

Patientupplevd vårdkvalité utifrån olika bakgrunder

Författare: Åsa Hjort Hedenberg

Ht 2013



Det är vår egen upplevelse av vår hälsa och vår ålder som har störst betydelse för vad vi tycker om vårdkvalitén. De yngre är betydligt mer missnöjda med vården än äldre. De som upplever sin hälsa som dålig är också mycket mer missnöjda än de som anser sig ha mycket god hälsa. Det har oftast inte någon betydelse om vi besöker vårdcentralen många eller få gånger under en period eller vilken utbildningsnivå den som söker vård har. Kön har en viss betydelse. Män är mer nöjda än kvinnor. Om patienten har svenska som modersmål eller inte har en ganska stor betydelse, för vad vi tycker om vårdkvalitén inom Primärvården, VGR. De som inte har svenska som modersmål är mer missnöjda med vården.

**Göteborgs Universitet
Handelshögskolan
Nationalekonomi med Statistik
Handledare: Kristofer Månsson
C-kurs i Statistik
15 Hp**

INNEHÅLL

1. SAMMANFATTNING.....	s.4
2. BAKGRUND.....	s.5
<i>Bakgrundsvariablerna (oberoende variabler)</i>	s.7
<i>Beroende variabler</i>	s.7
3. SYFTE OCH MÅLSÄTTNING.....	s.8
4. STATISTISK BAKGRUND.....	s.8
5. PARVISA JÄMFÖRELSE R MELLAN OBEROENDE VARIABLER.....	s.9
<i>Egenupplevd hälsa</i>	s.9
<i>Många besök på VC</i>	s.9
<i>Olika åldersgrupper</i>	s.9
<i>Modersmål</i>	s.10
<i>Utbildningsnivå</i>	s.10
6. VAL AV METOD.....	s.10
<i>Användning av olika modeller</i>	s.13
<i>Ordinal logistisk regression</i>	s.13
<i>Binär logistisk regression</i>	s.14
<i>Logistisk LASSO regression</i>	s.14
7. REDOVISNING AV RESULTAT / BILAGOR	s.15
<i>BILAGA 1; Parvisa jämförelser mellan oberoende variabler</i>	s.16
<i>BILAGA 2; Binär logistik regression</i>	s.28
<i>BILAGA 3; LASSO logistisk regression</i>	s.35
8. SLUTSATSER.....	s.49
9. REFERENSER.....	s.50
10. ETT PRAKTISKT RÄKNEEXEMPEL.....	s. 51

Ett särskilt tack till Linus Schiöler på Akademistatistik, som var så hjälpsam då vi skulle köra Logistisk LASSO i mjukvaran "R".

1. SAMMANFATTNING.

I uppsatsen har jag undersökt om och hur mycket olika bakgrundsfaktorer hos patienter påverkar hur nöjda de är med vårdkvaliteten. Uppsatsen begränsar sig till läkarbesök inom den offentliga Primärvården (Närhälsan) i VGR.

Vi har använt oss av två olika metoder för att analysera data från Nationell Patientenkät från hösten 2011. Dessa är Binär logistisk regression och LASSO logistisk regression. Vi har valt logistisk regression därför att den beroende variabeln inte har intervallskala. Resultaten från bearbetningar i mjukvaran SAS, från binär logistisk regression och från "R" LASSO regressionen har varit överensstämmande.

De koefficienter som är signifikanta har redovisats och i kombination med givna formler kan vi om man önskar göra en prediktering. Resultaten från Binär logistisk regression ger de mer "rättvisande" koefficienterna, medans koefficienterna eller redovisade modeller från LASSO är bäst lämpade att använda sig av vid prediktering, eftersom de krympta parametrarna ger en bättre precisering, vid prediktion.

Det är vår egen upplevelse av vår hälsa och ålder som har störst betydelse för vad vi tycker om vårdkvaliteten. De som upplever sin hälsa som dålig är mycket mer missnöjda än de som anser sig ha mycket god hälsa. De yngre är betydligt mer missnöjda med vården än äldre personer. Om patienten har svenska som modersmål eller inte har en ganska stor betydelse. De som inte har svenska som modersmål är mer missnöjda med vården än de med svenska som modersmål. Det har oftast inte någon betydelse om vi besöker Vårdcentralen många eller få gånger eller vilken utbildningsnivå den vårdsökande har. Kön har en viss betydelse. Män är mer nöjda än kvinnor.

Hur skiljer sig de två metoderna från varandra? Med Binär logistisk regression använder vi oss av vanlig Maximum Likelihood skattning. Se sid 28, bilaga 2. Med denna metod redovisas de mer "rättvisande" koefficienterna, standardavvikelse samt om de är signifikanta eller inte. Endast de signifikanta koefficienterna har tagits med. Vi har också använt oss av LASSO logistisk regression, se sid 35, bilaga 3. Med LASSO görs en Maximum likelihood skattning av de krympta parametrarna. I nästan samtliga fall har LASSO koefficienterna ett värde närmare 0 än de koefficienter vi får med binär logistisk regression. Med LASSO ökar bias i koefficienter men variansen minskar. Fördelen med LASSO blir då att koefficienterna är bättre lämpade för prediktering. I denna undersökning med data från Nationell Patientenkät, har vi en mycket stor mängd data. Vi har då en fördel av LASSO som väljer ut åt oss de variabler som har störst påverkan på y (nöjd patient). Vi kan slopa de mindre viktiga koefficienterna. En nackdel med LASSO och även andra krympningsmetoder är att det är svårt att beräkna standardavvikelse på krympta koefficienter, då krympningen leder till betydande bias. Genom Cross-validering eller Bootstrapping, kan vi genom repetitiva körningar, testa delmängder mot delmängder, få fram koefficienter men tyvärr bara en approximativ bild av variansen kring koefficienterna, p g a betydande bias.

Ju lägre lambda värde, desto högre värde på t , $L1$ norm, (se sid 35). Ju högre värde på t , se $L1$ Norm, (se till höger i figuren) desto fler koefficienter blir kvar i modellen och krympningen är mindre. När lambda värdet är högre, blir t värdet lägre, krympningen större och fler koefficienter antar värdet 0.

2. BAKGRUND

Det finns ett stort fokus idag på att Hälso- och sjukvård skall vara jämlik. För några decennier sedan talades det istället om att vården skall vara jämställd. Begreppet har vidgats; det skall inte spela någon roll vilken ålder, kön, social bakgrund, sexuell läggning eller födelseort man har. Enligt svensk lag skall vård ges på lika villkor för hela befolkningen.

Det jobbas på många olika håll för jämlik vård. Socialstyrelsen har som en del av sin definition av god vård att den skall vara jämlik. Regionen arbetar för jämlik vård genom KJV, *Kunskapscentrum för jämlik vård* och i Regionens årsredovisning kan vi läsa att ”hälsa inte skall vara beroende social eller geografisk hemvist eller ekonomi”. Även Regeringen (Socialdepartementet) och SKL (Sveriges Kommuner och Landsting) har under 2013 ingått en överenskommelse (dnr 11/6052) om ”Fortsatta insatser för att främja en mer jämlik hälso- och sjukvård, 2013”. I överenskommelsen kan vi läsa att ”det är idag känt att vården inte erbjuds eller tillgängliggörs på lika villkor för alla.” På Socialstyrelsens hemsida uttrycks att jämlik vård inte betyder att alla skall ha samma vård eller lika mycket, utan att varje person skall få vård och behandling utifrån sina behov.

Hur är det då ? Exempel på ojämlikheter i hälsa och vård

*”Dödligheten i förebyggbara och behandlingsbara sjukdomar är betydligt högre hos personer med enbart grundskoleutbildning jämfört med högskoleutbildade” (Hälso- och sjukvårdsrapport, Socialstyrelsen, 2009).

*”Utlandsfödda upplever att tillgängligheten till vården är sämre och de har lägre förtroende för sjukvården än svenskfödda” (Hälso- och sjukvårdsrapport, Socialstyrelsen, 2009).

*”Högutbildade med diagnosen epilepsi behandlas oftare av specialist jämfört med grundskoleutbildade” (Hälso- och sjukvårdsrapport, Socialstyrelsen, 2009).

*”Personer med psykiatrisk diagnos dör oftare i en behandlingsbar hjärt- och kärlsjukdom” (Öppna jämförelser och utvärdering av psykiatrisk vård, Socialstyrelsen, 2010).

Enligt Malena Lau, KJV, är risken större att behandlas med äldre mediciner om du är lågutbildad.

Det samlas in enormt mycket data inom t ex VGR. Det hade varit intressant att göra fler analyser av den data som samlas in. Exempelvis så har angivits i Sahlgrenska Universitetssjukhusets årsredovisning att kvinnor i snitt väntar 30 minuter längre än män på akutmottagningen. Det hade varit intressant att gå vidare och se om skillnaden i väntetid beror på att männen har andra åkommor än kvinnor, eller är det på att de är kvinnor som de får vänta längre.

Jag har i mitt arbete fått höra att män blir remitterade till specialist medan kvinnor blir sjukskrivna, trots samma diagnos. Det hade varit intressant att undersöka om det är så. Jag har också hört av dem som arbetar med Rehabiliteringsgarantin, att unga kvinnor mycket oftare är långtidssjuka än unga män. Det framkommer i mina resultat, längre fram, att unga kvinnor upplever sig ha mycket sämre hälsa än unga män.

VGR är först i landet med ett politiskt beslut om åtgärder för jämlik hälsa. Se ”Samling för social hållbarhet - åtgärder för jämlik hälsa”. (Beslut av den 24 september 2013, Regionfullmäktige).

Globalt och lokalt uppges skillnaderna öka. I nämnda beslut kan vi läsa att i det område i Göteborg med den högsta medelinkomsten, Saltholmen, jämfört med det området med lägsta medelinkomsten, Bergsjön, skiljer hela nio levnadsår. Det pekas på skillnader i levnadsvanor, skillnader i strukturella och sociala villkor.

Jag tror att det är viktigt med uppföljningssystem. Mig veterligt finns inte något strukturerad uppföljning av jämlik vård innefattande analys av data över tid, i VGR. Det är av vikt med en uppföljning så vi kan följa om arbetet med jämlik vård ger förändringar i praktiken.

Det har varit jättespännande att undersöka om olika bakgrunder hos patienterna, som redovisas i den Nationella Patientenkäten, är olika nöjda med vården. I min analys har jag kunnat se att utlandsfödda är mindre nöjda än svenskfödda. Kvinnor är mindre nöjda än män. De som upplever sig ha en dålig hälsa är betydligt mindre nöjda än de som anser sig ha utmärkt eller mycket god hälsa. Yngre är betydligt mindre nöjda med vården än äldre. Enlig Johan Frisack, Indikator, som genomför de Nationella Patientenkäterna, har det framkommit samma resultat, som de jag fick fram, i England och USA. I min analys kom jag fram till att *egenupplevd hälsa, ålder och om man har svenska eller inte som modersmål är de bakgrundsfaktorer som påverkar mest vad vi tycker om vården.*

Nationella Patientenkäter (NP studier) genomförs i alla Landsting/Regioner i landet, med några få undantag. NP är ett samarbetsprojekt mellan landstingen/Regioner och Sveriges Kommuner och Landsting (SKL). Institutet för kvalitetsindikatorer (Indikator) genomför dessa mätningar av den *patientupplevda vårdkvalitén*. Enkäterna är snarlika men anpassade efter vårdform; primärvård, somatisk öppen- och slutenvård samt psykiatriska öppen- och slutenvården. Enkäterna ser också något olika ut beroende på om de riktar sig till patienter som har besökt läkare eller sjuksköterska. Metoden har varit att i samtliga Landsting/Regioner välja ut fyra konsekutiva veckor. Under denna period gjordes ett slumpässigt urval av patienter som besökt vårdenheten (oftast Vårdcentral) . Det genomsnittliga urvalet per Primärvårdsenhet var 200 patienter. Den totala korrigerade svarsfrekvensen (totalt antal inkomna enkäter/nettourvalet (urval - bortfall)) för hela undersökningen var i genomsnitt 58,5%. Siffran för läkarbesök inom Västra Götalandsregionens (VGR) Primärvård (PV) var 53,4% hösten 2011. Returnerade enkäter i % var betydligt högre bland kvinnor än män.

Ett antal frågor i enkäterna har formulerats med svarsalternativet ”Ej aktuell” eller motsvarande. Det leder till att vi får många saknade värden. Samtliga enkäter med minst en besvarad fråga har inkluderats i undersökningen. Patienter under 40 år är underrepresenterade i resultaten. Dels är denna grupp ”sällanutnyttjare” av sjukvård och dels är svarsfrekvensen låg bland de yngre åldergrupperna.

Denna uppsats begränsar sig till de enkäter som genomförts inom VGR och som riktar sig till patienter som besökt läkare inom PV. Den bygger endast på enkätsvar från höstmätningen 2011 och omfattar 20 357 svar. Berörda frågor av enkäten framgår nedan. Enkäten är uppdelad i ett antal frågeområden såsom; inför besöket, på mottagningen, om läkaren, om provtagningar samt helhetsintryck. Som *y-variabler (beroende variabler)* har jag valt de åtta sista frågorna i enkäten (om helhetsintryck) som får representera de mer detaljerade frågorna i början av enkäten. Som *x-variabler (oberoende variabler)* finns sex bakgrundsvariabler sist i enkäten. Syftet med undersökningen är att påvisa om bakgrundsvariabler påverkar de beroende variablerna.

Bakgrundsvariablerna (oberoende variablerna) är följande

F1 I allmänhet, hur skulle du vilja säga att din hälsa är? 1/ Utmärkt 2/ Mycket bra 3/ Bra 4/ Någorlunda 5/ Dålig

F2 Förutom det här besöket – hur många gånger har du varit i kontakt med sjukvården de senaste sex månaderna? 1/ Ingen gång 2/ En gång 3/ Två eller tre gånger 4/ Fyra eller fler gånger

F3 Vilket år är du född? 1/ 0-15 år 2/ 16-44 år 3/ 45-64 år 4/ 65-74 år 5/ 75 år eller äldre

F4 Är du man eller kvinna? 1/ Man 2/ Kvinna

F5 Är svenska ditt modersmål? 1/ Ja 2/ Nej

F6 Vilken är din högsta avslutade utbildning? 1/ Universitet eller högskola 2/ Gymnasium eller likvärdigt 3/ Grundskola eller likvärdigt

Det som blir intressant att visa är hur dessa bakgrundsvariabler påverkar den patientupplevda vårdkvalitén (beroende variablerna)

De beroende variablerna är följande

E1 Hur värderar du som helhet den vård/ behandling du fick ?

E2 Kände du dig delaktig i beslut om din vård och behandling, så mycket som du önskade?

E3 Kände du att du blev bemött med respekt och på ett hänsynsfullt sätt?

E4 Fick du tillräcklig information om ditt tillstånd?

E5 Fick du veta vart du kunde vända dig om du behövde hjälp eller hade ytterligare frågor efter besöket?

E6 Hur bedömer du att samarbetet mellan personalen fungerade?

E7 Anser du att ditt aktuella behov av sjukvård blivit tillgodosett vid ditt besök på mottagningen?

E8 Skulle du rekommendera den här mottagningen till andra?

