) visar att beslutet att ha en bil eller inte varierar mera med ålder än andra faktorer. Beesley och Kain (1964) har i sin studie om amerikanska städer och bilinnehav använt bland annat familjens medelinkomst som förklarande variabler. Tanner (1979) har i en tvärsnittsstudie tagit befolkningstäthet, inkomst och medelålder som förklarande variabler.

Med avseende på de nämnda valda förklarande variabler i tidigare studier och tillgängliga data, har vi valt följande förklarande variabler. Datamaterialet förutom några förklaringsvariabler (kommer att nämnas) har hämtas huvudsakligen från SCB:s statistikdatabas. Det finns 26 kommuner i Stockholms län och data har hämtats ut från år 2006 till 2015.

Responsvariabeln är antal privatägda bilar per 1000 invånare efter kommun och år. Vi har tagit år som en faktor. Detta gör att responsvektorn blir en vektor av längd 260 (26x10). Detta är inte helt realistisk men vi har antagit att antalet nyregistrerade fysiska personbilar av ett år är oberoende av året innan. Detta antaganden behövs för vårt modell eftersom enligt modellantaganden i avsnitt 2.2 vi har N oberoende (normalfördelade) stokastiska variabler Y_1, \dots, Y_N , där $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, och där $\{\varepsilon_i, i = 1, \dots, N\}$ är normalfördelad slumpmässig felterm. Feltermerna kommer inte vara slumpmässig om Y_1, \dots, Y_N är beroende.

De förklarande variabler som används kan delas in enligt följande:

- *Socioekonomiska/demografiska* variabler som avser att redogöra för befolkning, befolkningstäthet, medelålder samt åldersmässig fördelning inom respektive kommun. Den åldersmässiga indelning görs i 10-årsklasser samt åldersgrupp 65 och uppåt. Dessutom inkluderar vi variabeln medelinkomst som beskriver ekonomisk styrka inom respektive kommun.
- *Geografiska* variabler som beskriver hur stor andel som pendlar över kommungräns och andel som bor och arbetar i samma kommun. Dessutom har vi variabeln *Avstånd* som avser att redogöra var en kommun ligger i relation till Stockholms centrum.
- Benispris

Responsvariabel:

Bilper1000. Antal fysiskt ägda personbilar i trafik per 1000 invånare efter kommun och år, det vill säga privatbilar. Personbilar ägda av juridiska personer inkluderas inte i det här data.

Förklarande variabler:

Invånare. Antal personer som är folkbokförda efter kommun och år.

Invånarekm2. Antal invånare per kvadratkilometer.

Medelinkomst. Medelinkomst efter kommun och år.

Medelålder. Medelålder efter kommun och år.

Boroarbetar. Antal förvärvsarbetande 16+ år som bor och arbetar inom samma kommun.

Utpendlare. Antal förvärvsarbetande 16+ år utpendlare över kommungräns.

Avstånd. Vägavstånd (i kilometer) från motsvarande kommuns centrum till Stockholms centrum (Sergels torg)².

Bensinpris. Bensinpriset under tidsperioden 2006-2015 (Källa: SPBI).

Ålder1. Antal personer i åldersgruppen 15-24.

Ålder2. Antal personer i åldersgruppen 25-34.

Ålder3. Antal personer i åldersgruppen 35-44.

Ålder4. Antal personer i åldersgruppen 45-54.

Ålder5. Antal personer i åldersgruppen 55-64.

Ålder6. Antal personer i åldersgruppen 65+ .

År. Detta motsvarar tidsperioden 2006-2015.

Antal personer i åldersgrupp under 15 år inkluderas inte som en separat åldersgrupp i förklaringsvariabler dock det inkluderas genom variabeln *Invånare*. Detta beror på att åldersgrupp under 15 har inte direkt påverkan på bilinnehavet men däremot deras inverkan på bilinnehavet speglas i andra åldersgrupper.

3.2 Bearbetning av data

För att lättare kunna jämföra mellan kommuner är det rimligt att ta antal per 1000 invånare för följande data:

²Hämtas från <https://www.avstandmellan.com>

Boroarbetarper1000. Antal förvärvsarbetande 16+ år som bor och arbetar i kommunen

Utpendlareper1000. Antal förvärvsarbetande 16+ år utpendlare över kommungräns per 1000 invånare.

Ålder1per1000. Antal personer i åldersgruppen 15-24 per 1000 invånare.

Ålder2per1000. Antal personer i åldersgruppen 25-34 per 1000 invånare.

Ålder3per1000. Antal personer i åldersgruppen 35-44 per 1000 invånare.

Ålder4per1000. Antal personer i åldersgruppen 45-54 per 1000 invånare.

Ålder5per1000. Antal personer i åldersgruppen 55-64 per 1000 invånare.

Ålder6per1000. Antal personer i åldersgruppen 65+ per 1000 invånare.

Vi har också bestämt att standardisera samtliga variabler. Standardisering gör att alla prediktorer hamnar på samma skala, och regressionskoefficienternas storlekar är direkt jämförbara sinsemellan. Därför är det enkelt att hitta de starkast förklarande variablerna. Tolkningen blir också annorlunda: allt annat lika så tolkas koefficienten för en given förklarande variabel som den skillnad i responsvariabeln som uppstår då den förklarande variabeln ökar med 1 standardavvikelse bort från medelvärdet.

Vi misstänker att vissa förklarande variabler kan vara korrelerade. Variabeln *Invånarekm2* som motsvarar befolkningstäthet kommer att möjligtvis vara korrelerad med *Avstånd* eftersom generellt befolkningstäthet är högre i centrala kommuner än de som ligger i inre/ytte förort. Dessutom misstänker vi att variablerna *Medelålder* och de åldersmässigt indelade variabler *Ålder1per1000, ..., Ålder6per1000* är korrelerade. Vi gissar också att variablerna *Boroarbetarper1000* och *Utpendlareper1000* är korrelerade med varandra eftersom (bortse från 16+ arbetslösa population) om ena avtar ökar den andra. I kommande avsnitt undersöker vi de samband och korrelationer.

