Stockholms universitet

Multipel logistisk regression Antagna till friidrottsgymnasier 2012-2015

Eleni Larsson

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2016:28
Matematisk statistik
December 2016

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Matematiska institutionen

Matematisk statistik
Stockholms universitet
Kandidatuppsats **2016:28**
http://www.math.su.se/matstat

Multipel logistisk regression
Antagna till friidrottsgymnasier 2012-2015

Eleni Larsson*

December 2016

Sammanfattning

Under åren 2011-2014 sökte 927 personer till friidrottsgymnasier i Sverige och av alla dessa antogs 453 stycken. De som ansöker till dessa gymnasier ansöker med olika förutsättningar, de är födda olika tid på året, tränar olika mycket, har haft skador och sjukdomar, tillhör olika biologiska kön, är olika långa och väger olika mycket. Alla dessa variabler kan påverka om en ansökande kommer bli antagen eller inte och det är detta vi undersöker i denna rapport. Vi vill ta reda på vilka variabler som påverkar sannolikheten för att en ansökande blir antagen och även hur denna variabel påverkar. Den mest intressanta variabeln att undersöka var när på året en ansökande var född.

Vi använde oss utav en multipel logistisk regressionsmodell för att undersöka hur dessa variabler påverkar. Det vi upptäckte var att när på året en ansökande är född inte har någon inverkan på om den ansökande blir antagen eller inte. Vi upptäckte däremot att kvinnor har enklare att bli antagna än män, att 1-2 pass per vecka i perioden oktober till december två år innan gymnasiet ökar sannolikheten för antagning mot 0 pass per vecka och att 3 eller flera pass under samma period minskar sannolikheten att bli antagen i jämförelse med 0 pass per vecka.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige. E-post: larsson.eleni@gmail.com. Handledare: Jan-Olov Persson och Gustav Alfelt.

Abstract

Between the years 2011-2014 there were 927 people who applied to athletics high schools in Sweden and 453 of those people got accepted. Applicants to these high schools apply with different preconditions, they are born in different months of the year, they practice differently, some of them are injured or have a disease, they have different biological sex and they vary in length and in weights. All these variables can influence on who of all these applicants that can be accepted and it is this we examine in this report. We want to find out which variables that affect the probability that an applicant is accepted and how they affect. The variable we are most interested in is the applicant's month of birth.

We used a multiple logistic regression model to examine how these variables affect the probability to be accepted. What we discovered was that the month an applicant was born in didn't affect the probability. We did however discovered that women are more likely to be accepted than men, that to train 1-2 times a week in the period October to December two years before high school increases the probability to be accepted compared to train 0 times per week and that train 3 or more times a week in the same period decreases the probability to be accepted compared to train 0 times a week.

Förord

Detta examensarbete omfattar 15 högskolepoäng och leder till en kandidatexamen i matematisk statistik vid Matematiska instutionen på Stockholms universitet.

Jag vill tacka mina handledare Jan-Olov Persson och Gustav Alfelt som genom ett stort engagemang bidragit med en värdefull handledning. Jag vill även tacka Anders Rydén som gav mig tillgång till data och för hans insikter inom området av friidrottsgymnasier. Till sist vill jag tacka min sambo Johan Ahlberg, min pappa Jörgen Karlsson och min vän Alva Backström, för deras stöd och värdefulla synpunkter.