3. SYFTE OCH MÅLSÄTTNING

Indikator, som har genomfört undersökningarna, har beräknat ett tal för varje Vårdcentral, som beskriver *den patientupplevda vårdkvalitén*. Detta tal skiljer sig mellan de olika Vårdcentralerna liksom bakgrundsvariablerna. Syftet med denna uppsats är att beräkna om och hur mycket de olika bakgrundsvariablerna påverkar den patientupplevda vårdkvalitén. Tanken är att dessa parametrar ska kunna användas för att *prediktera hur nöjda patienterna på de olika VC i VGR är*. Det blir en slags korrigering, hur nöjda borde patienterna vara utifrån VC's patients specifika bakgrundsvariabler och hur mycket avviker nöjdheten från det förväntade, med tanke på olika bakgrundsfaktorer. Alla som uppsöker en Vårdcentral borde få känna sig lika nöjd med den vård de fått. Min förhoppning är att denna uppsats skall kunna redogöra för vilka grupper som känner sig mer eller mindre nöjda.

4. STATISTISK BAKGRUND

Val av statistisk metod beror på syftet med undersökningen. Man måste fråga sig om avsikten är att genomföra kausala tolkningar av regressionskoefficienterna eller om man vill konstruera prediktionsmodeller.

Om syftet är att tolka koefficienterna bör man ha med alla relevanta bakgrundsvariabler. Då man p.g.a. multikollinearitet, dvs. korrelationer mellan bakgrundsvariablerna, kan bör man ha med alla variabler som påverkar den beroende variabler samt korrelerar med någon eller några av de oberoende variablerna. I praktiken är detta naturligtvis mycket svårt. Det kan finnas variabler som man känner till och som är svårsmätta. Det kan också finnas för oss okända variabler som korrelerar med såväl den beroende variabeln och de oberoende variablerna. Därför bör man vara försiktig med att dra kausala slutsatser från icke-experimentella data.

Vid prediktion är man primärt inte intresserad av att tolka regressionskoefficienterna. En prediktionsmodell kan innehålla variabler som endast indirekt är relaterade till den oberoende variabeln.

Man har kanske utlämnat en relevant bakgrundsvariabel. Om denna har en kraftig korrelation till en annan bakgrunds variabel som inte har en kausal inverkan på den oberoende variabeln kan den sistnämnda bakgrundsvariabeln ändå vara användbar vid prediktion.

Inkluderas för få bakgrundsvariabler vid prediktion riskerar man att få systematiska fel i sin prediktion. Om man tar med irrelevanta variabler ökar man det slumpmässiga felet och därmed minskar precisionen i prediktionen.

Det finns flera metoder för att öka precisionen i prediktion. En sådan metod är lasso som utvecklats av Tibshirani (1996).

Tanken med LASSO är att man introducerar ett bias i prediktionen för att minska det slumpmässiga felet. Detta kan leda till att mean square error vid prediktionen minskar.

LASSO är en s.k. krympningsmetod. En annan sådan metod är ridge regression. Denna krymper beloppet för samtliga koefficienter mot noll, dock i regel utan att några sättes lika med noll. Vid LASSO kommer vissa koefficienter vid krympningen att anta värdet noll. Detta är en fördel om man har många bakgrundsvariabler och man önskar konstruera en enkel prediktionsmodell med få bakgrundsvariabler med goda prediktiva egenskaper.

Här används LASSO för dikotoma variabler p.g.a. brist på programvara för ordinalvariabler. Därför har svaren på de 8 frågorna även dikotomiserats.

I denna uppsats är det 20 357 svar, vilket gör att koefficienterna lätt blir signifikanta. I tabellerna från binär logistisk regression är endast signifikanta parametrar tagits med. Vad gäller framräknade LASSO koefficienter, som inte avviker mycket från de förstnämnda, finns bias. Det blir då svårt att räkna på hur signifikanta dessa är. Alla koefficienter är redovisas vid LASSO regressionen.

5. PARVISA JÄMFÖRELSENER MELLAN OBEROENDE VARIABLER (se Bilaga 1)

Ett bra sätt att få en uppfattning om olika samband är att göra *parvisa jämförelser mellan x variablerna*, de olika bakgrundsfaktorerna. Vi ser om det eventuellt finns något samband. Då det finns många oberoende variabler är *multikollinearitet*, samband mellan de oberoende variablerna, snarare regel än undantag. Om det t ex finns en stark korrelation mellan två oberoende variabler, vet vi egentligen inte om det är den ena eller båda som har ett samband med y. Det kan vara så att en av dessa oberoende variabel inte har något samband alls med y, men att det finns stark korrelation mellan två eller flera oberoende variabler, som leder till att det går att mäta ett samband mellan en oberoende variabel och y, där det egentligen inte finns någon korrelation mellan x och y. I denna uppsats, liksom i enkäten, finns sex bakgrundsvariabler. Det är nästan alltid så att det finns andra bakgrundsvariabler, som vi inte känner till, som korrelerar sinsemellan, vilket ger en osäkerhet, särskilt i icke experimentella undersökningar som denna. Därför skall slutsatser dras med stor försiktighet.

Egenupplevd hälsa

Det kanske kan tyckas överflödigt att nämna, men det finns ett klart samband mellan dålig egenupplevd hälsa och många besök på vårdcentralen. I gruppen som upplever sin hälsa som ”Dålig”, är den andel som besökt vårdcentralen ”fyra gånger eller fler” de senaste sex månaderna 67%. Motsvarande siffra i gruppen som anser sin hälsa vara ”Utmärkt” är 4% (Tabell 1). Likaså finns ett tydligt samband mellan högre ålder och sämre egenupplevd hälsa (Tabell 2). Kvinnor har en sämre egenupplevd hälsa än män. Av männen anger 12 % att de har en ”Utmärkt” hälsa, motsvarande siffra för kvinnor är 9 %. Bland kvinnor uppger 35 % att deras hälsa är ”Någorlunda” eller ”Dålig” medan motsvarande siffra för männen är 28 % (Tabell 3). Det finns också en stor skillnad i egenupplevd hälsa beroende på om man har svenska som modersmål eller inte. Andelen hos de som har svenska som modersmål och uppger sin hälsa som ”Utmärkt” eller ”Mycket bra” är 36 %. Samma andel bland de som inte har svenska som modersmål är 26 % (Tabell 4). Högutbildade mår bättre än de med lägre utbildning. Bland de som har universitetsutbildning upplever 45 % att deras hälsa är ”Utmärkt” eller ”Mycket bra”. Motsvarande siffra för de med gymnasieutbildning är 34 % och för de med grundskoleutbildning 22 % (Tabell 5).

Många besök

Gruppen ”65-74 år” och ”75 år eller äldre” har den största andelen många besök under de senaste sex månaderna (Tabell 6). De kvinnor som besöker vårdcentralen gör det oftare än män (Tabell 7). De som inte har svenska som modersmål besöker också vårdcentralen oftare än de med svenska som modersmål (Tabell 8). De som har universitetsutbildning uppvisar en mindre andel besök där man har varit i kontakt med sjukvården många gånger, jämfört med de med kortare utbildning (Tabell 9).

Olika åldersgrupper

Kvinnor och män skiljer sig åt när i livet de söker vård. Av kvinnornas andel besök sker hela 25 % i åldersgruppen 16-44 år. Männens andel besök i den åldersgruppen är 16 %. 24% av männens alla besök sker i åldersgruppen ”65-74 år”. Motsvarande siffra för kvinnor är 18 %. För övrigt söker kvinnor mer sjukvård än män; hela 61 % av läkarbesöken görs av kvinnor (Tabell 10). Andelen läkarbesök, hos de som inte har svenska som modersmål, är större i de lägre åldrarna. Bland de som har svenska som modersmål är förhållandet det motsatt, då en större andel av besöken sker i

de äldre åldersgrupperna (Tabell 11). Bland de patienter som gör läkarbesök i PV är det betydligt fler som bara har grundskoleutbildning i de äldsta åldersgrupperna (Tabell 12).

Modersmål och utbildningsnivå

Bland de som har svenska som modersmål, görs 61 % av läkarbesöken inom PV av kvinnor. Motsvarande siffra för de som inte har svenska som modersmål är 63 % kvinnor och 37 % män (Tabell 13). Av de patienter som gjort läkarbesök inom VGR kan vi också nämna att 25 % av männen har högskoleutbildning, medan 30 % av kvinnorna har det (Tabell 14). Sist kan vi nämna att de som inte har svenska som modersmål också har något högre utbildning än de som har svenska som modersmål (Tabell 15).

Så här ser det ut om vi gör parvisa jämförelser. Som nämnts tidigare är det lätt att fel slutsatser dras, eftersom de olika bakgrundsvariablerna påverkar varandra. Således skulle t ex kvinnors högre utbildning kunna hänga samman med att kvinnor är yngre när de söker vård och att yngre har längre utbildning än äldre o s v. Vi måste vara försiktiga med att dra slutsatser även om vi har data som stöd, eftersom det kan finnas andra orsaker till att vi finner ett samband.

6. VAL AV METOD

I denna undersökning har jag valt två metoder: logistisk lasso (Least Absolute Shrinkage and Selection Operator) regression och binär logistisk regression. I vårt fall har vi för avsikt att kunna gå vidare senare och använda parametrarna för prediktering (och då är det lämpligt att finna en metod eller modell som krymper koefficienterna, minskar värdet).

När man vill prediktera ger modeller med färre och lägre värden alltid bättre resultat än OLS.

T ex ”Sparse multivariate regression with covariance estimation” utgiven av ”Department of Statistics University of Michigan” nämner tre olika ”sparse” modeller; Multivariate regression with covariance estimation (MRCE), lasso regression och Ridge regression. Där uppges att MRCE tom ger ännu effektivare prediktering än lasso och Ridge men bara i de fall då vi har mycket hög korrelation mellan de oberoende variablerna. Skillnaden mellan dessa tre metoder, vilka alla krymper koefficienterna, är inte så stor som skillnaden mot OLS ¹⁾. Nu vill vi i framtiden kunna använda koefficienterna för att prediktera och därför är en modell som krymper parametrarna aktuell. I de fall vi inte har för avsikt att bestraffa modellen är den optimala lösningen för β alltid $\hat{\beta}^{OLS}$.

Nu har vi i denna undersökning inte problem med alltför många ”covariates” eller prediktorer, men i de fall man har det är det särskilt lämpligt att välja bort Ordinary logistic regression och istället välja lasso ⁵⁾. Lasso är inte mindre lämplig när vi har få prediktorer.

Lasso krymper parametrarna mot 0, i vissa fall t o m till 0, vilket exkluderar dem från modellen. På så vis finner vi de parametrar som har störst inverkan på utfallet av y. Lasso krymper den totala summan av det absoluta värdet av modellens parametrar. Det finns en övre begränsning, t.

$$\sum_{k=1}^p |\beta_k| \leq t$$

$$t > 0$$

Med OLR strävar vi efter att maximera log-likelihood. Detsamma gäller för LLR, men med tillägget av *Lagrangian multiplier*, dvs vi försöker maximera det begränsade eller bestraffade log-likelihood. Vi vill finna ett optimalt lambda värde, dvs det som maximerar log-likelihood med LLR λ .

Formeln för OLR ser ut på följande sätt: λ

$$l(\beta) = \ln(L(\beta|y_1, \dots, y_n))$$

$$= \sum_{i=1}^n \left[\left(\frac{1}{1 + \exp(-\beta^T x'_i)} \right) + (1 - y_i) \ln \left(\frac{\exp(-\beta^T x'_i)}{1 + \exp(-\beta^T x'_i)} \right) \right]$$

$$= - \sum_{i=1}^n [(1 - y_i)\beta^T x'_i + \ln(1 + \exp(-\beta^T x'_i))]$$

Formeln för LLR ser ut på följande sätt:

$$l(\beta)_c = - \sum_{i=1}^n [(1 - y_i)\beta^T x'_i + \ln(1 + \exp(-\beta^T x'_i))] - \lambda \sum_{k=1}^p |\beta_k|$$

När vi funnit ett optimalt Lambda, λ , värde, efter en cross validering, kan det variera beroende på hur många undergrupper (t ex 4- eller 5-fold) som används. Cross validering går ut på att dela upp datamängden i ett antal delmängder. En delmängd skapar modell som sedan testas mot de övriga delmängderna. Det är inte lämpligt med t ex så mycket som 10 delmängder, 10-fold, om man har liten datamängd. Man kan även stratifiera delmängderna. Det går också att upprepa uppdelningarna av delmängder. Cross validering görs effektivt med hjälp av datorer.

$$K = n-1$$

Detta går inte alltid att följa, t ex om n är ”för stort”.

Mjukvaran ”R” som vi har använt oss av för att genomföra lasso regression, räknar ut med automatik, via cross validering, det optimala värdet på t, d v s krympningens storlek beror på t.

Målsättningen är att finna de parametrar som har störst betydelse för utfallet av y. Anledningen är att vi vill få en förenklad modell som tar med de koefficienter som påverkar y mest. Lasso ger tyvärr bias i koefficienter och koefficienternas standard avvikelser. Detta gör att det blir svårt att utvärdera om koefficienterna är signifikanta eller inte. Genom en särskild bootstrap funktion i ”R” försöker man avhjälpa problemet. Bootstrap går ut på att repetera, att göra många urval och jämföra dessa med varandra. Om olika urval stämmer väl överens, kan man göra en säkrare slutsats för en population.

Det finns även andra metoder, för att kalibrera en modell, såsom ”stepwise regression”.

Målet med denna metod är att finna en så reducerad modell som möjligt men som ändå ger en så god prediktering som möjligt. Oftast börjar man med en ”tom” modell, som bara innehåller y och

interceptet. Sedan tas koefficienterna med stegvis, eller reduceras till dess att vi funnit den bästa modellen. **3**

Vilka parametrar vi skall inkludera bestäms av att finna det lägsta BIC värdet, d v s Bayesian information criterion. Vi väljer den modell som ger det lägsta BIC värdet. Ju lägre BIC värde, desto närmare är modellen verkligheten.

$$\text{BIC} = n * \ln \frac{\text{RSS}}{n} + a * \ln * n$$

Det är denna regel som styr vilka koefficienter vi skall ta med.

I vår undersökning kan vi hoppa över detta steg, då vi istället använt oss av eventuell reducering av parametrar via lasso regression.

6. VAL AV METOD (forts)

I undersökningen har vi inga kontinuerliga svarsvärden på Y, utan ordinalskala respons. Därför använder vi oss av logistisk regression. Metoden gör det möjligt att analysera flertalet oberoende variabler. Vi har också ordinalskala på de oberoende variablerna. De går då att rangordna dem men vi vet inget om avstånden mellan dem.

Vi kan antingen välja att ha kvar ordinal skala och välja ordinal logistisk regression eller så kan vi transformera till binära oberoende variabler. I det senare fallet måste vi först undersöka avstånden mellan de oberoende variablernas svarsalternativ och transformera till binär skala, vilket blir en bedömning. En fördel med binär skala kan t ex vara att vi kan göra vissa exekveringar SAS. Liksom logistisk lasso, som vi kan genomföra i "R" kräver binär skala på variablerna. När vi använder oss av en metod som binär logistisk regression innebär det att vi inte predikterar värdet på y, utan P (y=1).

Nedan visas hur data har kodats om till binär skala, se sid 7, de olika frågorna med svarsalternativen

F1	F1	F2	F2	F3	F3
Ordinal	Binär	Ordinal	Binär	Ordinal	Binär
1	1	1	1	1	1
2	1	2	1	2	1
3	1	3	0	3	1
4	0	4	0	4	0
5	0			5	0

F4	F4	F5	F5	F6	F6
Ordinal	Binär	Ordinal	Binär	Ordinal	Binär
1	1	1	1	1	0
2	0	2	0	2	0
				3	1

Användning av olika modeller

Ordinal logistisk regression

Med ordinal logistisk regression transformeras inte några värden, utan man behåller den ursprungliga skalan. Av utrymmesskäl hoppar vi över Ordinal logistisk regression och nöjer oss med Binär logistisk Regression och LASSO logistisk regression. För den intresserade läsaren ges formlerna för Ordinal logistisk regression nedan.