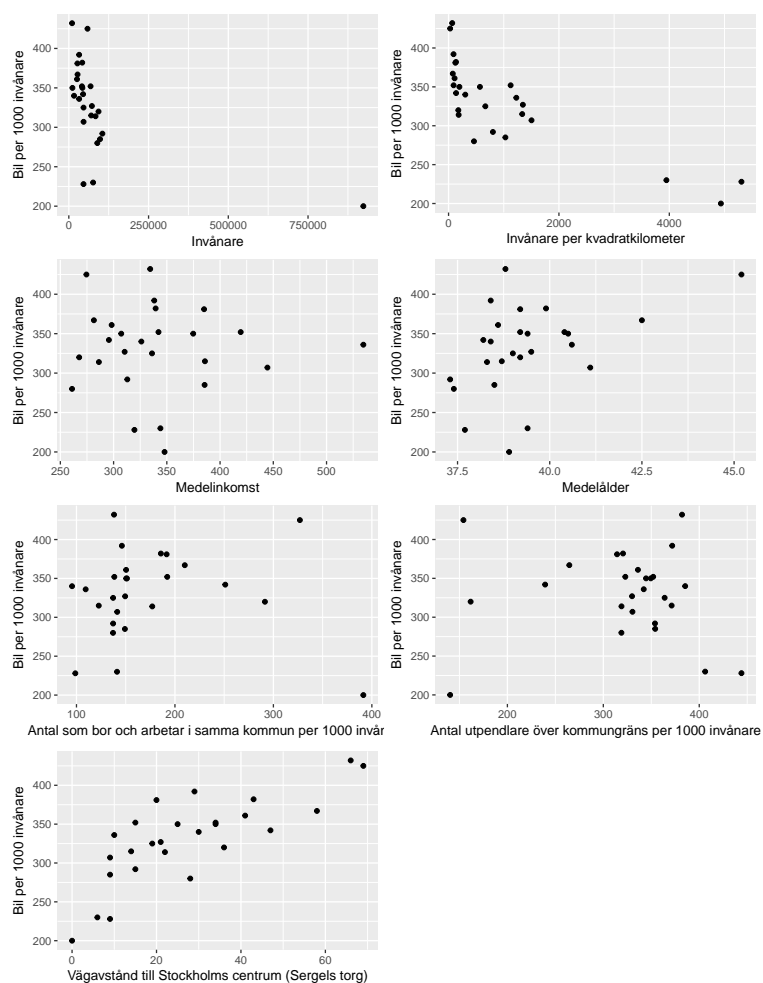
4 Analys av data och modellval

I det här examensarbetet ska vi använda oss av multipel linjär regression för att undersöka vilka variabler som påverkar och beskriver *biltäthet* i kommunerna i Stockholms län. Inledningsvis analyserar vi korrelationsmatrisen för den beroende och samtliga förklarande variabler. Vi tittar på korrelationen mellan de förklarande variabler och responsvariabeln i tabell 4.1. Det visar sig att variabler *Invånarekm2*, *Ålder2per1000* och *Avstånd* har signifikant korrelation med responsvariabeln *Bilper1000* ($r = -0.81$, $r = -0.78$ respektive $r = 0.76$).

| Förklarande variabler | Korrelation med responsvariabeln <i>Bilper1000</i> |
|---------------------------|--|
| <i>Invånare</i> | -0.55 |
| <i>Invånarekm2</i> | -0.81 |
| <i>Medelinkomst</i> | -0.04 |
| <i>Medelålder</i> | 0.25 |
| <i>Boroarbetarper1000</i> | -0.02 |
| <i>Utpendlareper1000</i> | -0.03 |
| <i>Avstånd</i> | 0.76 |
| <i>Bensinpris</i> | -0.04 |
| <i>Ålder1per1000</i> | 0.33 |
| <i>Ålder2per1000</i> | -0.78 |
| <i>Ålder3per1000</i> | -0.10 |
| <i>Ålder4per1000</i> | 0.41 |
| <i>Ålder5per1000</i> | 0.35 |
| <i>Ålder6per1000</i> | 0.33 |
| <i>År</i> | -0.02 |

Tabell 4.1: Korrelation mellan förklarande variabler och responsvariabeln *Bilper1000*.

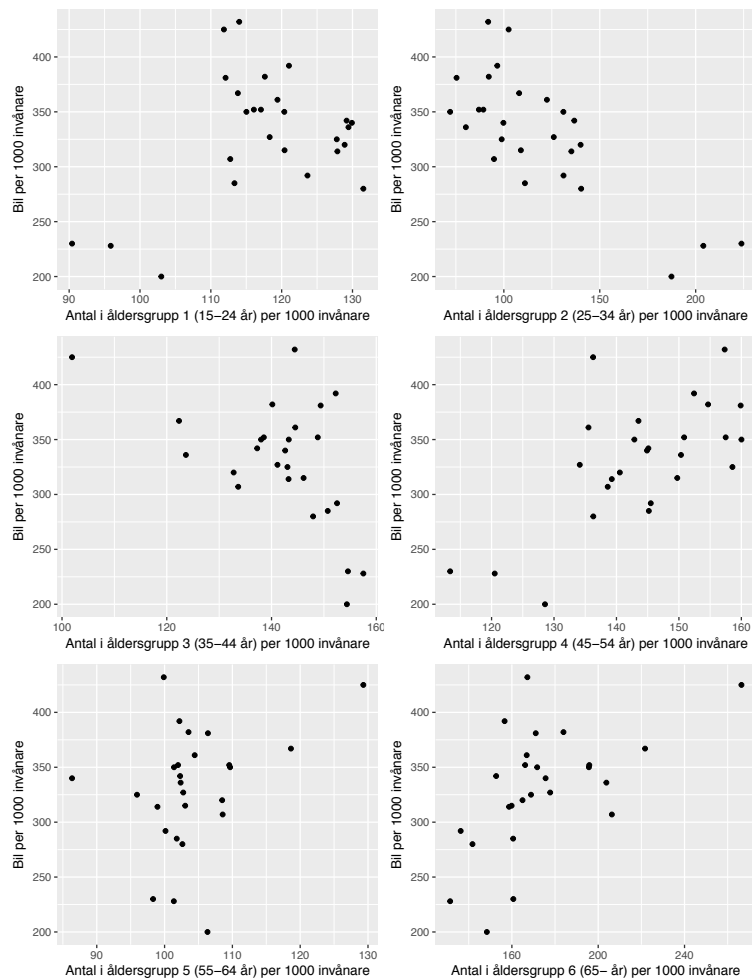
Vidare tittar vi på plot av förklarande variabler mot *Bilper1000* som responsvariabel (för år 2015) vilket är delad i två figurer och visas i Figur 4.1 och 4.2.