Innehåll

1 Introduktion **1**

1.1 Friidrottsgymnasier 1

1.2 Syfte 2

2 Data **2**

2.1 Variablerna 2

2.1.1 Kön 2

2.1.2 Antal träningspass per vecka under fyra olika perioder . . 3

2.1.3 Gymnasium 3

2.1.4 Födelsemånad 3

2.1.5 PT 3

2.1.6 Skada 4

2.1.7 Antagningsår 4

2.1.8 Antal_träningsår 4

2.1.9 Längd 4

2.1.10 Vikt 5

2.1.11 Redigeringar 5

2.2 Saknad data 5

3 Metod **6**

3.1 Multipel logistisk regression 6

3.2 Variansinflationsfaktor 7

3.3 Stegvis variabelselektion 7

3.3.1 Forward Selection 7

3.3.2 Stepwise regression 7

3.3.3 Backward elimination 8

3.4 Akaike information criterion 8

3.5 Hosmer-Lemeshow test 8

3.6 ROC 9

4 Analys och resultat **9**

4.1 Korrelation mellan de förklarande variablerna 9

4.2 Modellkonstruktion 12

4.2.1 Modell 1 12

4.2.2 Modell 2 15

4.3 Val av modell 16

4.4 Beskrivning av vald modell 17

5 Diskussion **18**

1 Introduktion

I Sverige kan man söka till speciella gymnasieskolor där det är möjligt att kombinera sin akademiska karriär med sin friidrottsliga, så kallade friidrottsgymnasier. Det som skiljer ansökande till friidrottsgymnasium ifrån varandra är i kön, antal träningar per vecka, antal år de tränat friidrott, vilket typ av friidrottsgymnasium de ansöker till, vilken månad de är födda, om de har en personlig tränare, om de varit skadade eller sjuka under en längre period, de är olika långa och väger olika mycket. Alla dessa variabler kan ha en inverkan på chansen att bli antagen och vi vill undersöka vilka som har det samt hur denna inverkan ter sig. Det Svenska Friidrottsförbundet är mest intresserade av att veta är om den månad en ansökande är född har en inverkan på att bli antagen. Det är troligt att den månad en ansökande är född har en inverkan då det visats att unga killar som är födda tidigt på året är överrepresenterade på toppnivå inom fotboll i Sverige. De visade att detta även gäller för flickor men då inte lika tydligt [1].

I denna rapport skall vi undersöka vilka av dessa variabler som påverkar chansen för antagning, vi skall även se om dessa variabler ökar eller minskar chansen till antagning för en ansökande. Detta kommer vi att göra genom att använda multipel logistisk regression vilket är en modell som hanterar binära responsvariabler.

1.1 Friidrottsgymnasier

Att kunna kombinera gymnasiestudier med sin idrottsliga satsning är för många ungdomar avgörande för en fortsatt idrottslig satsning under de övre tonåren. Skolverket har godkänt två varianter av specialidrott för gymnasielever, Riks Idrotts Gymnasium (RIG) samt Nationell Idrotts Utbildning (NIU). Svenska Friidrottsförbundet har sju stycken RIG samt 16 st NIU-skolor runt om i Sverige [2].

I de avtal Svenska Friidrottsförbundet har med sina RIG-orter skall lärartätheten vara en heltidsanställd lärare per tionde elev. Lärarna på RIG får bidrag utav Riksidrottsförbundet (RF) för att täcka upp en del av merkostnaden för lärarlöner. Samtliga lärare skall vara anställda av huvudmannen, det betyder av respektive RIG-orts gymnasieförvaltning. Av gymnasieutbildningens 2500 poäng är 400 p obligatoriskt för både RIG och NIU eleverna i ämnet specialidrott. Som utökad studiekurs kan eleverna läsa ytterligare 300 p i specialidrott. RF beslutar om antal platser på RIG och för närvarande har friidrotten 162 platser fördelat över tre år som Svenska Friidrottsförbundet delar ut på sina sju RIG [3].

NIU skiljer sig från RIG på det sätt att lärartätheten är lägre, en 100 % tjänst på tjugofyra stycken elever samt att det enbart erbjuds tre specialidrottstillfällen per vecka för eleverna. I de avtal som Svenska Friidrottsförbundet skriver med NIU-orterna är trepartsavtal mellan Svenska Friidrottsförbundet, NIU-kommunen samt en lokal förening för att garantera kvällsträning för eleverna. Svenska Friidrottsförbundet beslutar om antalet platser på NIU-orterna [4].

Antagningen till de två olika gymnasieformerna sker dock på samma sätt. De ansökande bedöms via sina idrottsliga resultat samt intervjuer och talangbedömning. Testerna och intervjuerna fortlöper fram till den 15:e januari samma år det är tänkt att en ansökande skall påbörja sin gymnasieutbildning. Efter denna ansökningsprocess får de ansökande besked om de blivit antagna eller inte på den idrottsliga delen [3] och [4].

1.2 Syfte

Syftet med denna rapport är att finna vilka faktorer som påverkar om en ansökande kommer in och hur dessa faktorer inverkar. Vi skall undersöka detta genom att använda logistisk regression. Genom att ta fram dessa faktorer och deras inverkan kan en uppfattning skapas av vilka ungdomar som med större sannolikhet kommer antas till ett friidrottsgymnasium. Det Svenska Friidrottsförbundet är mest intresserade av att veta är om de som är födda tidigt på året har en större chans att bli antagna än de som är födda senare, men andra faktorer kommer också att undersökas.