Kumulativa sannolikheter $P(Y \leq i) = p_1 + \dots + p_i$

Kumulativa odds $(Y \leq i) = \frac{P(Y \leq i)}{1 - P(Y \leq i)} = \frac{p_1 + \dots + p_i}{p_{i+1} + \dots + p_k}$

Kumulativ logit $(Y \leq i) = \ln \left(\frac{P(Y \leq i)}{1 - P(Y \leq i)} \right)$

Vi kan beräkna de kumulativa predikterade värdena

$$P(Y \leq i) = 1 / (1 + e^{-(\alpha_j + \beta x)})$$

I denna undersökning, har svarsvariablerna ordinal skala, d v s de går att rangordna. Då kan det vara att föredra de kumulativa sannolikheterna, såsom

$$\gamma_1 = \pi_1, \quad \gamma_2 = \pi_1 + \pi_2 \quad \dots, \quad \gamma_k \equiv 1$$

Binär logistisk regression (se Bilaga 2)

När vi mäter sambandet mellan X_j s med y vill vi kunna mäta y värdet $-\infty < y < \infty$. Därför väljer vi att mäta ett eventuellt linjärt samband mellan logit och x .

$$\text{Logit} = \ln \frac{p}{1-p} = \alpha + \beta x$$

Eller likvärdigt uttryckt

$$p = \frac{\exp^{\alpha + \beta x}}{1 + \exp^{\alpha + \beta x}}$$

I **Bilaga 2** redovisas tabeller över vad vi kommit fram till med användande av binär logistisk regression. Vi kan utläsa att modellerna är signifikanta och endast de signifikanta koefficienterna är redovisade. Om vi vill prediktera kan vi använda oss av ovanstående formel för att beräkna sannolikheten för respektive y variabel.

Logistisk LASSO Regression (LLR)(se Bilaga 3)

LLR är en metod som lämpar sig bäst att använda när vi har ett stort antal bakgrundsvariabler eller multikollinearitet. Metoden är också lämplig vid prediktering eftersom den ger ett mindre mean square error, vilket resulterar i en bättre prediktering. I denna undersökning har vi multikollinearitet och vi är intresserade av att senare kunna använda koefficienterna för prediktering. Lasso krymper parametrarna, vissa blir t o m 0, d v s de ingår inte längre i modellen. Lasso fokuserar på de viktigaste koefficienterna för utfallet av y . Lasso sätter en begränsning av den totala summan av koefficienterna, deras absoluta värde skall understiga t .

$$\sum_{k=1}^p |\beta_k| \leq t$$

$$t > 0$$

För att genomföra logistisk lasso har jag använt mig av mjukvaran ”R” 3). Jag har annars genomgående använt SAS systemet för de statistiska beräkningarna i denna uppsats. Det är ännu inte tekniskt möjligt att utföra logistisk lasso med SAS. När vi kör logistisk lasso i ”R” finner systemet ut det optimala värdet på t . Cross validering sker också med automatik (d v s modellen testas genom att data delas upp i undergrupper, en delmängd kan t ex forma modellen som sedan valideras mot övriga delmängder för att finna en optimal modell) 2).

Det finns ett förhållande mellan t och λ , som är $2/\lambda^2$. D v s ökar λ minskar t . Ju högre värde på λ desto fler koefficienter antar värdet 0. Om $\lambda = 0$ är lasso detsamma som OLS. Bestraffningen eller minskningen av de absoluta värdena på koefficienterna ger bias till koefficienterna och till standardavvikelseerna. Prediktering blir som sagt ändå säkrare med lasso än med OLR, p g a att variansen minskar. Ett stort problem är emellertid att det finns bias i skattningarna, vilket gör det svårt att finna ut om koefficienterna är signifikanta eller inte. För att lösa detta finns en funktion i ”R” med bootstrap 2).

Paketet Cv.glmnet i R levererar två olika värden på ”bästa” λ . :”lambda.min” som ger bästa möjliga prediktering och ett mer heuristiskt värde ”lambda.1se”, vilket värde är :”lambda.min” + ett medelfel för:”lambda.min” 4). Detta sista värde ger en förenklad modell. **Se Bilaga 5.** I vår undersökning är lambda värdena genomgående mycket låga men vi har större standardavvikelse. Med dessa låga lambda värden får vi höga värden på t och för alla sju beroende variabler nedan ingen reducering av koefficienter med ”labda.min” men vi får reducering av koefficienterna, förenklade modeller med ”lambda.1st” 4).

Det är intressant att jämföra koefficienterna från binär logistisk regression med de från logistisk lasso regression (lasso) 2). Vi kan se att koefficienterna krympt mot noll när vi använder lasso, i samtliga fall utom ett, y variabeln E2, x variabeln F5, där lasso koefficienten har ett mycket litet men högre värde än motsvarande binär koefficient, 0.5023 mot 0.5002. I alla andra fall kan vi se hur nära parameter värdena ligger varandra, lasso hela tiden något snävare mot noll.

7. REDOVISNING AV RESULTAT / BILAGOR

1; Parvisa jämförelser mellan oberoende variabler

2; Binär logistisk regression

3; LASSO logistisk regression

Bilaga 1**Parvisa jämförelser mellan oberoende variabler**

F1 I allmänhet, hur skulle du vilja säga att din hälsa är? 1/ Utmärkt 2/ Mycket bra 3/ Bra 4/ Någorlunda 5/ Dålig

F2 Förutom det här besöket–hur många gånger har du varit i kontakt med sjukvården de senaste sex månaderna?

1/ Ingen gång 2/ En gång 3/ Två eller tre gånger 4/ Fyra eller fler gånger

TABELL 1	..	F2 1	F2 2	F2 3	F2 4	Totalt	%
F1 1	Frekvens	686	586	572	229	2073	10.43
	Rad %	33.09	28.27	27.59	11.05		
	Kolumn %	24.37	14.53	7.65	4.14		
F1 2	Frekvens	1079	1312	1706	712	4809	24.21
	Rad %	22.44	27.28	35.48	14.81		
	Kolumn %	38.33	32.52	22.80	12.86		
F1 3	Frekvens	792	1452	2822	1617	6683	33.64
	Rad %	11.85	21.73	42.23	24.20		
	Kolumn %	28.13	35.99	37.72	29.21		
F1 4	Frekvens	227	619	2077	2182	5105	25.70
	Rad %	4.45	12.13	40.69	42.74		
	Kolumn %	8.06	15.35	27.76	39.41		
F1 5	Frekvens	31	65	304	796	1196	6.02
	Rad %	2.59	5.43	25.42	66.56		
	Kolumn %	1.10	1.61	4.06	14.38		
	Totalt	2815	4034	7481	5536	19866	
	%	14.17	20.31	37.66	27.87		
						*saknar svar 491	

F1 I allmänhet, hur skulle du vilja säga att din hälsa är? 1/ Utmärkt 2/ Mycket bra 3/ Bra 4/ Någorlunda 5/ Dålig

F3 Vilket år är du född? 1/ 0-15 år 2/ 16-44 år 3/ 45-64 år 4/ 65-74 år 5/ 75 år eller äldre

TABELL 2	..	F3 1	F3 2	F3 3	F3 4	F3 5	Totalt	%
F1 1	Frekvens	602	623	520	211	99	2055	10.44
	Rad %	29.29	30.32	25.30	10.27	4.82		
	Kolumn %	36.40	14.52	8.44	5.20	2.81		
F1 2	Frekvens	632	1416	1492	822	411	4773	24.24
	Rad %	13.24	29.67	31.26	17.22	8.61		
	Kolumn %	38.21	33.01	24.22	20.25	11.65		
F1 3	Frekvens	323	1362	2137	1613	1190	6625	33.64
	Rad %	4.88	20.56	32.26	24.35	17.96		
	Kolumn %	19.53	31.75	34.69	39.74	33.74		
F1 4	Frekvens	86	671	1529	1216	1560	5062	25.71
	Rad %	1.70	13.26	30.21	24.02	30.82		
	Kolumn %	5.20	15.64	24.82	29.96	44.23		
F1 5	Frekvens	11	218	483	197	267	1176	5.97
	Rad %	0.94	18.54	41.07	16.75	22.70		
	Kolumn %	0.67	5.08	7.84	4.85	7.57		
	Totalt	1654	4290	6161	4059	3527	19691	
	%	8.40	21.79	31.29	20.61	17.91		
							*Saknar svar 666	

F1 I allmänhet, hur skulle du vilja säga att din hälsa är? 1/ Utmärkt 2/ Mycket bra 3/ Bra 4/ Någorlunda 5/ Dålig
 F4 Är du man eller kvinna? 1/ Man 2/ Kvinna

TABELL 3	..	F4 1	F4 2	Totalt	%
F1 1	Frekvens	971	1107	2078	10.42
	Rad %	46.73	53.27		
	Kolumn %	12.45	9.11		
F1 2	Frekvens	2021	2797	4818	24.15
	Rad %	41.95	58.05		
	Kolumn %	25.92	23.01		
F1 3	Frekvens	2654	4059	6713	33.65
	Rad %	39.54	60.46		
	Kolumn %	34.04	33.40		
F1 4	Frekvens	1769	3373	5142	25.77
	Rad %	34.40	65.60		
	Kolumn %	22.69	27.75		
F1 5	Frekvens	382	817	1199	6.01
	Rad %	31.86	68.14		
	Kolumn %	4.90	6.72		
	Totalt	7797	12153	19950	
	%	39.08	60.92		
				*saknar svar 407	

F1 I allmänhet, hur skulle du vilja säga att din hälsa är? 1/ Utmärkt 2/ Mycket bra 3/ Bra 4/ Någorlunda 5/ Dålig

F5 Är svenska ditt modersmål? 1/ Ja 2/ Nej

TABELL 4	..	F5 1	F5 2	Totalt	%
F1 1	Frekvens	1879	195	2074	10.43
	Rad %	90.60	9.40		
	Kolumn %	10.87	7.50		
F1 2	Frekvens	4334	478	4812	24.19
	Rad %	90.07	9.93		
	Kolumn %	25.07	18.38		
F1 3	Frekvens	5720	964	6684	33.61
	Rad %	85.58	14.42		
	Kolumn %	33.09	37.06		
F1 4	Frekvens	4427	697	5124	25.76
	Rad %	86.40	13.60		
	Kolumn %	25.61	26.80		
F1 5	Frekvens	928	267	1195	6.01
	Rad %	77.66	22.34		
	Kolumn %	5.37	10.27		
	Totalt	17288	2601	19889	
%		86.92	13.08		
				*Saknar svar 468	

F1 I allmänhet, hur skulle du vilja säga att din hälsa är? 1/ Utmärkt 2/ Mycket bra 3/ Bra 4/ Någorlunda 5/ Dålig
 F6 Vilken är din högsta avslutade utbildning? 1/ Universitet eller högskola 2/ Gymnasium eller likvärdigt
 3/ Grundskola eller likvärdigt

TABELL 5	..	F6 1	F6 2	F6 3	Totalt	%
F1 1	Frekvens	670	660	404	1734	9.11
	Rad %	38.64	38.06	23.30		
	Kolumn %	12.67	9.16	6.18		
F1 2	Frekvens	1692	1791	1039	4522	23.76
	Rad %	37.42	39.61	22.98		
	Kolumn %	31.99	24.85	15.90		
F1 3	Frekvens	1744	2549	2263	6556	34.45
	Rad %	26.60	38.88	34.52		
	Kolumn %	32.97	35.36	34.63		
F1 4	Frekvens	953	1755	2335	5043	26.50
	Rad %	18.90	34.80	46.30		
	Kolumn %	18.02	24.35	35.73		
F1 5	Frekvens	230	453	494	1177	6.18
	Rad %	19.54	38.49	41.97		
	Kolumn %	4.35	6.28	7.56		
	Totalt	5289	7208	6535	19032	
	%	27.79	37.87	34.34		
					*saknar svar 1325	

F2 Förutom det här besöket – hur många gånger har du varit i kontakt med sjukvården de senaste sex månaderna?

1/ Ingen gång 2/ En gång 3/ Två eller tre gånger 4/ Fyra eller fler gånger

F3 Vilket år är du född? 1/ 0-15 år 2/ 16-44 år 3/ 45-64 år 4/ 65-74 år 5/ 75 år eller äldre

TABELL 6	..	F3 1	F3 2	F3 3	F3 4	F3 5	Totalt	%
F2 1	Frekvens	370	653	972	498	299	2792	14.19
	Rad %	13.25	23.39	34.81	17.84	10.71		
	Kolumn %	22.48	15.25	15.75	12.25	8.51		
F2 2	Frekvens	408	823	1263	834	658	3986	20.26
	Rad %	10.24	20.65	31.69	20.92	16.51		
	Kolumn %	24.79	19.22	20.47	20.52	18.74		
F2 3	Frekvens	598	1503	2236	1624	1435	7396	37.59
	Rad %	8.09	20.32	30.23	21.96	19.40		
	Kolumn %	36.33	35.09	36.23	39.96	40.86		
F2 4	Frekvens	270	1304	1700	1108	1120	5502	27.96
	Rad %	4.91	23.70	30.90	20.14	20.36		
	Kolumn %	16.40	30.45	27.55	27.26	31.89		
	Totalt	1646	4283	6171	4064	3512	19676	
	%	8.37	21.77	31.36	20.65	17.85		
							*Saknar svar 681	

F2 Förutom det här besöket – hur många gånger har du varit i kontakt med sjukvården de senaste sex månaderna?

1/ Ingen gång 2/ En gång 3/ Två eller tre gånger 4/ Fyra eller fler gånger

F4 Är du man eller kvinna? 1/ Man 2/ Kvinna

TABELL 7	..	F4 1	F4 2	Totalt	%
F2 1	Frekvens	1299	1524	2823	14.16
	Rad %	46.01	53.99		
	Kolumn %	16.68	12.55		
F2 2	Frekvens	1628	2418	4046	20.30
	Rad %	40.24	59.76		
	Kolumn %	20.91	19.91		
F2 3	Frekvens	2873	4635	7508	37.67
	Rad %	38.27	61.73		
	Kolumn %	36.89	38.16		
F2 4	Frekvens	1987	3569	5556	27.87
	Rad %	35.76	64.24		
	Kolumn %	25.52	29.38		
	Totalt	7787	12146	19933	
	%	39.07	60.93		
				*saknar svar 424	

F2 Förutom det här besöket – hur många gånger har du varit i kontakt med sjukvården de senaste sex månaderna?

1/ Ingen gång 2/ En gång 3/ Två eller tre gånger 4/ Fyra eller fler gånger

F5 Är svenska ditt modersmål? 1/ Ja 2/ Nej

TABELL 8	..	F5 1	F5 2	Totalt	%
F2 1	Frekvens	2552	264	2816	14.17
	Rad %	90.63	9.38		
	Kolumn %	14.77	10.17		
F2 2	Frekvens	3577	454	4031	20.29
	Rad %	88.74	11.26		
	Kolumn %	20.71	17.50		
F2 3	Frekvens	6471	1017	7488	37.68
	Rad %	86.42	13.58		
	Kolumn %	37.46	39.19		
F2 4	Frekvens	4675	860	5535	27.86
	Rad %	84.46	15.54		
	Kolumn %	27.06	33.14		
	Totalt	17275	2595	19870	
	%	86.94	13.06		
				*Saknar svar 487	

F2 Förutom det här besöket – hur många gånger har du varit i kontakt med sjukvården de senaste sex månaderna?