Figur 4.1: Plot av förklarande variabler mot *Bilper1000* som responsvariabel år 2015.

I Figur 4.1 för variabeln *Invånare* ser vi en observation som avviker stort från de övriga vilket är Stockholms kommun med högsta antal invånare i länet. För variabeln *Invånarekm2* ser vi tre observationer som avviker från de övriga. De är kommunerna Stockholm, Solna och Sundbyberg som har tydlig avvikelse. Det finns inte en tydlig samband mellan *Invånare* och *Bilper1000* medan *Invånarekm2* verkar följa ett nedåtgående mönster.

För variabeln *Medelinkomst* ser man inte en tydlig linjär samband. För *Medelålder* ser man ett positivt samband. För variabelerna *Boroarbetarper1000* och *Utpendlareper1000* kan vi inte se ett tydligt samband. För variabeln *Avstånd* ser man att det finns ett tydligt positivt samband.



Figur 4.2: Plot av förklarande variabler mot *Bilper1000* som responsvariabel år 2015.

I Figur 4.2 verkar variablerna i åldersgrupper 2 och 3 följa ett nedåtgående mönster. För andra åldersgrupper 4 kan man inte se något tydligt samband.

Resultatet av Tabell 4.1 och Figur 4.1 och 4.2 visar att det finns ett tydligt samband mellan de förklarande variablerna *Invånarekm2*, *Ålder2per1000* (negativ korrelation) och responsvariabeln *Bilper1000*. Dessutom finns det positivt samband mellan *Avstånd* (vägavstånd i kilometer från motsvarande kommuns centrum till Stockholms centrum) och *Bilper1000*. Sambandet mellan responsvariabeln och andra förklaringsvariabler är inte tydligt. Vi noterar de tre kommunerna Stockholm, Solna och Sundbyberg (punkter med lägsta y-värde) avvikelser från övriga kommunerna och dess påverkan på resultatet.

4.1 Multikollinearitet

Vidare kollar vi hur det ligger till med korrelationen mellan våra förklarande variabler. Som vi nämnde i förra avsnitt misstänker vi att variablerna *Invånare* och de åldersmässigt indelade variabler *Ålder1per1000*, ..., *Ålder6per1000* är korrelerade. Dessutom gissar vi att variablerna *Utpendlareper1000* och *Boroarbetarper1000* är korrelerade med varandra.

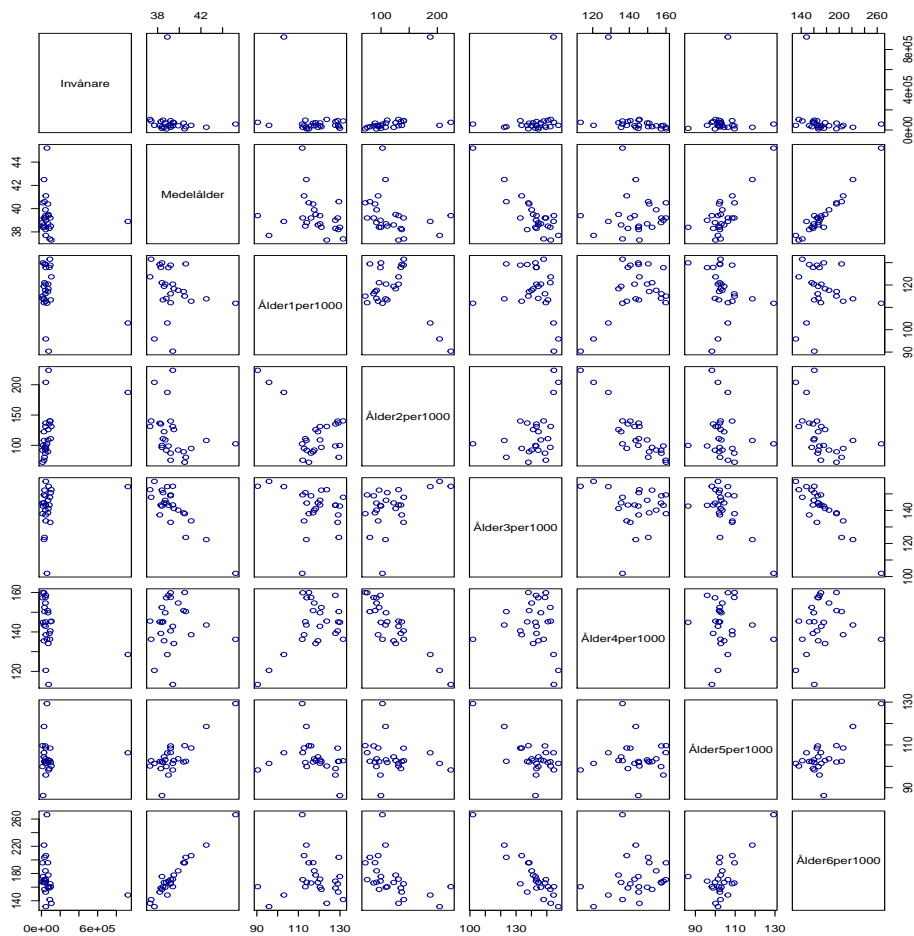
Vi börjar med att göra en multipel regression med samtliga förklarande variabler (Summary i Figur 6.1 i bilaga) och tittar närmare på VIF-värdet för dessa variabler. Resultatet sammanställs i Tabell 4.2.