2 Data

Data som rapporten grundas på kommer från Svenska Friidrottsförbundet. Det består av insamlade uppgifter om 927 personer som ansökt till friidrottsgymnasier under åren 2011-2014. Av dessa kom 453 personer in, dvs. 49 % blev antagna. Genom sin ansökan till friidrottsgymnasierna svarar de på ett antal frågor om sin person och träning. Dessa svar ligger till grund för data som används. Data innehåller även information om huruvida en ansökande kom in eller ej. Det gör att data delas in i en binär indelning av antagen och nekad antagning. Tyvärr saknas information kring om en ansökande blivit nekad p.g.a. för låga betyg eller om hen tackat nej till sin erbjudna plats, men detta berörs mer i diskussionen.

2.1 Variablerna

Data består utav 927 observationer med värden för 13 stycken förklarande variabler som är antingen kategoriska eller kontinuerliga samt en binär responsvariabel. I Tabell 1 ser vi summeringsstatistik över de olika variablerna så att vi kan skapa oss en första inblick i hur de inverkar på responsvariabeln. Responsvariabeln förklarar om en ansökande till ett friidrottsgymnasium blivit antagen eller inte och är döpt till Ansökande.

2.1.1 Kön

Den kategoriska variabeln Kön består av uppgifterna om en ansökandes biologiska kön. Det är betydligt fler kvinnor som ansöker till friidrottsgymnasier än män, 523 stycken kvinnor mot 377 män. I samhällsdebatten har en individs biologiska kön en central roll och även om vi inte är säkra på om könet har en inverkan på antagningen till ett friidrottsgymnasium väljer vi ändå att undersöka det.

2.1.2 Antal träningspass per vecka under fyra olika perioder

När ansökningen sker anger de ansökande hur många träningar de har utfört per vecka under fyra olika perioder. Perioderna är mars till april året innan de börjar studera på gymnasiet, oktober till december två år innan de börjar på gymnasiet, mars till april två år innan gymnasiet och oktober till december tre år innan de börjar gymnasiet. Dessa uppgifter består variablerna Vår_träning2, Höst_träning2, Vår_träning1 och Höst_träning1 av. Att träna flera gånger i veckan kan tänkas påverka eftersom en person som tränar mera bör besitta en bättre teknik samt en starkare och uthålligare kropp. Däremot är det under hösten som en friidrottare utför sin tyngsta träning och därför tros höstträning ha en större inverkan på en ansökandes chans till antagning än vårträningen.

2.1.3 Gymnasium

Det existerar för närvarande 23 stycken friidrottsgymnasier i Sverige och en ansökande har i sin ansökan angett vilket gymnasium de ansöker till i första hand, andra hand och tredje hand. Dessa 23 gymnasier tillhör två kategorier, nämligen RIG och NIU. Det är troligen svårare för en ansökande att bli antagen till ett RIG än ett NIU, då RIG:en är mer specialiserade med högre tränartäthet på färre antal gymnasier. Eftersom det kan existera en skillnad mellan de två olika typerna utav friidrottsgymnasier har vi skapat variabeln Gymnasier som beskriver om en ansökande ansökt till RIG, NIU eller till både RIG och NIU.

2.1.4 Födelsemånad

Som vi nämnde i inledningen har man visat att ungdomar som är födda tidigt på året är överrepresenterade på toppnivå inom fotboll i Sverige [1]. Med tanke på detta är det troligt att vilken månad en ansökande är född påverkar hens chans att bli antagen. Personerna som ansökte till friidrottsgymnasierna angav ej vilken månad de är födda i men dessa uppgifter går att utläsa från deras personnummer. Vi har därigenom skapat variabeln Födelsemånad. När data undersöktes såg vi att 60.7 % var födda på den första halvan av året och att 38.2 % under sista halvan av året.

2.1.5 PT

De som ansöker har angett om de har en personlig tränare eller inte och utav dessa uppgifter har variabeln PT skapats. Ansökande som har en personlig tränare och inte tränar med en grupptränare har troligen fått mer individuellt inriktad hjälp som kan vara till deras fördel. Det är färre ansökande som har en personlig tränare än de som har en grupptränare, 34.3 % har en personlig tränare och 64.2 % har det inte.

2.1.6 Skada

Vissa utav de som ansöker har haft längre uppehåll i sin träning p.g.a. en skada eller sjukdom. Ett längre avbrott i träningen skulle kunna göra att de inte besitter samma tekniska förmåga eller styrka som sina medsökande och därmed ha en minskad chans för att bli antagna. Vi har därför skapat variabeln Skada som består av uppgifter om en ansökande varit skadad eller sjuk under en längre period. Det är 36.0 % som har varit skadade eller sjuka, dvs. majoriteten av de ansökande har inte varit skadade eller sjuka under en längre period.