1/ Ingen gång 2/ En gång 3/ Två eller tre gånger 4/ Fyra eller fler gånger

F6 Vilken är din högsta avslutade utbildning? 1/ Universitet eller högskola 2/ Gymnasium eller likvärdigt

3/ Grundskola eller likvärdigt

TABELL 9	..	F6 1	F6 2	F6 3	Totalt	%
F2 1	Frekvens	841	992	799	2632	13.84
	Rad %	31.95	37.69	30.36		
	Kolumn %	15.90	13.75	12.27		
F2 2	Frekvens	1163	1437	1231	3831	20.15
	Rad %	30.36	37.51	32.13		
	Kolumn %	21.98	19.92	18.91		
F2 3	Frekvens	1899	2683	2585	7167	37.69
	Rad %	26.50	37.44	36.07		
	Kolumn %	35.90	37.19	39.71		
F2 4	Frekvens	1387	2103	1895	5385	28.32
	Rad %	25.76	39.05	35.19		
	Kolumn %	26.22	29.15	29.11		
	Totalt	5290	7215	6510	19015	
	%	27.82	37.94	34.24		
					*saknar svar 1342	

F3 Vilket år är du född? 1/ 0-15 år 2/ 16-44 år 3/ 45-64 år 4/ 65-74 år 5/ 75 år eller äldre
 F4 Är du man eller kvinna? 1/ Man 2/ Kvinna

TABELL 10	..	F4 1	F4 2	Totalt	%
F3 1	Frekvens	828	820	1648	8.33
	Rad %	50.24	49.76		
	Kolumn %	10.74	6.80		
F3 2	Frekvens	1271	3022	4293	21.71
	Rad %	29.61	70.39		
	Kolumn %	16.48	25.05		
F3 3	Frekvens	2328	3862	6190	31.30
	Rad %	37.61	62.39		
	Kolumn %	30.19	32.01		
F3 4	Frekvens	1851	2227	4078	20.62
	Rad %	45.39	54.61		
	Kolumn %	24.00	18.46		
F3 5	Frekvens	1434	2135	3569	18.05
	Rad %	40.18	59.82		
	Kolumn %	18.59	17.69		
	Totalt	7712	12066	19778	
	%	38.99	61.01		
				*Saknar svar 579	

F3 Vilket år är du född? 1/ 0-15 år 2/ 16-44 år 3/ 45-64 år 4/ 65-74 år 5/ 75 år eller äldre

F5 Är svenska ditt modersmål? 1/ Ja 2/ Nej

TABELL 11	..	F5 1	F5 2	Totalt	%
F3 1	Frekvens	1372	280	1652	8.38
	Rad %	83.05	16.95		
	Kolumn %	8.00	10.92		
F3 2	Frekvens	3622	658	4280	21.71
	Rad %	84.63	15.37		
	Kolumn %	21.12	25.66		
F3 3	Frekvens	5170	988	6158	31.24
	Rad %	83.96	16.04		
	Kolumn %	30.15	38.53		
F3 4	Frekvens	3689	374	4063	20.61
	Rad %	90.79	9.21		
	Kolumn %	21.51	14.59		
F3 5	Frekvens	3297	264	3561	18.06
	Rad %	92.59	7.41		
	Kolumn %	19.22	10.30		
	Totalt	17150	2564	19714	
%		86.99	13.01		
				*saknar svar 643	

F3 Vilket år är du född? 1/ 0-15 år 2/ 16-44 år 3/ 45-64 år 4/ 65-74 år 5/ 75 år eller äldre

F6 Vilken är din högsta avslutade utbildning? 1/ Universitet eller högskola 2/ Gymnasium eller likvärdigt
3/ Grundskola eller likvärdigt

TABELL 12	..	F6 1	F6 2	F6 3	Totalt	%
F3 1	Frekvens	343	299	273	915	4.85
	Rad %	37.49	32.68	29.84		
	Kolumn %	6.52	4.17	4.24		
F3 2	Frekvens	1531	2131	621	4283	22.70
	Rad %	35.75	49.75	14.50		
	Kolumn %	29.12	29.72	9.64		
F3 3	Frekvens	1941	2681	1528	6150	32.60
	Rad %	31.56	43.59	24.85		
	Kolumn %	36.92	37.39	23.73		
F3 4	Frekvens	912	1237	1887	4036	21.39
	Rad %	22.60	30.65	46.75		
	Kolumn %	17.35	17.25	29.31		
F3 5	Frekvens	530	822	2130	3482	18.46
	Rad %	15.22	23.61	61.17		
	Kolumn %	10.08	11.46	33.08		
	Totalt	5257	7170	6439	18866	
	%	27.86	38.00	34.13		
					*Saknar svar 1491	

F4 Är du man eller kvinna? 1/ Man 2/ Kvinna

F5 Är svenska ditt modersmål? 1/ Ja 2/ Nej

TABELL 13	..	F5 1	F5 2	Totalt	%
F4 1	Frekvens	6814	978	7792	38.99
	Rad %	87.45	12.55		
	Kolumn %	39.24	37.34		
F4 2	Frekvens	10550	1641	12191	61.01
	Rad %	86.54	13.46		
	Kolumn %	60.76	62.66		
	Totalt	17364	2619	19983	
	%	86.89	13.11		
				*saknar svar 374	

F4 Är du man eller kvinna? 1/ Man 2/ Kvinna

F6 Vilken är din högsta avslutade utbildning? 1/ Universitet eller högskola 2/ Gymnasium eller likvärdigt
3/ Grundskola eller likvärdigt

TABELL 14	..	F6 1	F6 2	F6 3	Totalt	%
F4 1	Frekvens	1833	2824	2715	7372	38.57
	Rad %	24.86	38.31	36.83		
	Kolumn %	34.58	38.97	41.35		
F4 2	Frekvens	3468	4422	3851	11741	61.43
	Rad %	29.54	37.66	32.80		
	Kolumn %	65.42	61.03	58.65		
	Totalt	5301	7246	6566	19113	
	%	27.74	37.91	34.35		
					*Saknar svar 1244	

F5 Är svenska ditt modersmål? 1/ Ja 2/ Nej

F6 Vilken är din högsta avslutade utbildning? 1/ Universitet eller högskola 2/ Gymnasium eller likvärdigt
3/ Grundskola eller likvärdigt

TABELL 15	..	F6 1	F6 2	F6 3	Totalt	%
F5 1	Frekvens	4565	6183	5787	16535	86.78
	Rad %	27.61	37.39	35.00		
	Kolumn %	86.43	85.63	88.32		
F5 2	Frekvens	717	1038	765	2520	13.22
	Rad %	28.45	41.19	30.36		
	Kolumn %	13.57	14.37	11.68		
	Totalt	5282	7221	6552	19055	
	%	27.72	37.90	34.38		
					*Saknar svar 1302	

BILAGA 2

E1 Hur värderar du som helhet den vård/ behandling du fick?

BINÄR LOGISTISK REGRESSION

Probability modeled is E1=1

Number of Observation Read	20357
Number of Observation Used	18495

Response profile

ordered value	E1	Tot Freq
1	0	6361
2	1	12134

Testing global null hypothesis: BETA=0

Test	Chi Square	df	Pr >ChiSq
Likelihood Ratio	597.7665	5	<.0001
Score	598.1169	5	<.0001
Wald	577.2447	5	<.0001
<i>*modellen är signifikant</i>			

Analysis of Maximum Likelihood Estimates

Parameter	Df	estimate	std error	Pr >ChiSq
intercept	1	0.1473	0.0563	0.0090
F1	1	0.5685	0.0340	<.0001
F3	1	-0.4741	0.0354	<.0001
F4	1	0.1029	0.0327	0.0017
F5	1	0.4780	0.0450	<.0001
F6	1	-0.1000	0.0355	0.0049
<i>*signifikanta parametrar är medtagna</i>				

E2 Kände du dig delaktig i beslut om din vård och behandling, så mycket som du önskade?

BINÄR LOGISTISK REGRESSION

<i>Probability modeled is E2=1</i>	
Number of Observation Read	20357
Number of Observation Used	19176

Response profile

ordered value	E2	Tot Freq
1	0	7173
2	1	12003

Testing global null hypothesis: BETA=0

Test	Chi Square	df	Pr >ChiSq
Likelihood Ratio	682.3645	4	<.0001
Score	677.9795	4	<.0001
Wald	652.3384	4	<.0001
<i>*modellen är signifikant</i>			

Analysis of Maximum Likelihood Estimates

Parameter	df	estimate	std error	Pr >ChiSq
intercept	1	0.0434	0.0500	0.3854
F1	1	0.5306	0.0344	<.0001
F2	1	0.0934	0.0336	0.0055
F3	1	-0.5530	0.0328	<.0001
F5	1	0.5002	0.0442	<.0001
<i>*signifikanta parametrar är medtagna</i>				

E3 Kände du att du blev bemött med respekt och på ett hänsynsfullt sätt?

BINÄR LOGISTISK REGRESSION

<i>Probability modeled is E3=1</i>	
<i>Number of Observation Read</i>	20357
<i>Number of Observation Used</i>	19268

Response profile

<i>ordered value</i>	<i>E3</i>	<i>Tot Freq</i>
1	0	3555
2	1	15713

Testing global null hypothesis: BETA=0

<i>Test</i>	<i>Chi Square</i>	<i>df</i>	<i>Pr >ChiSq</i>
Likelihood Ratio	726.6917	5	<.0001
Score	699.9220	5	<.0001
Wald	666.8626	5	<.0001
<i>*modellen är signifikant</i>			

Analysis of Maximum Likelihood Estimates

<i>Parameter</i>	<i>df</i>	<i>estimate</i>	<i>std error</i>	<i>Pr >ChiSq</i>
intercept	1	1.4423	0.0630	<.0001
F1	1	0.5334	0.0421	<.0001
F2	1	0.1115	0.0427	0.0090
F3	1	-0.9594	0.0446	<.0001
F4	1	0.1521	0.0398	0.0001
F5	1	0.2954	0.0518	<.0001
<i>*signifikanta parametrar är medtagna</i>				

E4 Fick du tillräcklig information om ditt tillstånd?

BINÄR LOGISTISK REGRESSION

<i>Probability modeled is E4=1</i>	
<i>Number of Observation Read</i>	20357
<i>Number of Observation Used</i>	16710

Response profile

ordered value	E4	Tot Freq
1	0	6346
2	1	10364

Testing global null hypothesis: BETA=0

Test	Chi Square	df	Pr >ChiSq
Likelihood Ratio	636.1204	6	<.0001
Score	629.8350	6	<.0001
Wald	607.2239	6	<.0001
<i>*modellen är signifikant</i>			

Analysis of Maximum Likelihood Estimates

Parameter	df	estimate	std error	Pr >ChiSq
intercept	1	0.0624	0.0590	0.2909
F1	1	0.7044	0.0367	<.0001
F2	1	0.0711	0.0362	0.0497
F3	1	-0.4610	0.0368	<.0001
F4	1	0.1467	0.0337	<.0001
F5	1	0.1270	0.0473	0.0073
F6	1	0.1863	0.0371	<.0001
<i>*signifikanta parametrar är medtagna</i>				

E6 Hur bedömer du att samarbetet mellan personalen fungerade?

BINÄR LOGISTISK REGRESSION

Probability modeled is E6=1

Number of Observation Read	20357
Number of Observation Used	14523

Response profile

ordered value	E6	Tot Freq
1	0	5057
2	1	9466

Testing global null hypothesis: BETA=0

Test	Chi Square	df	Pr >ChiSq
Likelihood Ratio	386.1329	5	<.0001
Score	383.3486	5	<.0001
Wald	373.0518	5	<.0001

**modellen är signifikant*

Analysis of Maximum Likelihood Estimates

Parameter	df	estimate	std error	Pr >ChiSq
intercept	1	0.3584	0.0625	<.0001
F1	1	0.4456	0.0382	<.0001
F3	1	-0.5326	0.0399	<.0001
F4	1	0.1144	0.0364	0.0017
F5	1	0.3457	0.0493	<.0001
F6	1	-0.1027	0.0396	0.0096

**signifikanta parametrar är medtagna*

E7 Anser du att ditt aktuella behov av sjukvård blivit tillgodosett vid ditt besök på mottagningen?

BINÄR LOGISTISK REGRESSION

Probability modeled is E7=1

Number of Observation Read	20357
Number of Observation Used	18335

Response profile

ordered value	E7	Tot Freq
1	0	5583
2	1	12752

Testing global null hypothesis: BETA=0

Test	Chi Square	df	Pr >ChiSq
Likelihood Ratio	925.9688	6	<.0001
Score	916.5380	6	<.0001
Wald	868.9900	6	<.0001

**modellen är signifikant*

Analysis of Maximum Likelihood Estimates

Parameter	df	estimate	std error	Pr >ChiSq
intercept	1	0.2444	0.0585	<.0001
F1	1	0.7223	0.0367	<.0001
F2	1	0.1612	0.0371	<.0001
F3	1	-0.6380	0.0376	<.0001
F4	1	0.0872	0.0343	0.0111
F5	1	0.4379	0.0464	<.0001
F6	1	0.1634	0.0377	<.0001

**signifikanta parametrar är medtagna*

E8 Skulle du rekommendera den här mottagningen till andra?