| Variabel | VIF-värde |
|---------------------------|-----------|
| <i>Invånare</i> | 6.32 |
| <i>Invånarekm2</i> | 14.82 |
| <i>Ålder1per1000</i> | 5.25 |
| <i>Ålder2per1000</i> | 82.11 |
| <i>Ålder3per1000</i> | 23.91 |
| <i>Ålder4per1000</i> | 28.27 |
| <i>Ålder5per1000</i> | 65.92 |
| <i>Ålder6per1000</i> | 442.37 |
| <i>Medelålder</i> | 367.73 |
| <i>Medelinkomst</i> | 6.51 |
| <i>Utpendlareper1000</i> | 41.03 |
| <i>Boroarbetarper1000</i> | 44.98 |
| <i>Avstånd</i> | 4.08 |
| <i>Bensinpris</i> | 3.71 |
| <i>År</i> | 8.54 |

Tabell 4.2: VIF-värdet för full modell.

Från Tabell 4.2 kan vi se att VIF-värdet för alla variabler förutom *Avstånd* och *Bensinpris* är högre än 5. Detta är en signal för multikollinearitet.

Genom att plotta de olika förklarande variablerna mot varandra kan vi också upptäcka multikollinearitet. I figur 4.3 ser vi plot av variabler *Invånare*, *Medelålder* samt olika åldersgrupper per 1000 invånare. Det verkar att det finns ett tydligt samband mellan *Ålder6per1000* och de två förklaringsvariablerna *Medelålder* och *Ålder3per1000*.

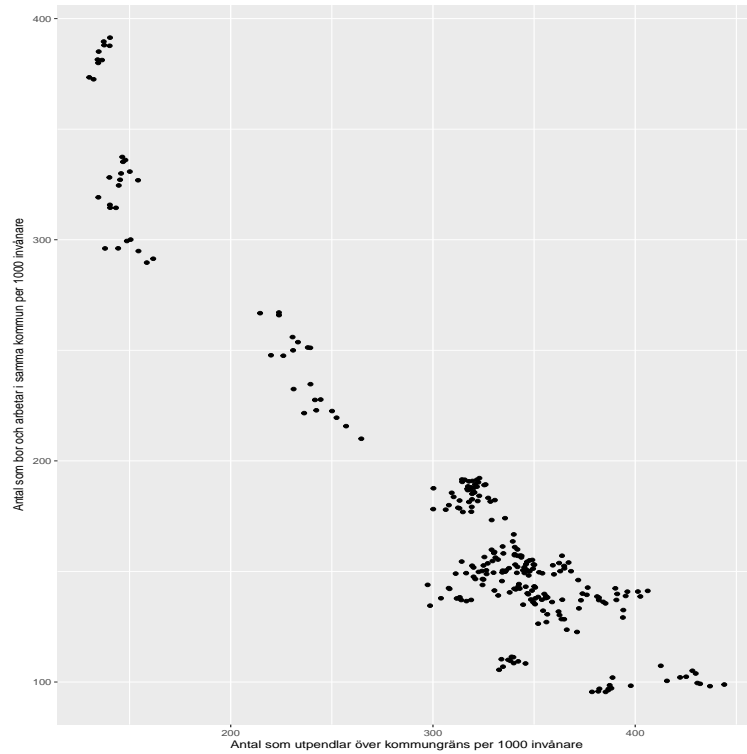


Figur 4.3: Variablerna *Medelålder*, *Ålder1per1000*, *Ålder2per1000*, *Ålder3per1000*, *Ålder4per1000*, *Ålder5per1000* och *Ålder6per1000* mot varandra år 2015.

Vi vill eliminera den/de förklarande variabler som är mest korrelerad med de andra. *Medelålder* är den variabel som påverkas av alla de åldersmässigt indelade variabler *Ålder1per1000*, ..., *Ålder6per1000*. Därför väljer vi att eliminera variabeln *Medelålder*. Variabeln *Invånare* också har ett tydligt samband med alla de åldersmässigt indelade variabler *Ålder1per1000*, ..., *Ålder6per1000* och vi väljer att eliminera *Invånare* också.

De två andra variablerna *Boroarbetarper1000* och *Utpendlareper1000* är starkt korrelerade ($r = -0.95$). Dessutom är *Boroarbetarper1000* och *Utpendlareper1000* var för sig svagt korrelerade med responsvariabeln *bilper1000* ($r = -0,02$ och $r = -0,03$). plot av variablerna mot varandra i figur 4.4 visar också ett tydligt samband. Vi väljer att behålla variabeln *Utpendlare* och

eliminera variabeln *Boroarbetarper1000*.



Figur 4.4: Variablen *Utpendlareper1000* mot *Boroarbetarper1000* år 2006-2015.

Så efter analys av multikollinearitet kommer vi fram till en reducerad modell där de korrelerade variablerna *Invånare*, *Medelålder* och *Boroarbetarper1000* är eliminerade. Vi gör som tidigare en multipel regression. Resultatet av VIF-värdet visas i Tabell 4.3. och vi ser att VIF-värdena har blivit mycket lägre. Det här modell kallar vi modell 1. Modell 1 har $AIC= 2212.089$.

| Variabel | VIF-värde |
|--------------------------|-----------|
| <i>Invånarekm2</i> | 8.92 |
| <i>Ålder1per1000</i> | 4.68 |
| <i>Ålder2per1000</i> | 23.25 |
| <i>Ålder3per1000</i> | 13.29 |
| <i>Ålder4per1000</i> | 5.26 |
| <i>Ålder5per1000</i> | 7.81 |
| <i>Ålder6per1000</i> | 11.15 |
| <i>Medelinkomst</i> | 6.47 |
| <i>Utpendlareper1000</i> | 1.81 |
| <i>Avstånd</i> | 3.71 |
| <i>Bensinpris</i> | 3.21 |
| <i>År</i> | 7.18 |

Tabell 4.3: VIF-värdet för modellen utan variablerna *Invånare*, *Medelålder* och *Boroarbetarper1000*.