2.1.7 Antagningsår

Data som vi har är över flera år, 2011-2014. Som med variabeln Kön är vi här inte säkre på om vi kommer finna någon inverkan på chansen för att bli antagen. Vi har ändå valt att undersöka om året en ansökande är tänkt att börja på ett friidrottsgymnasium har en påverkan. Variabeln som innehåller informationen om vilket år en ansökande är tänkt att börja studera på ett friidrottsgymnasium är döpt till Antagningsår.

2.1.8 Antal_träningsår

I ansökan till friidrottsgymnasiet har de ansökande angivit vilket år de började med friidrott. En del ansökande har inte angivit ett exakt år de började med friidrott och vi har valt att använda det senare året om de angivit två tänkbara startår. Vi är inte intresserade av att se om vilket år en ansökande började träna friidrott har en inverkan på chansen för antagning eftersom de ansökt fyra olika år. Det vi däremot finner intressant är att se om en person som tränat friidrott längre har en ökad chans för antagning. En person som tränat längre bör besitta bättre teknik, styrka och uthållighet. Vi kan skapa en sådan variabel genom att subtrahera året de började med friidrott från året det är tänkt att de skall börja studera på ett friidrottsgymnasium. Variabeln som innehåller denna information har vi valt att döpa till Antal_träningsår. En ansökande har i genomsnitt tränat friidrott runt 6 år innan skolstart.

2.1.9 Längd

Det diskuteras ofta inom friidrotten att vara lång är till en fördel. P.g.a. detta så väljer vi att undersöka om längden inverkar på chansen för antagning. Variabeln med de ansökandes längd heter Längd i denna rapport och vi har valt att använda medelvärdet för de personer som angivit sin längd i ett intervall. Vi har även valt att mäta längden i centimeter. En ansökande är i genomsnitt 173 cm lång.

2.1.10 Vikt

Inom flertalet grenar är en lätt kroppsvikt en fördel och därför väljer vi att lägga in informationen om en ansökandes vikt i variabeln Vikt. Vi har valt att använda medelvärdet om en person angivit sin vikt i ett intervall och enheten kilogram används. En ansökande väger i genomsnitt 60.5 kg.

2.1.11 Redigeringar

Förutom ändringar i data som nämnts i beskrivningen av vardera variabel har vi lagt till NA där ingen information angivits. Flera personer lämnade in uppgifter två gånger samma år och där har den senast inlämnade informationen ansetts som den korrekta.

Tabell 1: Tabell med summeringstatestik för de 927 observationerna över alla förklarande variabler. Värden med $*$ bakom sig representerar medelvärde.

Variabel		Ansökande	Antagna	Icke antagna	Antal saknade observationer
Kön	Man Kvinna	40.7 % 56.4 %	36.0 % 60.5 %	45.1 % 52.5 %	27
Vår_träning2	0-6	4.06 $*$	4.00 $*$	4.12 $*$	0
Höst_träning2	0-6	3.77 $*$	3.65 $*$	3.88 $*$	0
Vår_träning1	0-6	3.42 $*$	3.34 $*$	3.49 $*$	0
Höst_träning1	0-6	2.98 $*$	2.95 $*$	3.02 $*$	0
Gymnasium	RIG NIU Båda	44.2 % 28.3 % 26.8 %	45.7 % 28.5 % 25.2 %	42.8 % 28.1 % 28.3 %	7
Födelsemånad	0-12	5.60 $*$	5.48 $*$	5.72 $*$	10
PT	Personlig tränare Grupptränare	34.3 % 64.2 %	32.2 % 66.0 %	36.3 % 62.4 %	14
Skada	Skadad Frisk	36.0 % 62.0 %	37.3 % 59.8 %	34.8 % 64.1 %	18
Antagningsår	2012 2013 2014 2015	23.3 % 24.9 % 24.7 % 27.1 %	26.0 % 24.5 % 23.4 % 26.0 %	20.7 % 25.3 % 25.9 % 28.1 %	0
Antal_träningsår	0-12	6.43 $*$	6.35 $*$	6.50 $*$	24
Längd	151-197	173 $*$	172 $*$	173 $*$	9
Vikt	37-110	60.5 $*$	60.6 $*$	60.3 $*$	16

2.2 Saknad data

I Tabell 1 ses att vi har variabler som innehåller saknad data. Då detta kommer leda till problematik under analysen väljer vi att handskas med dessa avbrott på SAS:s standardsätt. SAS undersöker om det existerar saknad data bland variablerna i en uppbyggd modell och raderar sedan raderna där saknad data finns [5].