BINÄR LOGISTISK REGRESSION

<i>Probability modeled is E8=1</i>	
<i>Number of Observation Read</i>	20357
<i>Number of Observation Used</i>	18411

Response profile

ordered value	E8	Tot Freq
1	0	5312
2	1	13099

Testing global null hypothesis: BETA=0

Test	Chi Square	df	Pr >ChiSq
Likelihood Ratio	546.0969	5	<.0001
Score	538.5481	5	<.0001
Wald	522.5654	5	<.0001
<i>*modellen är signifikant</i>			

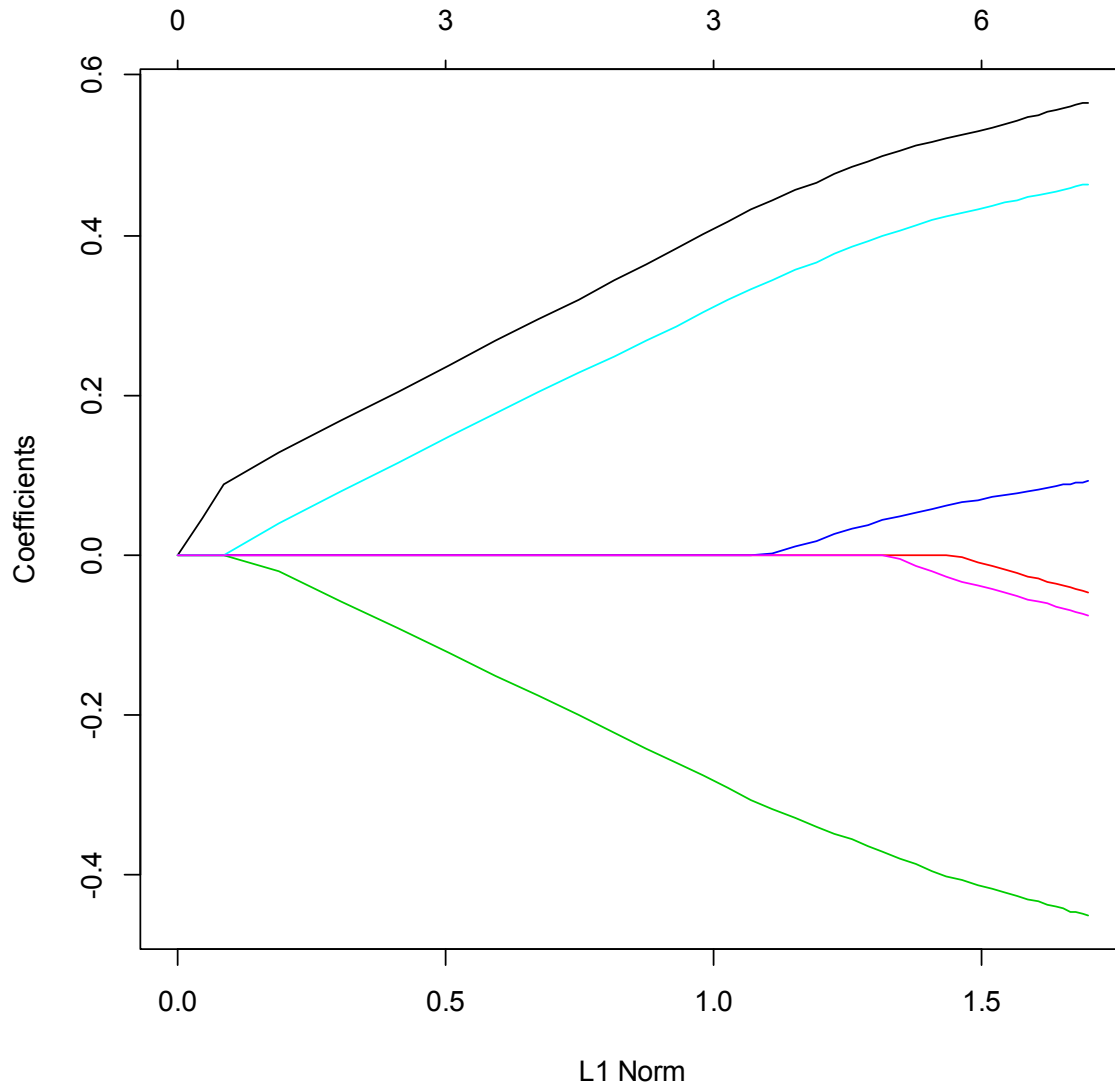
Analysis of Maximum Likelihood Estimates

Parameter	df	estimate	std error	Pr >ChiSq
intercept	1	0.6248	0.0591	<.0001
F1	1	0.5126	0.0356	<.0001
F3	1	-0.5892	0.0377	<.0001
F4	1	0.1388	0.0345	<.0001
F5	1	0.2640	0.0469	<.0001
F6	1	0.1038	0.0378	0.0060
<i>*signifikanta parametrar är medtagna</i>				

Bilaga 3; LASSO logistisk regression

Resultat för E1 (Hur värderar du som helhet den vård/behandling du fick ?)

F1 Svart, F2 Röd, F3 Limegrön, F4 Kornblå, F5 Turkos, F6 Cerise

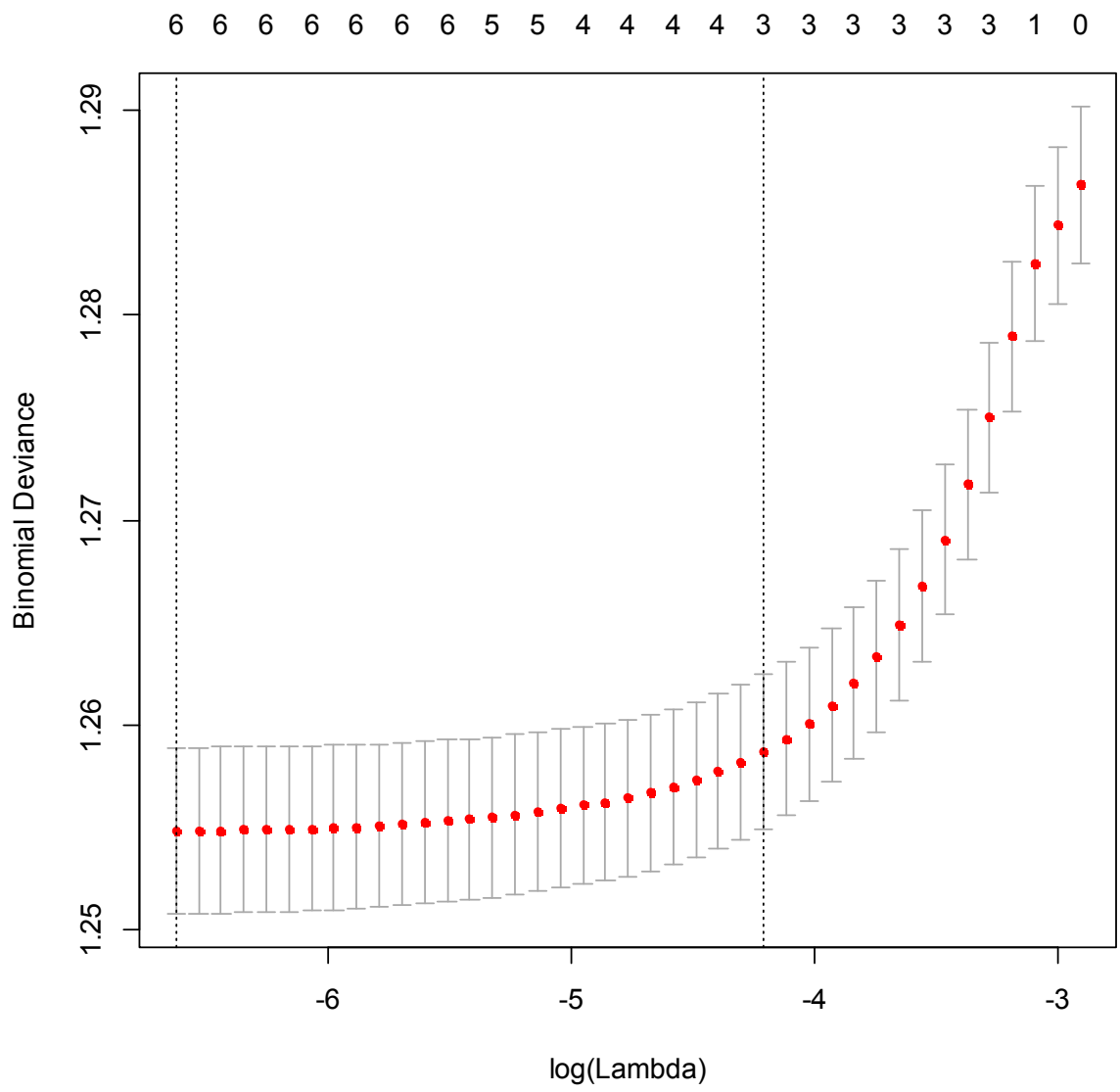


Lambda.min 0.001320948

E1	Koefficient
(Intercept)	0.16057090
F1	0.56633790
F2	-0.04759665
F3	-0.45156157
F4	0.09259794
F5	0.46375376
F6	-0.07499222

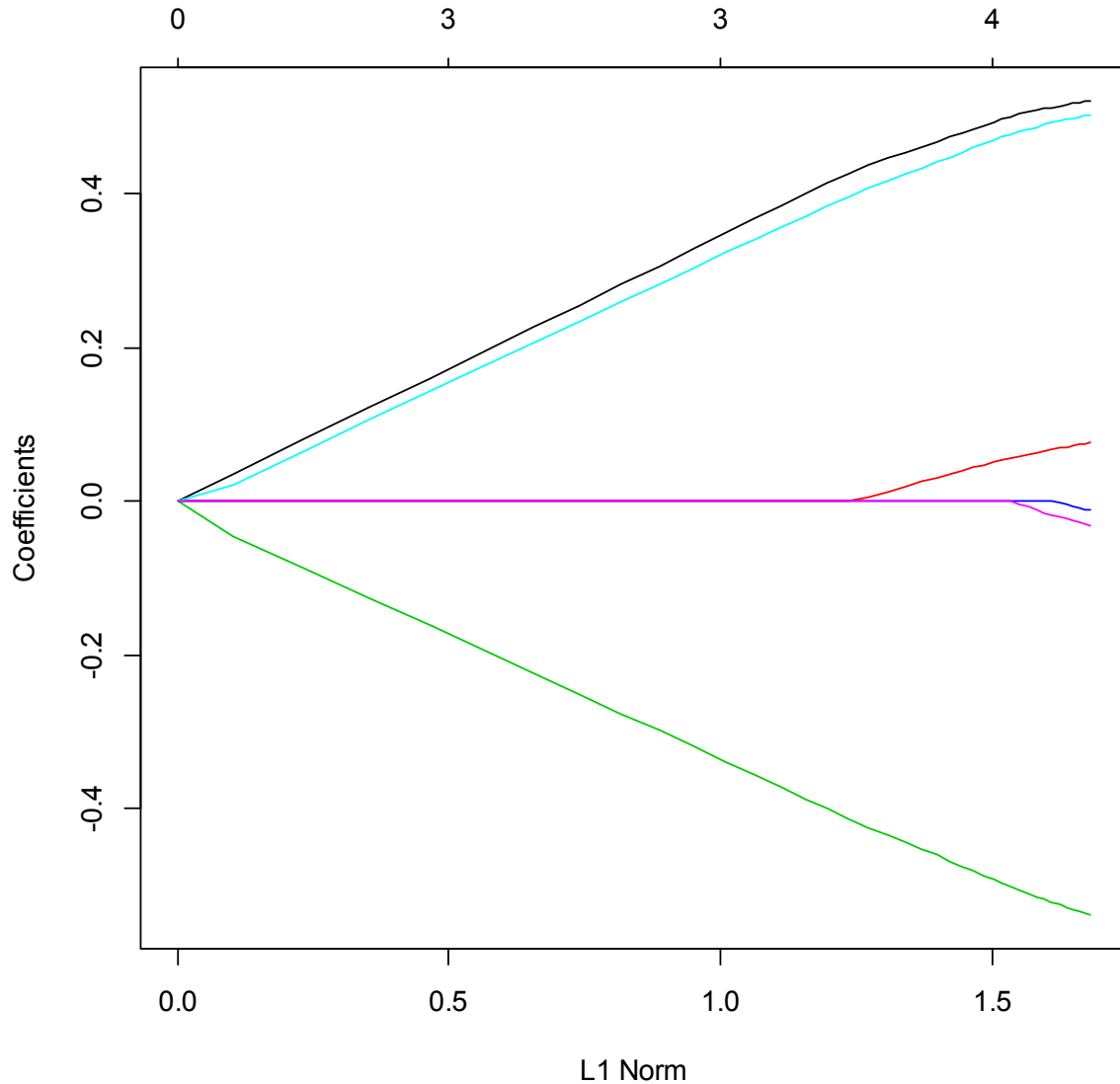
Lambda.1se 0.01483853

E1	Koefficient
(Intercept)	0.2760693
F1	0.4165815
F2	.
F3	-0.2918839
F4	.
F5	0.3183458
F6	.



24)

Resultat för E2 (Kände du dig delaktig i beslut om din vård och behandling, så mycket som du önskade?) F1 Svart, F2 Röd, F3 Limegrön, F4 Kornblå, F5 Turkos, F6 Cerise

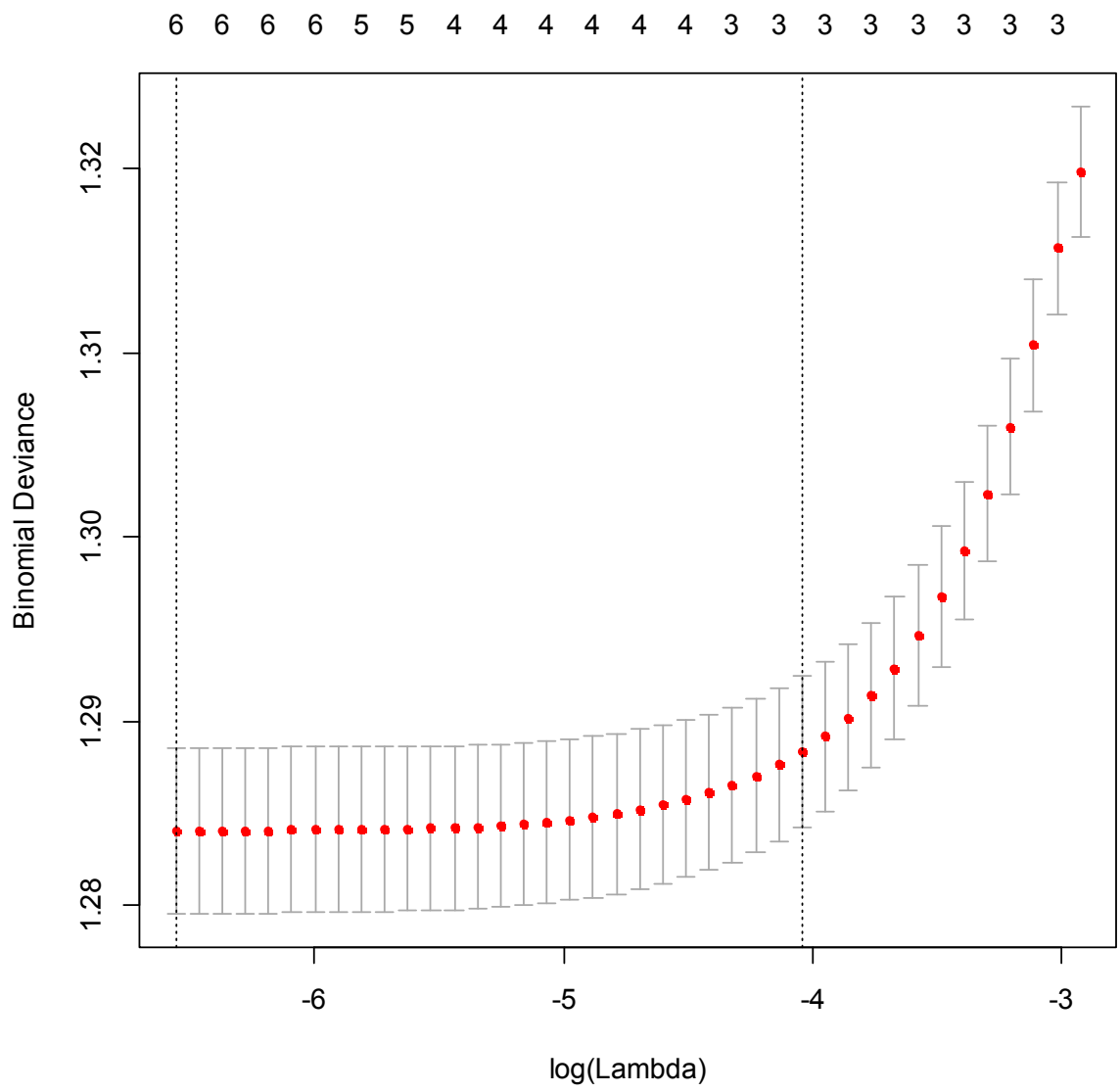


Lambda.min 0.001421456

E2	Koefficient
(Intercept)	0.06801465
F1	0.52113178
F2	0.07581624
F3	-0.53813030
F4	-0.01158406
F5	0.50236429
F6	-0.03140019