4.2 Modellval

I det här avsnitt ska vi använda oss av samtliga metoder för variabelselektion från avsnitt 2.7.

4.2.1 Modell baserad på Backward elimination

Baserad på proceduren Backward elimination gör vi en multipel regression med alla variabler. Sedan eliminerar vi den variabeln med högre p -värdet. Proceduren fortsätter tills alla variabler som är kvar är signifikanta.

Vi börjar med modell 1 som vi har från förra delen. Bland de alla icke-signifikanta variabler har *Ålder5per1000* högsta p -värdet. Vi eliminerar det och får modell 2. Nästa steg eliminera vi variabel *Ålder2per1000* som har högst p -värdet och bildar vi modell 3. I modell 3 är alla variabler förutom *Ålder1per1000* signifikanta. Vi eliminerar *Ålder1per1000* och bildar modell 4. Proceduren slutar här eftersom alla variabler är statistisk signifikanta på 5 procent signifikansnivå.

Baserad på proceduren Backward elimination har vi fått modell 4 med 9 predikator. Variablerna som visar sig vara statistisk signifikanta på 5 procent signifikansnivå är *Invånarekm2*, *Medelinkomst*, *Utpendlareper1000*, *Avstånd*, *Ålder3per1000*, *Ålder4per1000*, *Ålder6per1000*, *År* och *Bensinpris*.

4.2.2 Modell baserad på Forward selection

I denna procedur utgår från en modell utan variabel och adderar en variabel i taget till modellen. Vi börjar med variabel som har högsta korrelation med responsvariabeln vilket är *Invånarekm2* ($r = -0.81$). Sedan i varje steg lägger vi en variabel som är mest signifikant i test av om motsvarande parameter är noll (lägst p -värde) vilket blir variabeln *Avstånd* i första steg. Vi noterar att ibland i ett steg finns det flera variabler som kan väljas eftersom de alla är signifikanta och vi väljer den variabel som ger minst ökning i förklaringsgraden. Det visar sig att Forward selection proceduren ger samma resultat som Backward elminiation proceduren.

4.2.3 Modell baserad på AIC kriterium

Som vi beskrev i avsnitt 2.7 är Akaike's Information Criterion (AIC) ett annat jämförelsemått som kan användas för att bestämma lämpligaste modell för att beskriva data. I programmet R finns en funktion `stepAIC` från MASS paketet vilket använder AIC som en kriterium. Vi väljer att använda denna funktion. Det visar sig att den här funktion ger samma resultat som de två förra metoderna.

4.2.4 Vald modell

Så baserad på alla tre metoder fick vi en modell med 9 predikator och resultatet av regressionsmodellen visas i tabell 3. Variablerna som visar sig vara statistisk signifikanta på 5 procent signifikansnivå är *Invånarekm2*, *Medelinkomst*, *Utpendlareper1000*, *Avstånd*, *Ålder3per1000*, *Ålder4per1000*, *Ålder6per1000*, *År* och *Bensinpris*.

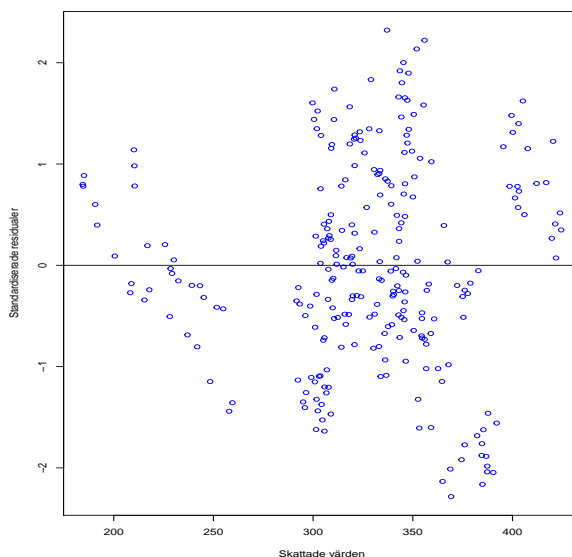
Innan vi ska visa resultatet av parameterskattningarna ska vi använda oss av residualplottar för att bedöma huruvida modellen är lämplig. Detta görs för att kolla om modell antagandena för linjär regression är uppfyllda.

5 Modellvalidering, resultat och diskussion

5.1 Modellvalidering

För att undersöka om den valda modellen i förra avsnittet är lämplig kan man titta närmare på olika residualplottar och kolla om modell antaganden för linjär regression är uppfyllt. Annars kan vi inte dra något korrekt slutsats om modellen inte är lämplig.

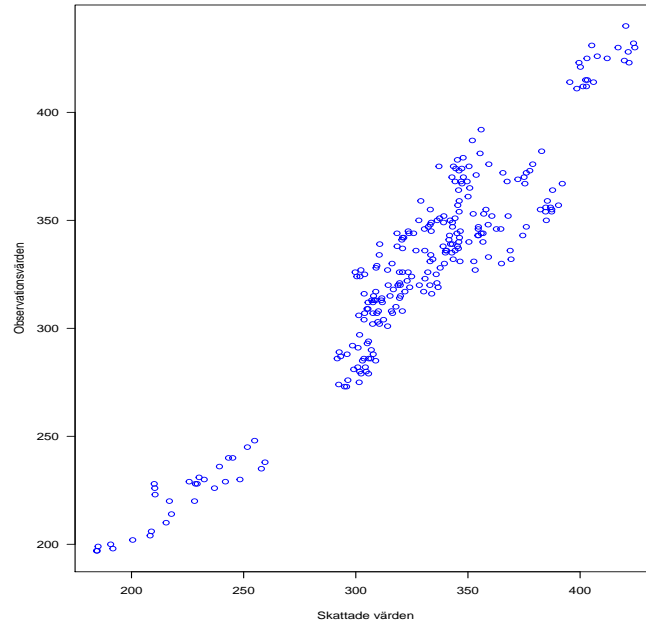
Vi börjar att plotta residensualerna mot de skattade värdena samt förklarande variabler. Detta görs för att kontrollera antagandena om att residualerna har väntevärde 0 och konstant varians (homoskedasticitet). Istället för residualerna väljer vi standardiserade residualer genom att dela de med sina skattade standardavvikelse.