3 Metod

I denna del utav rapporten kommer en beskrivning ske utav metoderna som vi senare använder i delen Analys och resultat.

3.1 Multipel logistisk regression

När responsvariabeln är binär och flertalet förklarande variabler existerar kan en multipel logistisk regressionsmodell användas. Denna modell beräknar sannolikheten att ett utfall inträffar givet att vi vet värdet på de förklarande variablerna och värdena på parametrarna. Denna sannolikhet betecknas som $\pi(\mathbf{x})$ och beräknas genom formeln:

$$\pi(\mathbf{x}) = \frac{exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_j x_j)}{1 + exp(\alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_j x_j)}. \quad (1)$$

Här är α interceptet och β_j är effekten den förklarande variabeln x_j har på $\pi(\mathbf{x})$. När alla förklarande variabler förutom x_j hålls konstanta ökar $\pi(\mathbf{x})$ om $\beta_j > 0$ när x_j ökar och om $\beta_j < 0$ minskar $\pi(\mathbf{x})$ när x_j ökar.

En multipel logistisk regressionsmodell besitter inte ett linjärt samband mellan $\pi(\mathbf{x})$ och de förklarande variablerna men detta går att få om log-oddset (logit) tas på bägge sidor utav Funktionen (1) [6, s. 119-120].

Logit är logaritmen utav oddset (Ω). För att beräkna oddset divideras sannolikheten att en händelse inträffar (π) med sannolikheten att en händelse inte inträffar ($1 - \pi$).

$$\Omega = \frac{\pi}{1 - \pi}.$$

När oddset är större än 1 är det mer troligt att en händelse inträffar än att den inte gör det och om oddset är mindre än 1 gäller det omvända. Om till exempel $\Omega = 2$ är det två gånger troligare att en händelse sker än att den inte gör det [6, s. 44-45].

Genom att använda logit på Funktion (1) fås:

$$logit(\pi(\mathbf{x})) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_j x_j. \quad (2)$$

För att kunna tolka β_j i Funktion (2) tas exponenten utav bägge sidor, detta resulterar i att $exp(\beta_j)$ är odds-kvoten mellan två olika nivåer och en enhets ökning på den förklarande variabeln x_j om alla andra förklarande variabler hålls konstanta [6, s. 164-165].

Odds-kvoten fås genom att dividera två stycken odds med varandra [6, s. 44-45].

$$\text{Odds-kvot} = \frac{\Omega_1}{\Omega_2} = \frac{\pi_1/(1-\pi_1)}{\pi_2/(1-\pi_2)}.$$

3.2 Variansinflationsfaktor

Om det existerar kollineära eller nästan kollineära förklaringsvariabler kan förväntade signifikanta förklarande variabler ses som icke-signifikanta. Korrelerade förklaringsvariabler kan upptäckas genom att använda variansinflationsfakorn (VIF). Denna räknas ut genom formeln:

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

där R_j^2 är förklaringsgraden som visar hur mycket av variationen i en förklarande variabel som beskrivs utav övriga förklarande variabler i en multipel linjär regressionsmodell. VIF tar värdet 1 om det inte existerar någon korrelation mellan en förklarande variabel och de övriga förklarande variablerna. Ett VIF-värde så nära 1 som möjligt är därför att föredra [7, s. 71-73].

3.3 Stegvis variabelselektion

Det existerar stegvisa procedurer för att välja ut vilka förklarande variabler som bör vara med i modellen. Dessa procedurer eliminerarar eller adderar förklarande variabler stegvis tills ett visst stoppkriterium är uppfyllt. Vi kommer att använda oss utav tre av dem, forward selection, backward elimination samt stepwise regression. De tre procedurerna behöver dock inte leda till samma resultat [7, s. 69-70] .

3.3.1 Forward Selection

Denna procedur utgår ifrån en modell med enbart intercept, dvs. inga förklarande variabler existerar. I varje steg läggs en ny förklarande variabel till. Den variabel som tillkommer är den variabel som är mest signifikant av de kvarvarande variablerna. Denna process fortsätter tills det inte existerar några variabler som uppfyller signifikansnivån som angivits på förhand [7, s. 70] .