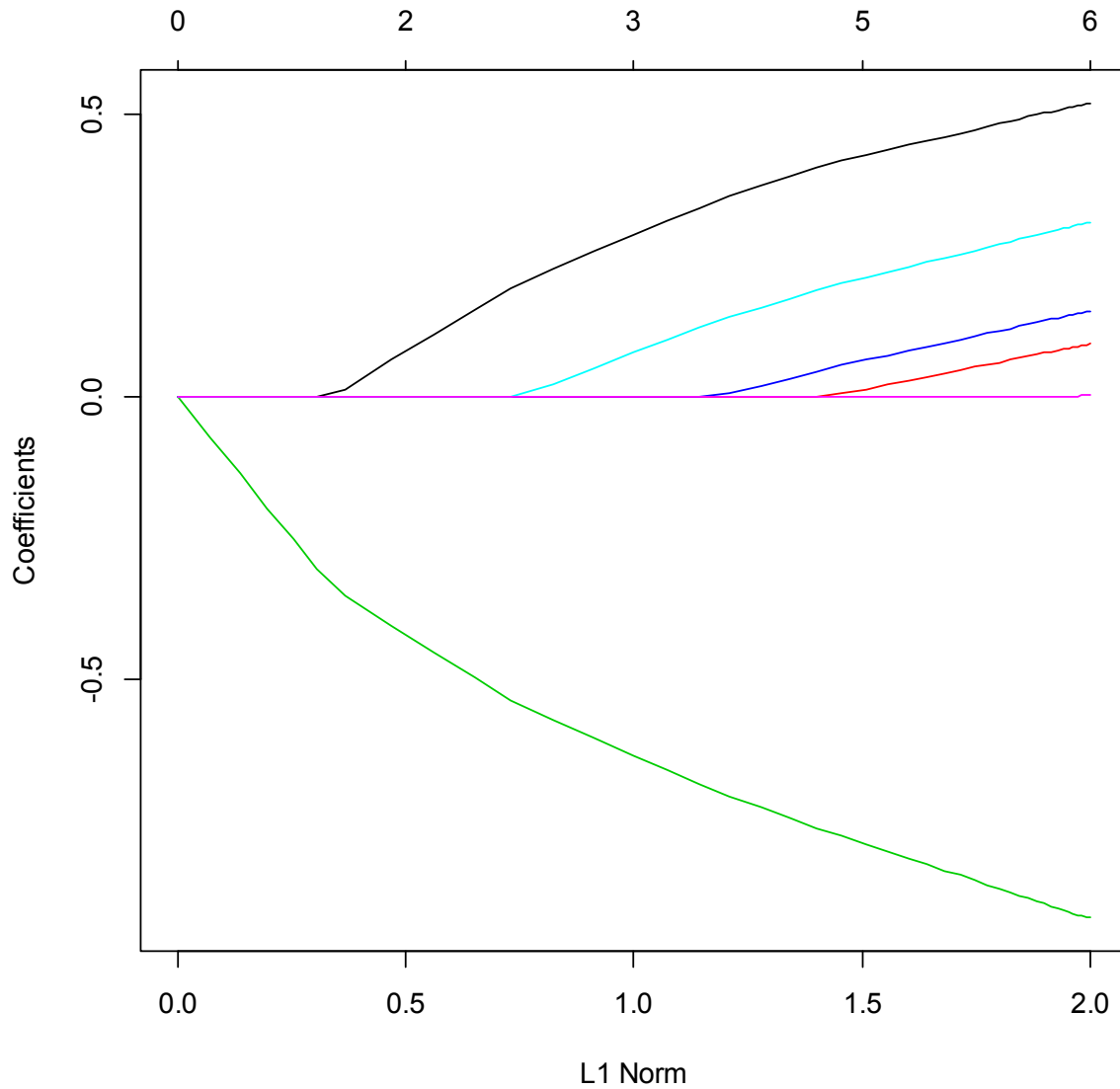
Lambda.1se 0.01752437

E2	Koefficient
(Intercept)	0.2039903
F1	0.3668881
F2	.
F3	-0.3558198
F4	.
F5	0.3403301
F6	.



2) 4)

Resultat för E3 (Kände du att du blev bemött med respekt och på ett hänsynsfullt sätt?)
 F1 Svart, F2 Röd, F3 Limegrön, F4 Kornblå, F5 Turkos, F6 Cerise

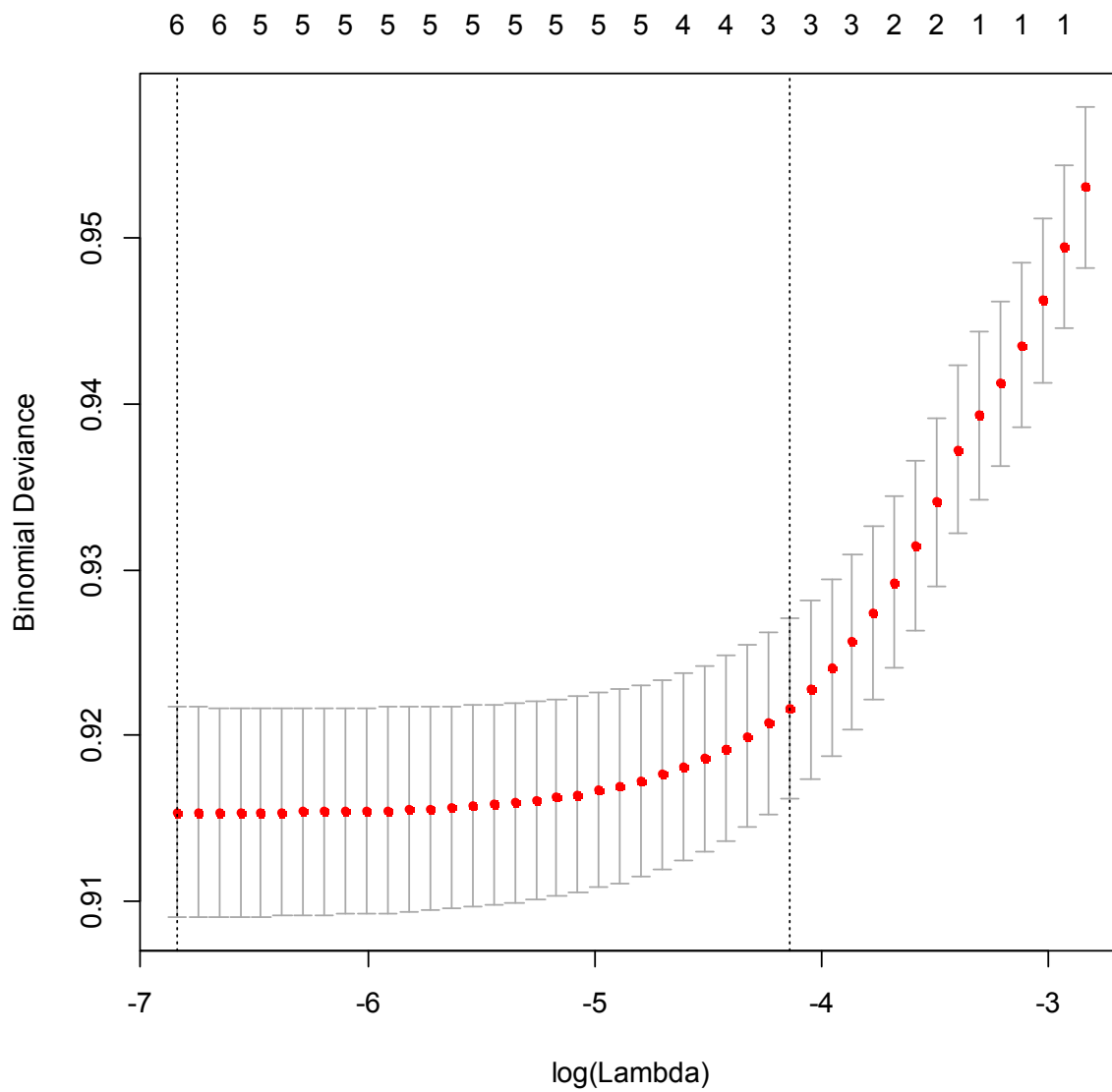


Lambda.min 0.001069544

E3	Koefficient
(Intercept)	1.421585627
F1	0.520180457
F2	0.092851238
F3	-0.922506497
F4	0.151187380
F5	0.308427600
F6	0.00435333

Lambda.1se 0.01588238

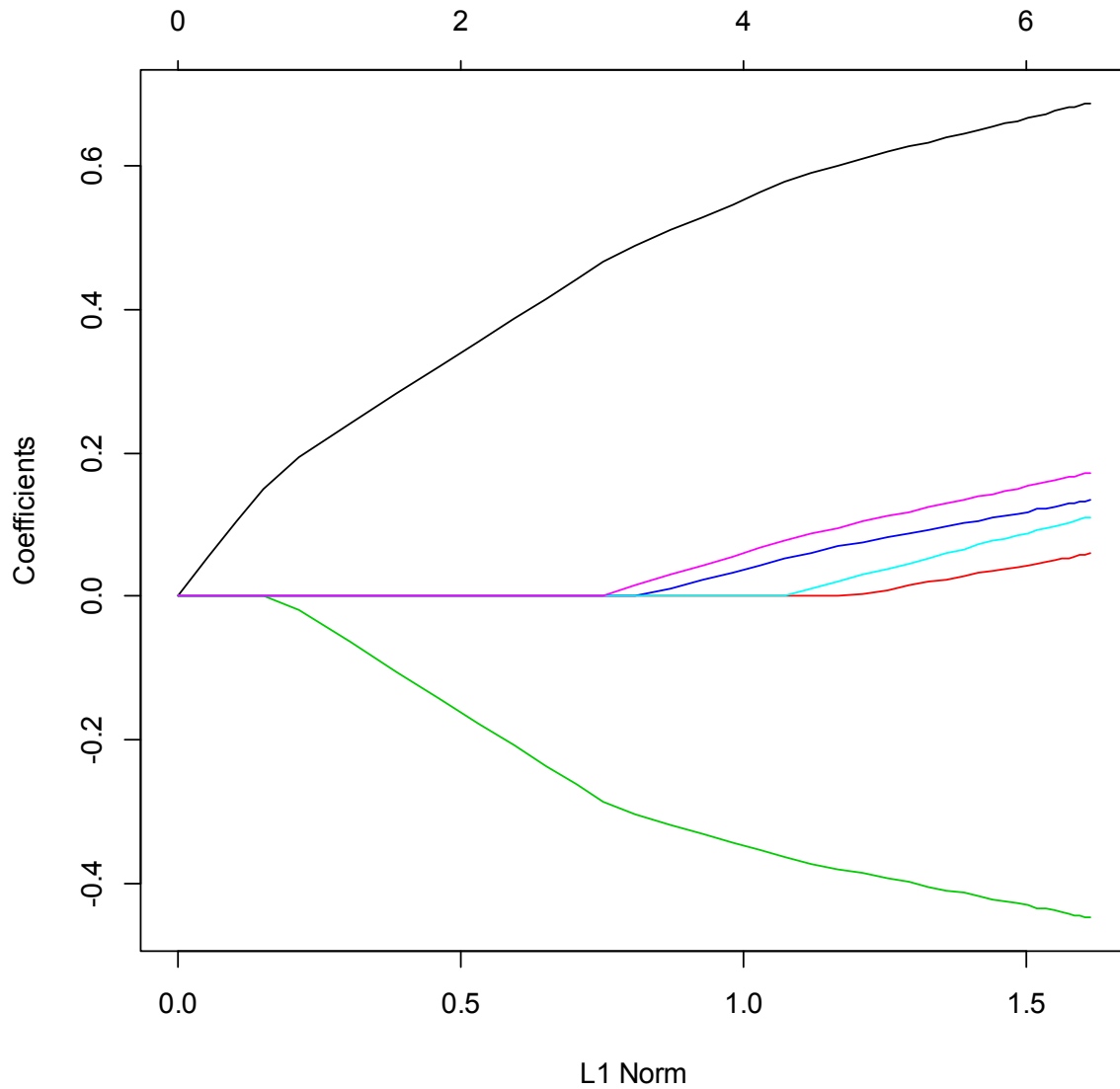
E3	Koefficient
(Intercept)	1.6303201
F1	0.3112051
F2	.
F3	-0.6613846
F4	.
F5	0.1006052
F6	.



24)

Resultat för E4 (Fick du tillräcklig information om ditt tillstånd?)

F1 Svart, F2 Röd, F3 Limegrön, F4 Kornblå, F5 Turkos, F6 Cerise

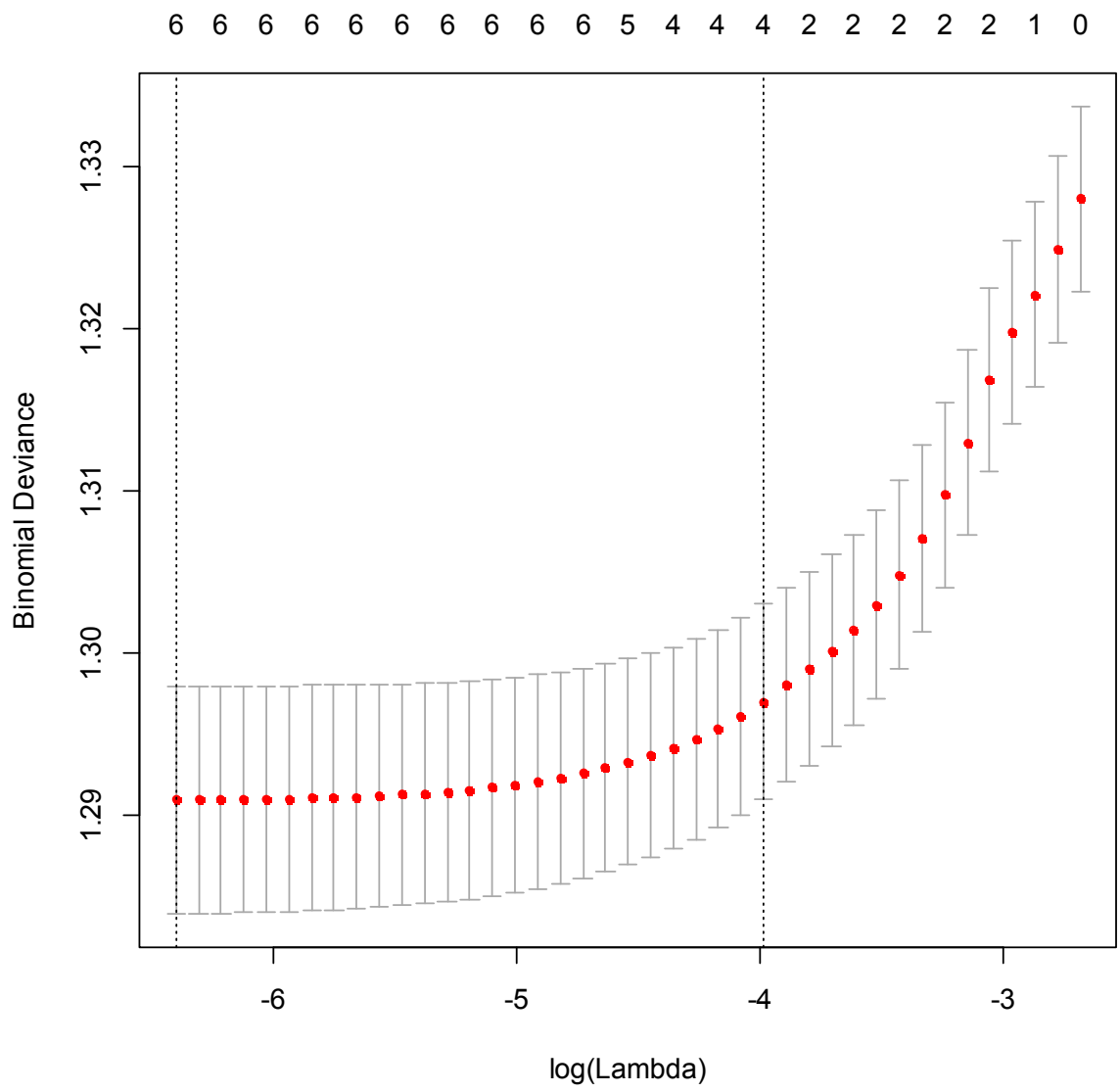


Lambda.min 0.001655555

E4	Koefficient
(Intercept)	0.09209304
F1	0.68831814
F2	0.05932311
F3	-0.44824521
F4	0.13448097
F5	0.11037285
F6	0.17205138