Figur 5.1: Residualerna plottad mot skattade värden.

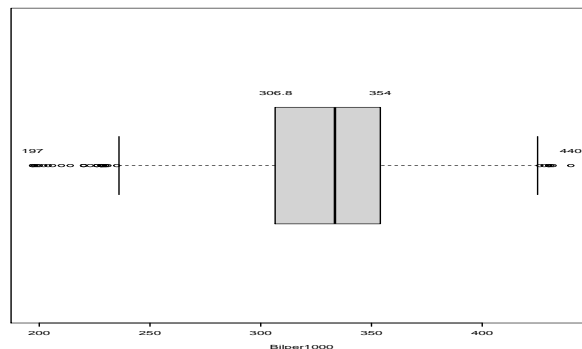
Figur 5.1 visar residualernas plot mot skattade värden. Residualernas ser ut vara utspridda runt 0. Men de är mer utspridda i mitten och mindre utspridda på kanterna vilket kan betyda att variansen inte är konstant. De flesta punkterna har samlas mellan skattade värdena 300 till 350. Detta leder till mer undersökning av responsvariabeln.

Vi plottar också y-värdena (observationer) mot skattade värden vilket visas i Figur 5.2. Det verkar finnas likadana mellanrum i plottade observationsvärdena mot skattade värdena.



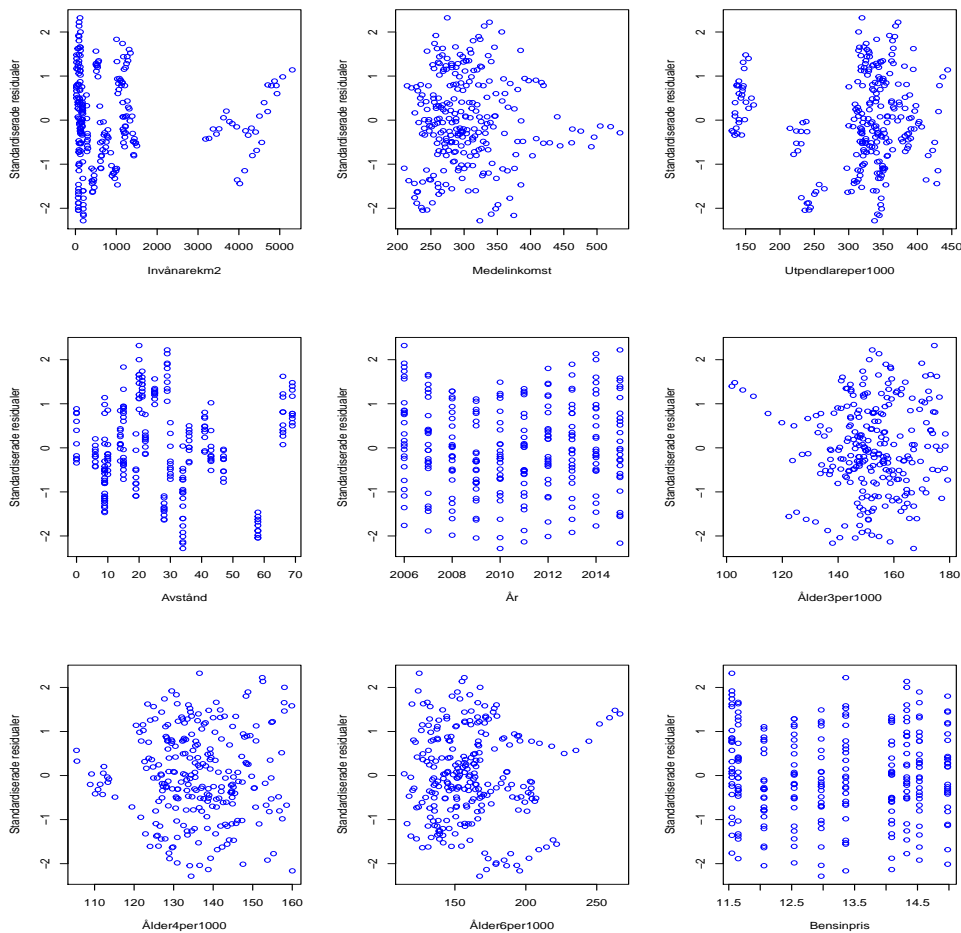
Figur 5.2: Observationsvärden plottad mot skattade värden.

Vi kollar responsvariabelns utspridning genom att plotta dessa lådagram i Figur 5.3. Som vi ser är responsvariabelns övre och under kvartil 306.8 respektive 354 vilket inte är ett stort intervall. Responsvariabelns värden är inte jämnt utspridda och det finns flera extremvärden. Detta är kommunerna med minsta eller högsta biltäthet. Detta är en förklaring för mellanrummet som finns i residualernas plott mot skattade värden i Figur 5.2. Residualplotten skulle vara mer jämnt utspridda om vi hade jämnare responsvariabelvärden.