3.3.2 Stepwise regression

Denna procedur fungerar som forward selection med undantaget att den i varje steg undersöker om variablerna i modellen fortfarande är signifikant skilda från noll. Om dessa inte är det elimineras de ur modellen. Signifikansnivån är även här angiven på förhand [7, s. 70-71].

3.3.3 Backward elimination

Till skillnad från i forward selection utgår backward elimination ifrån en modell som innehåller alla förklarande variabler. I denna procedur utesluts i varje steg en variabel. Variabeln utesluts om det visar sig att variabeln inte har någon signifikant inverkan på responsvariabeln. Om det existerar flera sådan variabler utesluts den variabeln som är minst signifikant. Denna process fortgår tills alla variabler inverkar på responsvariabeln med den från början givna signifikansnivån [7, s. 70].

3.4 Akaike information criterion

För att avgöra hur bra en modell är kan man använda sig utav Akaike information criterion (AIC). AIC är ett mått på hur väl en modell passar data.

$$AIC = -2(\text{maximerad log likelihood} - \text{antal variabler}).$$

Modellen som har lägst AIC värde är modellen som oftast välj. Däremot är en så enkel modell som möjligt att föredra vilket kan leda till att en modell med aningen högre AIC värde väljs istället [6, s. 212].

3.5 Hosmer-Lemeshow test

För att avgöra hur väl en modell passar data kan man använda sig utav ett Hosmer-Lemeshow test. I Hosmer-Lemeshow test skapas g stycken grupper, där antalet grupper beror på antalet förklarande variabler. Om antalet grupper underskrider 6 stycken ger testet ofta ett icke pålitligt resultat.

Den logiska regressionsmodellen ger oss estimerade sannolikheter att den binära responsvariabeln Y kommer anta värdet 1 för olika värden på de förklarande variablerna. Alla möjliga sannolikheter från sannolikheten 0 till sannolikheten 1 kan delas in i intervall som går i stigande ordning tex. ([0-0.2[, [0.2-0.4[, [0.4-0.6[, [0.6-0.8[, [0.8-1]). De g grupper som skapas i Hosmer-Lemeshow test består utav antalet estimerade värden i de olika sannolikhets intervallen. Antalet estimerade sannolikheter i vardera grupp bör dock överstiga 5 för att testet skall ge ett pålitligt resultat. De värden som fås fram används sedan för att beräkna Hosmer-Lemeshow goodness-of-fit statistika $\hat{C}$

$$\hat{C} = \sum_{k=1}^{g} \frac{(o_k - n'_k \overline{\pi}_k)^2}{n'_k \overline{\pi}_k (1 - \overline{\pi}_k)}.$$

Här är n'_k värdet på antalet estimerade sannolikheter i grupp k, $\overline{\pi}_k$ är medelvärdet på de estimerade sannolikheterna i grupp k och o_k är antalet gånger responsvariabeln antar värdet 1 i grupp k.

Om modellen är en modell som passar data väl är $\hat{C}$ Chi-Två fördelat med $g-2$ frihetsgrader [8, s. 147-156].

3.6 ROC

För att undersöka en modells prediktiva förmåga kan man använda en receiver operating characteristic (ROC) kurva. Denna kurva skapas genom att plotta sensitiviteten (sannolikheten för positivt utfall när positivt utfall är det korrekta) som funktion av 1-specificitet (sannolikheten för negativt utfall när negativt utfall är det korrekta) för alla tänkbara brytpunkter mellan 0 och 1 [6, s. 224].

En brytpunkt kan beskrivas genom att vi låter:

$$\hat{Y} = \begin{cases} 1 & \text{Om } \hat{\pi}(\mathbf{x}) \geq \pi_0 \\ 0 & \text{Annars} \end{cases}$$

Här är $\hat{Y}$ det skattade värdet på responsvariabeln, $\hat{\pi}(\mathbf{x})$ den skattade sannolikheten som fås utav modellen och π_0 är en brytpunkt [8, s. 160-164] och [9, s. 11].

ROC kurvan som skapas går från punkten $(0,0)$ till punkten $(1,1)$ och formen är ofta konkav [6, s. 224].

För att mäta den prediktiva förmågan beräknas arean under grafen, där större area betyder bättre prediktiv förmåga. Approximativt kan man anta, enligt [8, s. 162]:

Arean $= 0.5$:	Modellen saknar prediktiv förmåga.
$0.7 \leq$ Arean < 0.8 :	Modellens prediktiva förmåga anses vara acceptabel.
$0.8 \leq$ Arean < 0.9 :	Modellens prediktiva förmåga anses vara god.
Arean ≥ 0.9 :	Modellens prediktiva förmåga anses vara utmärkt.