Lambda.1se 0.01859725

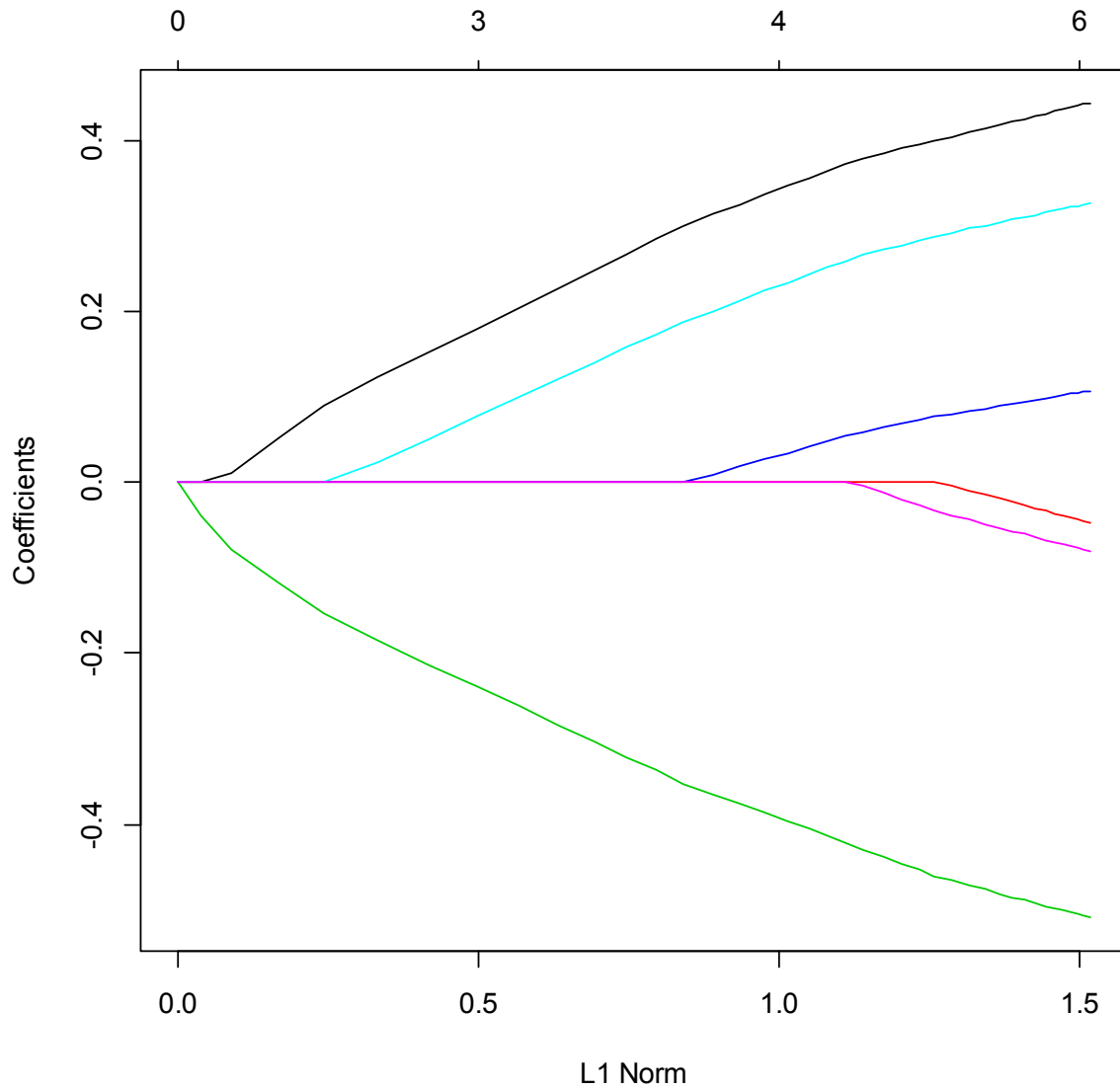
E4	Koefficient
(Intercept)	0.33415742
F1	0.51052169
F2	.
F3	-0.31809853
F4	0.01145751
F5	.
F6	0.03048153



24)

Resultat för E6 Hur bedömer du att samarbetet mellan personalen fungerade?

F1 Svart, F2 Röd, F3 Limegrön, F4 Kornblå, F5 Turkos, F6 Cerise

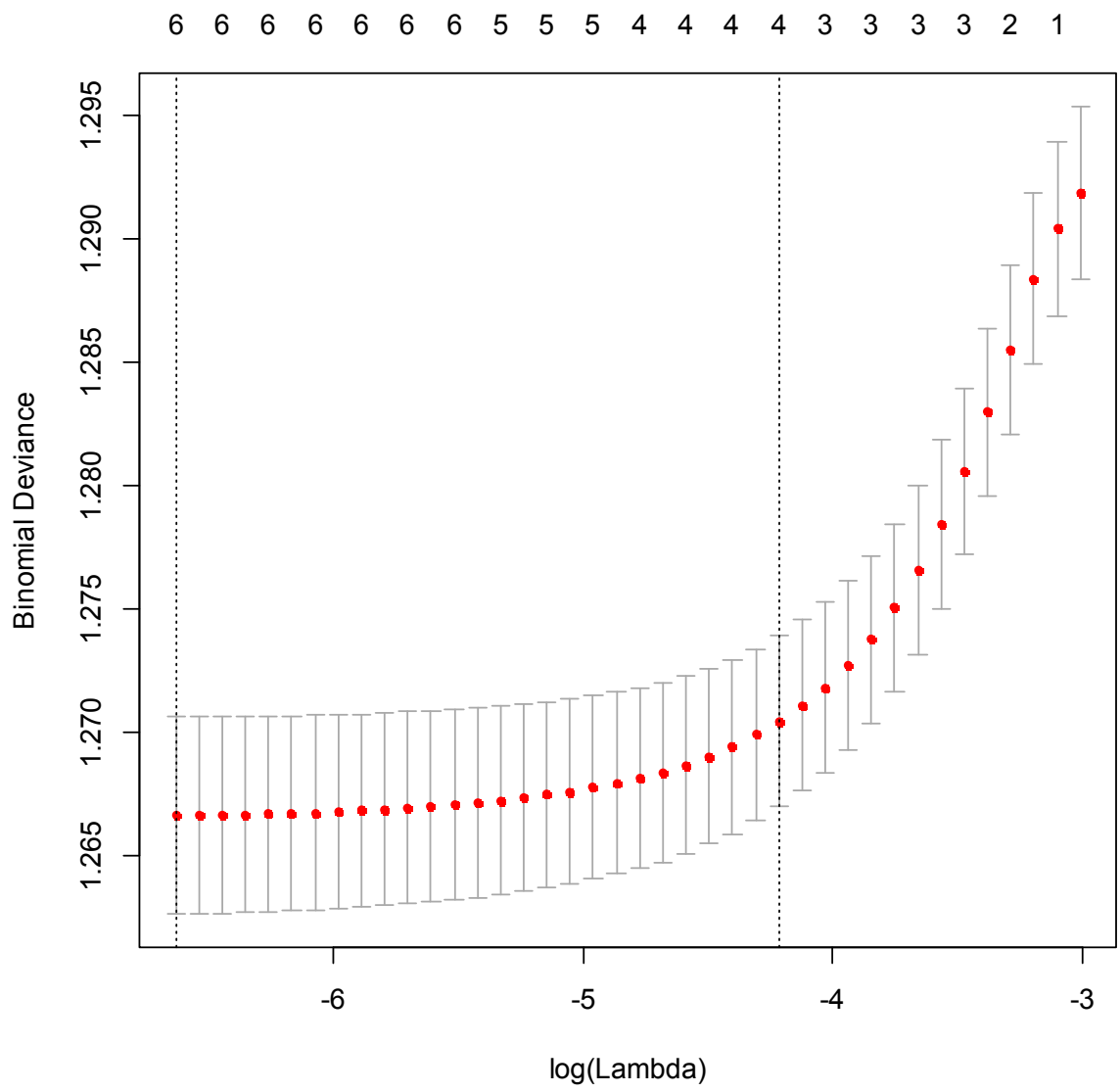


Lambda.min 0.001313721

E6	Koefficient
(Intercept)	0.37202195
F1	0.44537135
F2	-0.04685758
F3	-0.50979716
F4	0.10771153
F5	0.32740560
F6	-0.08017078

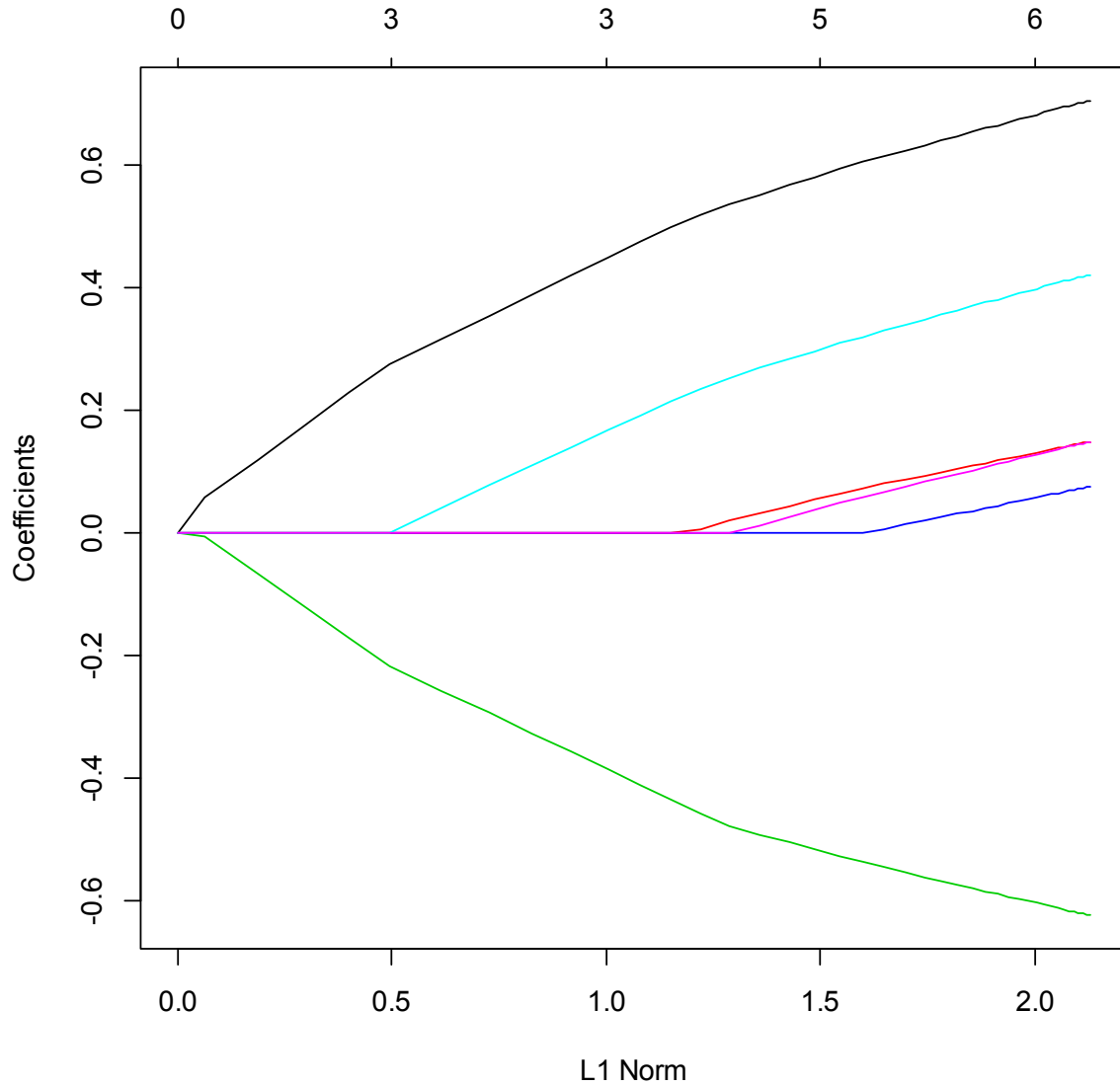
Lambda.1se 0.01475734

E6	Koefficient
(Intercept)	0.4843871986
F1	0.3009555011
F2	.
F3	-0.3522261210
F4	0.0001704529
F5	0.1886284006
F6	.