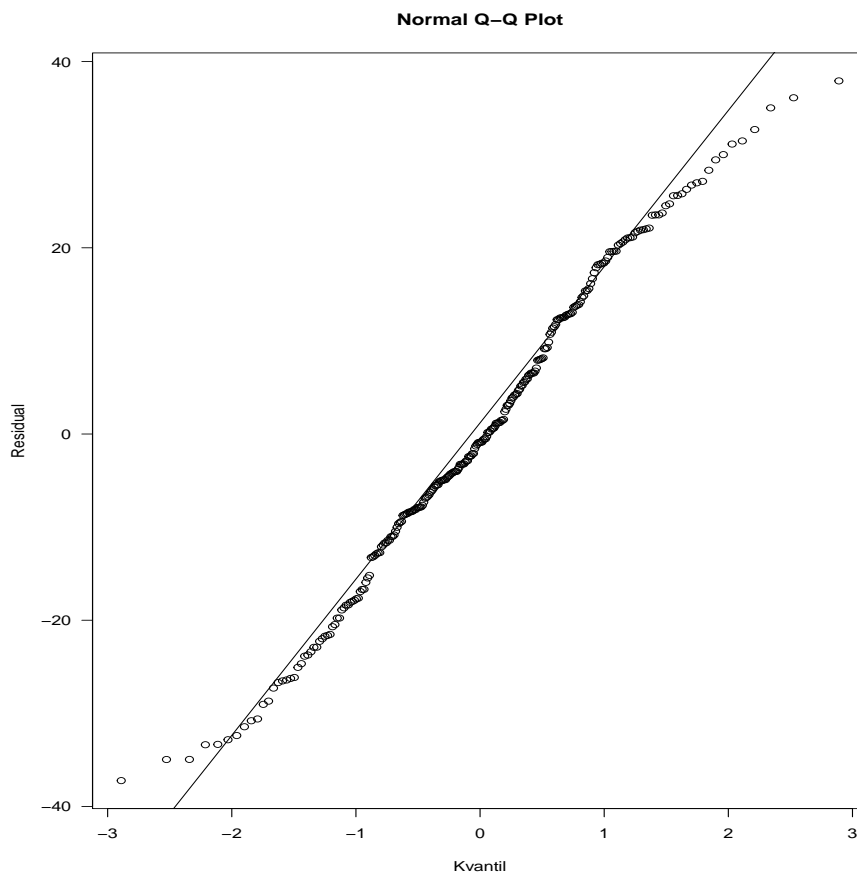
Figur 5.3: Lådagram för responsvariabel *Bilper1000*.

Figur 5.4 visar residualernas plot mot förklarande variabler. Residualernas ser ut vara utspridda runt 0 och residualernas variation tycks vara konstant.



Figur 5.4: Residualerna plottad mot förklarande variabler.

Dessutom måste vi undersöka om antagandet om feltermernas fördelning är uppfyllt, det vill säga att residualerna ska vara approximativt normalfördelade med väntevärde 0. För att undersöka detta använder vi R-funktion QQ-plot där punkterna ska följa en rät linje approximativt, vilket de gör enligt Figur 5.5



Figur 5.5: Normalfördelningsplott.

Genom att kolla olika antaganden drar vi slutsats att våra modell är lämplig och vi kommer att presentera resultatet i nästa del.

5.2 Resultat

Nu ska vi titta närmare på resultatet av våra modell. Variablerna som visar sig vara statistisk signifikanta på 5 procent signifikansnivå är *Invånarekm2*, *Medelinkomst*, *Utpendlareper1000*, *Avstånd*, *Ålder3per1000*, *Ålder4per1000*, *Ålder6per1000*, *År* och *Bensinpris*.

Vi får följande samband mellan *Bilper1000* i valda modellen (Summary i Figur 6.2 i bilaga):

$$\begin{aligned}
 \text{Bilper1000} = & 326.527 - 18.617 \text{ Invånarekm2} + 9.233 \text{ Medelinkomst} + 7.104 \\
 & \text{Utpendlareper1000} + 7.104 \text{ Avstånd} + 9.476 \text{ Ålder3per1000} + 10.397 \text{ Ålder4per1000} \\
 & + 15.308 \text{ Ålder6per1000} - 6.094 \text{ År} - 6.545 \text{ Bensinpris}.
 \end{aligned}$$

Lutningskoefficienterna är som vi förväntade oss. Tolkning av regressionskoefficienter är att bland de alla faktorer har avstånd störst inverkan på biltäthet. Faktorerna medelinkomst, antal förvärvsarbetande utpendlare per 1000 invånare, avstånd, antal i åldersgrupperna 3, 4 och 6 har positivt effekt på biltäthet. Däremot har befolkningstäthet, år (tid) och bensinpris negativ effekt på biltäthet.

Parameterskattningar och förklaringsgraden för den slutliga modellen redovisas i Tabell 5.1.

| Variabel | Parameterskattning | p -värde |
|--------------------------|--------------------|----------------------|
| <i>Intercept</i> | 326.527 | < 0.001 |
| <i>Invånarekm2</i> | -18.617 | < 0.001 |
| <i>Ålder3per1000</i> | 9.476 | < 0.001 |
| <i>Ålder4per1000</i> | 10.397 | < 0.001 |
| <i>Ålder6per1000</i> | 15.308 | < 0.001 |
| <i>Medelinkomst</i> | 9.233 | < 0.001 |
| <i>Utpendlareper1000</i> | 7.104 | < 0.001 |
| <i>Avstånd</i> | 30.186 | < 0.001 |
| <i>Bensinpris</i> | -6.545 | < 0.001 |
| <i>År</i> | -6.094 | < 0.001 |
| $\hat{\sigma} = 16.54$ | $R^2 = 0.9003$ | $R_{adj}^2 = 0.8967$ |

Tabell 5.1: Parameterskattningar i slutliga modellen.

Vi kan titta på VIF-värdena i tabell 5.2 vilket visar att alla variablerna har fått VIF-värdet under 5.

| Variabel | VIF-värde |
|--------------------------|-----------|
| <i>Invånarekm2</i> | 2.54 |
| <i>Ålder3per1000</i> | 4.74 |
| <i>Ålder4per1000</i> | 2.64 |
| <i>Ålder6per1000</i> | 4.99 |
| <i>Medelinkomst</i> | 3.28 |
| <i>Utpendlareper1000</i> | 1.668 |
| <i>Avstånd</i> | 3.18 |
| <i>Bensinpris</i> | 2.97 |
| <i>År</i> | 4.23 |

Tabell 5.2: VIF-värdet för slutliga modellen.