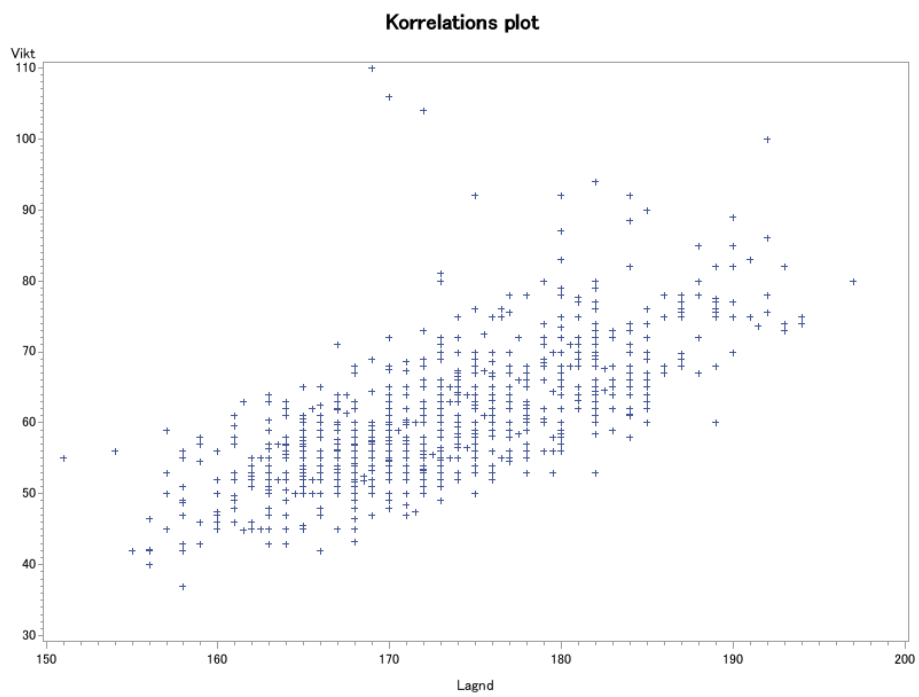
4 Analys och resultat

Vi skall i denna del tillämpa metoderna som är beskrivna i delen Metod för att konstruera, välja och analysera modeller.

4.1 Korrelation mellan de förklarande variablerna

Om det existerar nästan korrelerade förklarande variabler kan en förklarande variabel som skulle ha en signifikant inverkan på responsvariabeln ge ett icke-signifikant resultat. Det är p.g.a. detta som vi vill undersöka om det förekommer nästan korrelerade förklarande variabler i vårt data. Vi undersöker detta genom att studera korrelations-plottar mellan de förklarande variablerna Längd, Vikt, Födelsemånad och Antal_träningsår; tabeller mellan de förklarandevariablerna Höst_träning1, Höst_träning2, Vår_träning1, Vår_träning2, Kön, Gymnasium, Antagningsår och Skada; samt genom att se på alla de förklarande variablernas VIF-värden.

Det är enbart i korrelations ploten mellan förklarande variablerna Längd och Vikt som vi finner en påtaglig korrelation. I Figur 1 ser vi att de ansökande som är längre tenderar till att väga mer. Vi löser detta korrelations problem genom beräkna de ansökandes BMI vilket är ett mått på en persons vikt (kg) i förhållande till hens längd (m) i kvadrat [10]. BMI kan ha en inverkan på chansen för att bli antagen eftersom en lång och lätt kropp anses vara en fördel i flertalet friidrottsgrenar. Vi döper den förklarande variabeln med information om BMI till BMI.



Figur 1: Korrelations-plot över de ansökandes längd och vikt.

I Tabell 2 och 3 kan vi se att det existerar en korrelation mellan variablerna Höst_träning1 och Vår_träning1 samt mellan Höst_träning2 och Vår_träning2. Höstträningen tror vi har en större inverkan på responsvariabeln och därför väljer vi att ta bort de förklarande variablerna Vår_träning1 och Vår_träning2. Från de andra tabellerna finner vi ingen påtaglig korrelation.

Tabell 2: 7x7 tabell över variablerna Höst_träning1 och Vår_träning1.

Höst_träning1\Vår_träning1	0	1	2	3	4	5	≥6
0	31	0	4	3	2	0	1
1	0	7	12	9	2	0	0
2	1	4	103	119	4	2	0
3	0	2	1	191	134	9	0
4	0	2	0	5	108	49	7
5	0	1	1	1	10	48	14
≥6	0	0	0	0	1	5	24

Tabell 3: 7x7 tabell över variablerna Höst_träning2 och Vår_träning2.