24)

Resultat för E7 (Anser du att ditt aktuella behov av sjukvård blivit tillgodosett vid ditt besök på mottagningen?) F1 Svart, F2 Röd, F3 Limegrön, F4 Kornblå, F5 Turkos, F6 Cerise

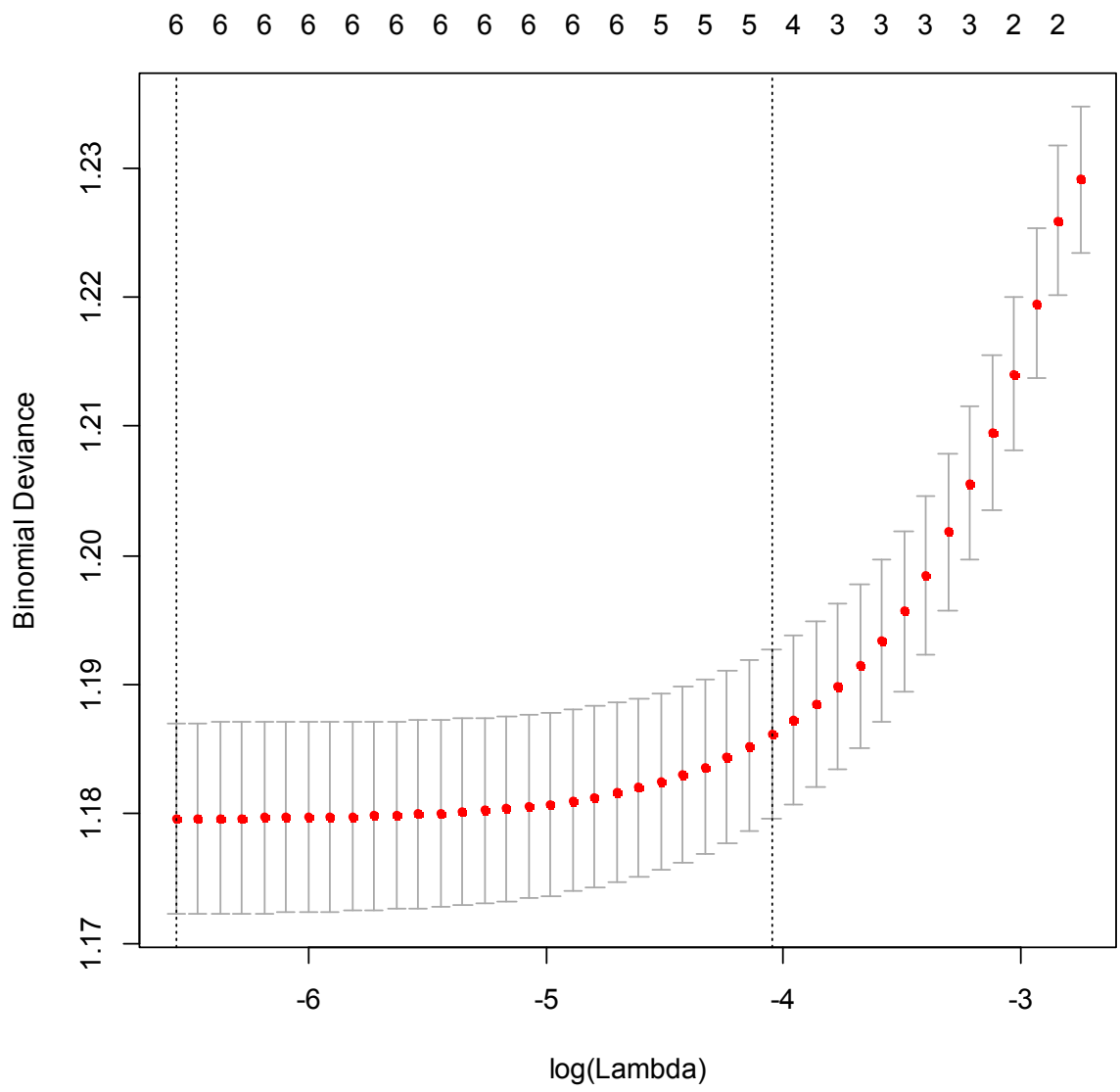


Lambda.min 0.001413343

E7	Koefficient
(Intercept)	0.27071347
F1	0.70694359
F2	0.14950425
F3	-0.62465390
F4	0.07550894
F5	0.42253336
F6	0.14950234

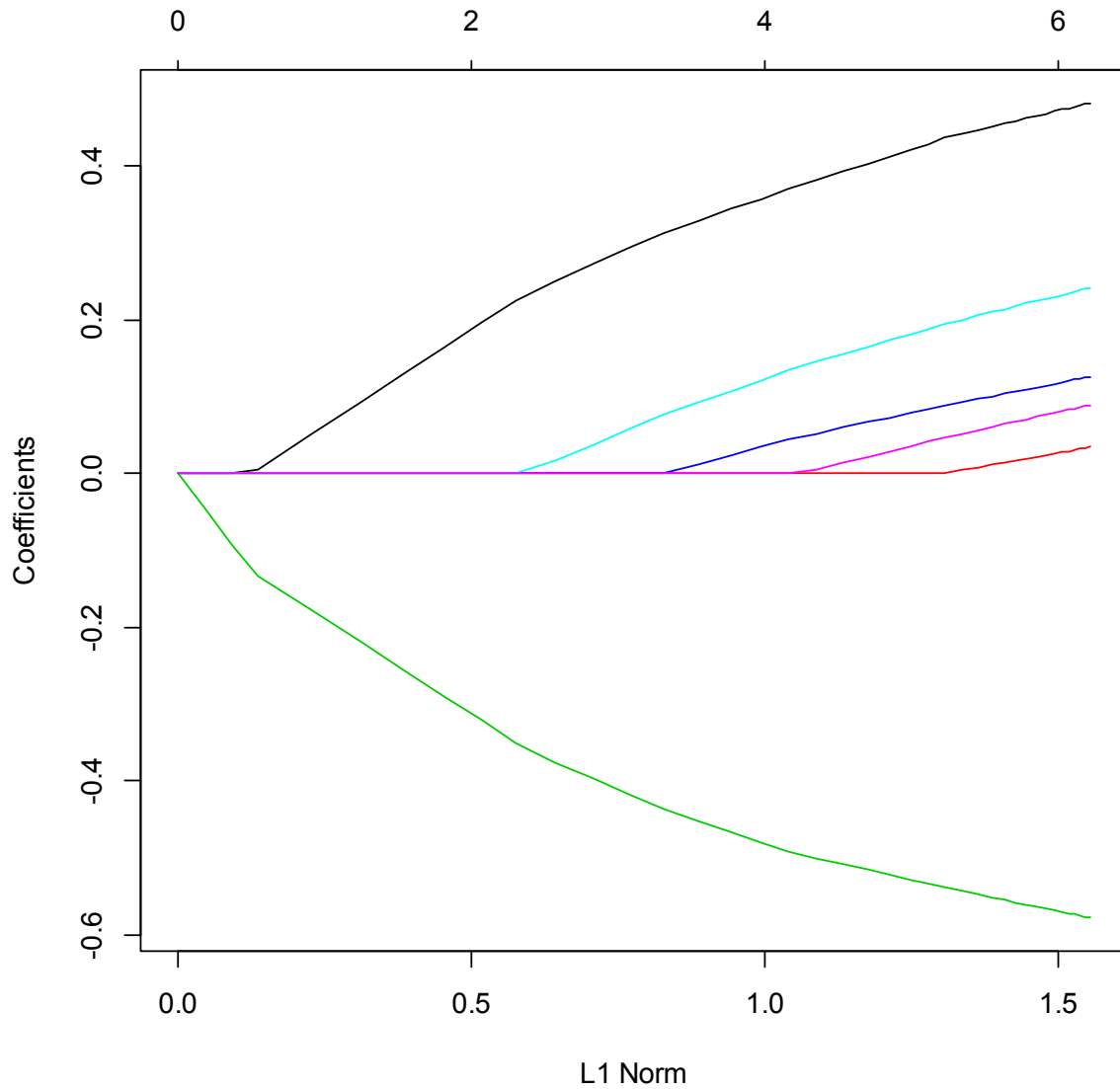
Lambda.1se 0.01742435

E7	Koefficient
(Intercept)	0.54625755
F1	0.53641761
F2	0.02086832
F3	-0.47738185
F4	.
F5	0.25282012
F6	.



2) 4)

Resultat för E8 Skulle du rekommendera den här mottagningen till andra?
 F1 Svart, F2 Röd, F3 Limegrön, F4 Kornblå, F5 Turkos, F6 Cerise

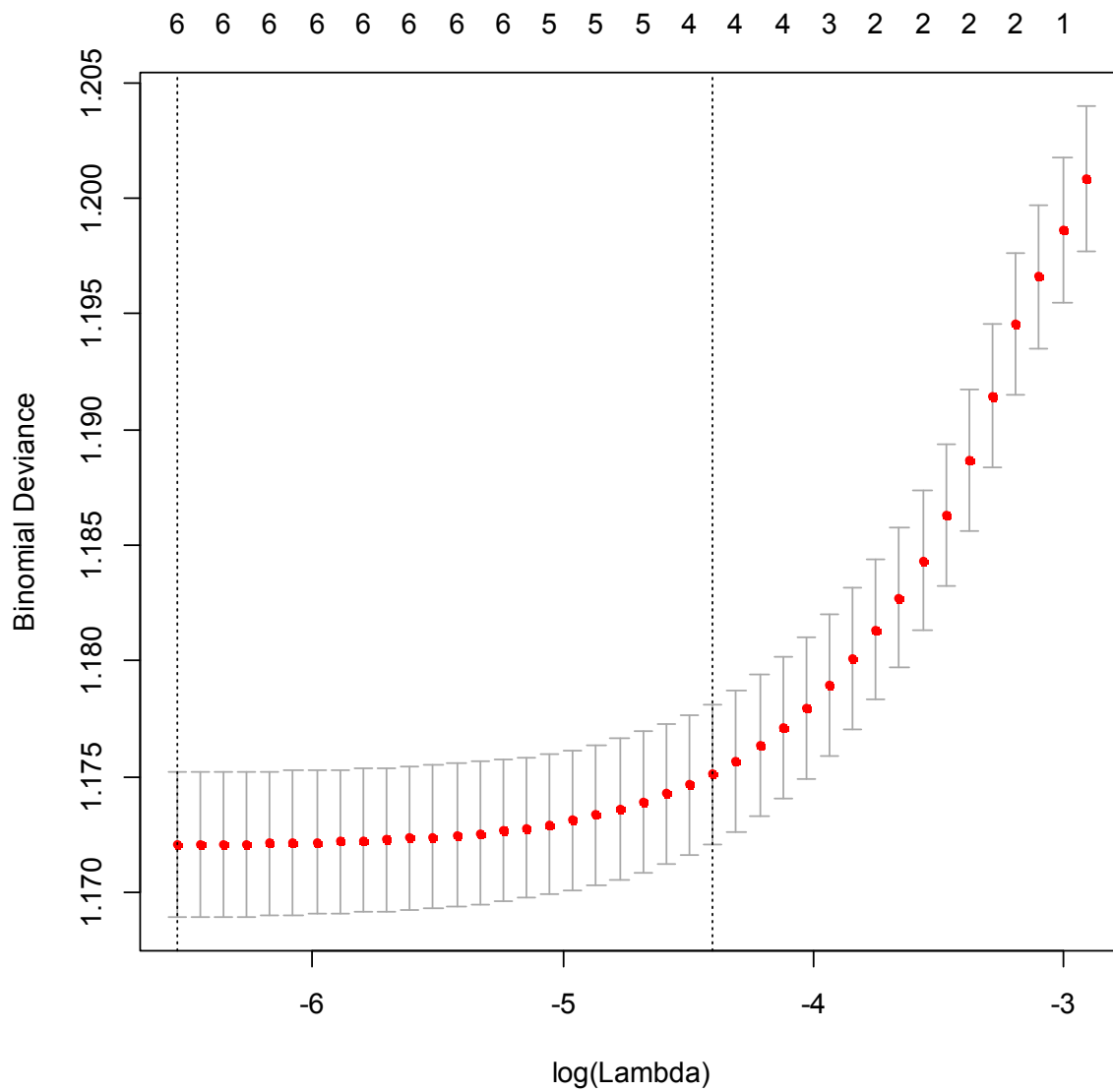


Lambda.min 0.001442923

E8	Koefficient
(Intercept)	0.65522207
F1	0.48275569
F2	0.03447708
F3	-0.57811243
F4	0.12610081
F5	0.24225376
F6	0.08918959

Lambda.1se 0.01226129

E8	Koefficient
(Intercept)	0.84927986
F1	0.35796050
F2	.
F3	-0.47995339
F4	0.03460757
F5	0.12163885
F6	.



24

8. SLUTSATSER

Genom resultaten i uppsatsen kan vi konstatera att *olika bakgrunder hos patienter påverkar patientupplevd vårdkvalité*. Därför kan det vara missvisande att jämföra hur nöjda patienter är på olika VC om vi inte tar hänsyn till olika bakgrunder hos patienterna. Det har framkommit att patienter som upplever att de har dålig hälsa också är mest missnöjda med vården. Sedan kommer ålder, där yngre inte är lika nöjda som äldre patienter. En tredje bakgrundsvariabel som påverkar nöjdheten är om man har svenska som modersmål eller inte. De som har svenska som modersmål är mer nöjda än de som inte har svenska som modersmål. Enligt uppgift från Indikator, görs även i England och USA korrigering för egenupplevd hälsa och ålder. I VGR uppvisas samma mönster som i USA och i England, med tanke på de två bakgrundsvariablerna, ålder och egenupplevd hälsa.

Jag har använt två olika metoder, binär logistisk regression samt LASSO logistisk regression. Jag räknade fram koefficienterna för binär logistisk regression för att se vilka bakgrunder hos patienterna som starkast påverkar patientupplevd vårdkvalité. Anledningen till att använda LASSO är att välja ut de bakgrundsfaktorer som påverkar mest. LASSO minskar variansen (ökar då precisionen) och är särskilt lämpade för att senare kunna användas vid prediktering. Alla bakgrundsfaktorer skall inte tas med, eftersom de inte påverkar lika mycket, eller inte alls. Det spelar t ex oftast inte någon roll om patienten besökt VC ofta eller inte. Utbildningsnivå har heller inte någon större påverkan.

9. REFERENSER

1/Adam J Rothman, Elizaveta Levina and Ji Zhu

"Sparse Multivariate Regression With Covariance Estimation"

Journal of Computational and Graphical Statistics

Volume 19, Number 4, Pages 947–962

DOI:

10.1198/jcgs.2010.09188

2/ LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE

"The LASSO Logistic Regression Model: Modifications to aid causality assessment for Adverse Events Following Immunization"

3/ Heide Garcia and Peter Filzmoser,

"Multivariate Statistical Analysis using the R Package chemometrics"

Department of Statistics and Probability Theory

Vienna University of Technology, Austria

P.Filzmoser@tuwien.ac.at

November 7, 2011

4/ Sanford Weisberg

"Variable Selection and Regularization"

October 15, 2012

5/ Robert Tibshirani and Michael Saunders,

Stanford University, USA

Saharon Rosset,

IBM T. J. Watson Research Center, Yorktown Heights, USA

Ji Zhu

University of Michigan, Ann Arbor, USA

and Keith Knight

University of Toronto, Canada

[Received September 2003. Final revision August 2004]

"Sparsity and smoothness via the fused lasso"

J. R. Statist. Soc. B (2005)

67, Part 1, pp. 91–108

Webb adresser se

- 1) <http://www.stat.lsa.umich.edu/~jizhu/pubs/Rothman-JCGS10.pdf>
- 2) http://www.lshtm.ac.uk/library/MSc_MS/2009-10/491467.PDF
- 3) <http://cran.r-project.org/web/packages/chemometrics/vignettes/chemometrics-vignette.pdf>
- 4) <http://users.stat.umn.edu/~sandy/courses/8053/handouts/lasso.pdf>
- 5) <http://www.stanford.edu/group/SOL/papers/fused-lasso-JRSSB.pdf>

Litteratur

P. McCULLAGH and J. A. NELDER, Generalized Linear Models,
Second edition 1989

11. ETT PRAKTISKT RÄKNEEXEMPEL

Med hjälp av bakgrundsdata kan man räkna fram sannolikheten för att en patient har en positiv patientupplevd vårdkvalité.

Ett exempel;

E1 Hur värderar du som helhet den vård/behandling du fick ?

Patient 1; med dålig egenupplevd hälsa, har besökt vårdcentralen ofta, är ung, är kvinna, har svenska som modersmål och har universitetsutbildning.

$$P = \frac{\exp^{\alpha+\beta x}}{1+\exp^{\alpha+\beta x}} \quad (P = \text{sannolikheten})$$

$$P = \frac{2,71828^{0,1473+0,5685(0)-0,4741(1)+0,1029(0)+0,4780(1)-0,10(0)}}{1+2,71828^{0,1473+0,5685(0)-0,4741(1)+0,1029(0)+0,4780(1)-0,10(0)}} = 0,4622$$

Patient 2; Utmärkt hälsa, har inte besökt VC tidigare, är gammal, är man, har svenska som modersmål, har universitetsutbildning.

$$P = \frac{2,71828^{0,1473+0,5685(1)-0,4741(0)+0,1029(1)+0,4780(1)-0,10(0)}}{1+2,71828^{0,1473+0,5685(1)-0,4741(0)+0,1029(1)+0,4780(1)-0,10(0)}} = 0,7853$$