5.3 Diskussion

I det här examensarbete har vi undersökt vilka variabler som beskriver den kommunala variation som förekommer i Stockholms läns fordonsstatistik. Resultat kan sammanfattas så här:

Bland alla faktorer har vägvstånd till Stockholms centrum (Sergels torg) störst inverkan på biltäthet. Faktorerna medelinkomst och antal förvärvsarbetande utpendlare per 1000 invånare har stark positiv påverkan på biltäthet. Dessutom har åldersmässigt fördelade variabler olika påverka på biltäthet. Det visade sig att åldersgrupp 3 (35-45), åldersgrupp 4 (45-54) och åldersgrupp 6 (65+) har signifikant och positivt effekt på biltäthet. Däremot har befolkningstäthet, utveckling i tid (från år 2006 till 2015) och bensinpris negativ effekt på biltäthet.

I resultatet kom vi bland annat fram till att befolkningstäthet förklarar en stor del av variationen i biltäthet mellan olika kommuner. Detta verkar vara rimligt av olika skäl. En orsak är att kollektivtrafik körs oftare i befolkningstäta områden vilket gör att vissa beslutar sig för att inte köpa bil. Den andra orsaken kan vara tillgång till parkeringsplatser som är svårare för boende i tätbefolkade området. I vårt modell kom vi fram att avstånd har största påverkan på biltäthet vilket återigen kan förklaras med tillgången till lokaltrafik samt svårigheten att finna parkering i innerstaden. Av de olika åldersgrupper som vi inkluderade i vårt studie hade åldersgrupperna 3, 4 och 6 starkare påverkan än resterande åldersgrupper. Detta är intressant, men jag ser ingen självklar anledning till det.

Undersökningen skulle förbättras med att inkludera de variabler som också kan anses ha stor påverkan på biltäthet men det fanns inte tillräcklig data för det. Bland de mest relevanta faktorer är bebyggelse och hur tätt olika kommuner är byggda. Antal flerbostadshus och småhus i området skulle möjligen förklara variation mellan kommunernas biltäthet stort.

6 Bilaga

```
Residuals:
  Min      1Q  Median      3Q      Max
-37.791 -9.691 -0.482  9.647 32.573

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    326.5269    0.8844 369.211 < 2e-16 ***
std_Invånare    -4.2432    2.2284  -1.904 0.058062 .
std_Invånarekm2 -13.1578    3.4115  -3.857 0.000147 ***
std_Medelinkomst  7.1853    2.2605   3.179 0.001671 **
std_Medelålder  -31.8138   16.9921  -1.872 0.062365 .
std_Boroarbetarper1000 55.1248    5.9428   9.276 < 2e-16 ***
std_Utpendlareper1000 58.9202    5.6757  10.381 < 2e-16 ***
std_Avstånd     25.4798    1.7902  14.233 < 2e-16 ***
scale(ar)      -11.0622    2.5900  -4.271 2.79e-05 ***
std_Ålder1per1000  2.5454    2.0295   1.254 0.210972
std_Ålder2per1000 -11.0448    8.0293  -1.376 0.170219
std_Ålder3per1000  -9.1688    4.3330  -2.116 0.035354 *
std_Ålder4per1000  5.7680    4.7111   1.224 0.222004
std_Ålder5per1000  3.2945    7.1943   0.458 0.647405
std_Ålder6per1000 36.5390   18.6370   1.961 0.051068 .
std_Bensinpris  -9.9248    1.7060  -5.817 1.87e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.26 on 244 degrees of freedom
Multiple R-squared:  0.9277,    Adjusted R-squared:  0.9232
F-statistic: 208.7 on 15 and 244 DF,  p-value: < 2.2e-16
```

Figur 6.1: Modellen med samtliga förklaringsvariabler


```

Residuals:
    Min       1Q   Median       3Q      Max
-37.225 -10.123  -0.895  12.512  37.923

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    326.527     1.026 318.237 < 2e-16 ***
std_invkm2     -18.617     1.639 -11.359 < 2e-16 ***
std_medink       9.233     1.863  4.957 1.32e-06 ***
std_utpendlper1000 7.104     1.334  5.327 2.23e-07 ***
std_avstand     30.186     1.832 16.479 < 2e-16 ***
scale(ar)      -6.094     2.114 -2.883 0.004287 **
std_a3per1000   9.476     2.238  4.235 3.22e-05 ***
std_a4per1000  10.397     1.671  6.221 2.07e-09 ***
std_a6per1000  15.308     2.296  6.667 1.66e-10 ***
std_bpris      -6.545     1.771 -3.696 0.000269 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.54 on 250 degrees of freedom
Multiple R-squared:  0.9003,    Adjusted R-squared:  0.8967
F-statistic: 250.8 on 9 and 250 DF,  p-value: < 2.2e-16

```

Figur 6.2: Slutliga modellen

7 Referenser

Andersson, P. Tyrcha, J., Notes in Econometrics, 2016.

Beesley, M. E. Kain, J. F. (1964) Urban form, car ownership and public policy: an appraisal of 'traffic in towns?' Urban Studies , 1(2), pp. 174203.

Fahrmeir, L., Kneib, T., Lang, S. Marx, B. (2013). Regression [Elektronisk resurs] : Models, Methods and Applications. Berlin, Heidelberg: Springer Berlin Heidelberg.

Jones, S. R. Tanner, J. C. (1979) Car ownership and public transport, TRRL Report 464, Transport and Road Research Laboratory, Crowthorne.

Matstoms, Pontus. (2002) Modeller och prognoser för regionalt bilinnehav i Sverige. VTI rapport 476.

SCB statistikdatabasen: [Elektronisk resurs] Statistical database. (2006-2015). Stockholm: Statistiska centralbyrån.

Sundberg, R., Lineara Statistiska Modeller, 2016

Svenska Petroleum och Biodrivmedel Institutet (SPBI) [Elektronisk resurs]: Utveckling av försäljningspris för bensin, dieselbränsle och etanol

Vagland, Å. Pyddoke, R. (2006). Hur hushållen anpassar sig till ändrade kostnader för bilinnehav och bilanvändning [Elektronisk resurs] : essä och kunskapsöversikt. Linköping: VTI., VTI rapport 545.