Höst_träning2\Vår_träning2	0	1	2	3	4	5	≥6
0	8	0	5	3	1	0	0
1	0	7	3	2	2	1	0
2	0	2	30	42	6	2	0
3	0	1	9	123	103	15	0
4	1	2	3	17	181	105	11
5	0	1	0	7	20	141	27
≥6	0	1	1	1	4	12	54

När vi studerar VIF-värdena i Tabell 4 med alla de förklarande variablerna som presenteras i sektionen 2.1 ser vi samma resultat som vi fann för korrelations-plottarna och i tabellerna. Vi vill även undersöka om det finns några nästan korrelerade förklarande variabler efter att vi tagit bort förklarande variablerna Vikt, Vår_träning1 och Vår_träning2 samt lagt till den förklarande variabeln BMI. Det undersöks genom att beräkna VIF-värdena med den nya uppsättningen av förklarande variabler. Vi ser i Tabell 4 att alla förklarande variablers VIF-värden nu ligger runt 1.

Tabell 4: Tabell över VIF-värdena för de förklarande variablerna. Kolumnen VIF 1 innehåller VIF-värdena med de ursprungliga förklarande variablerna. Kolumnen VIF 2 innehåller VIF-värdena med de nya förklarande variablerna. De variabler som ej VIF-värde beräknas på är angivna med x.

Variable	VIF 1	VIF 2
Kön	1.69	1.69
Vår_träning2	2.33	x
Höst_träning2	3.06	1.67
Vår_träning1	3.74	x
Höst_träning1	2.97	1.71
Gymnasium	1.03	1.03
Födelsemånad	1.02	1.02
PT	1.05	1.04
Skada	1.04	1.04
Antagningsår	1.03	1.03
Antal_träningsår	1.13	1.12
Längd	2.59	1.63
Vikt	2.15	x
BMI	x	1.05

4.2 Modellkonstruktion

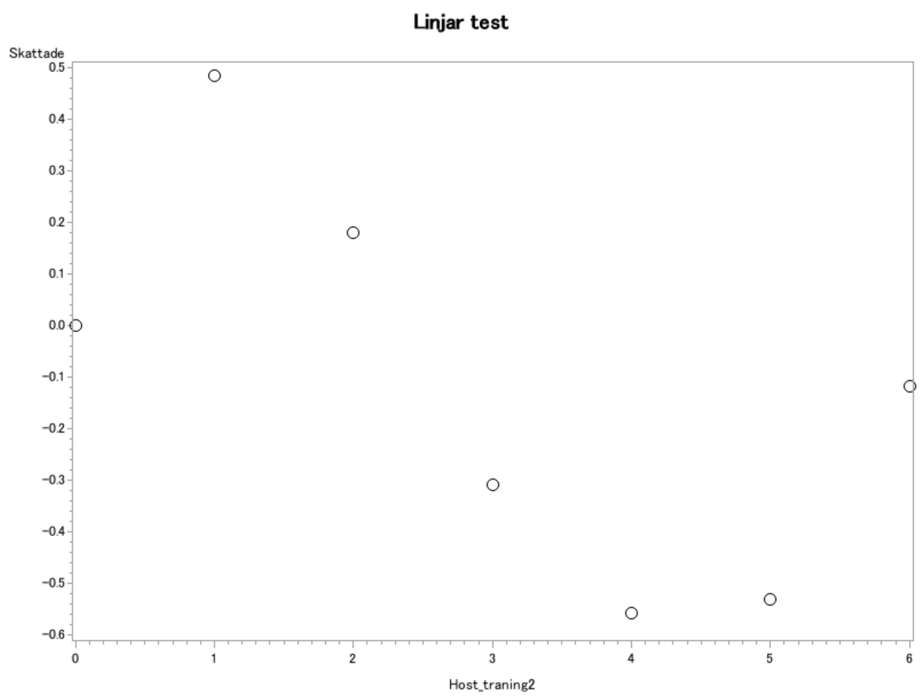
I denna del av rapporten skall vi fokusera på att skapa våra modeller. Från dessa modeller som skapas kommer vi välja ut den modell som vi anser är den som beskriver vårt data bäst. Vi kommer använda oss utav forward regression, backward elimination, stepwise selection samt sunt förnuft för att såväl skapa som välja en modell.

4.2.1 Modell 1

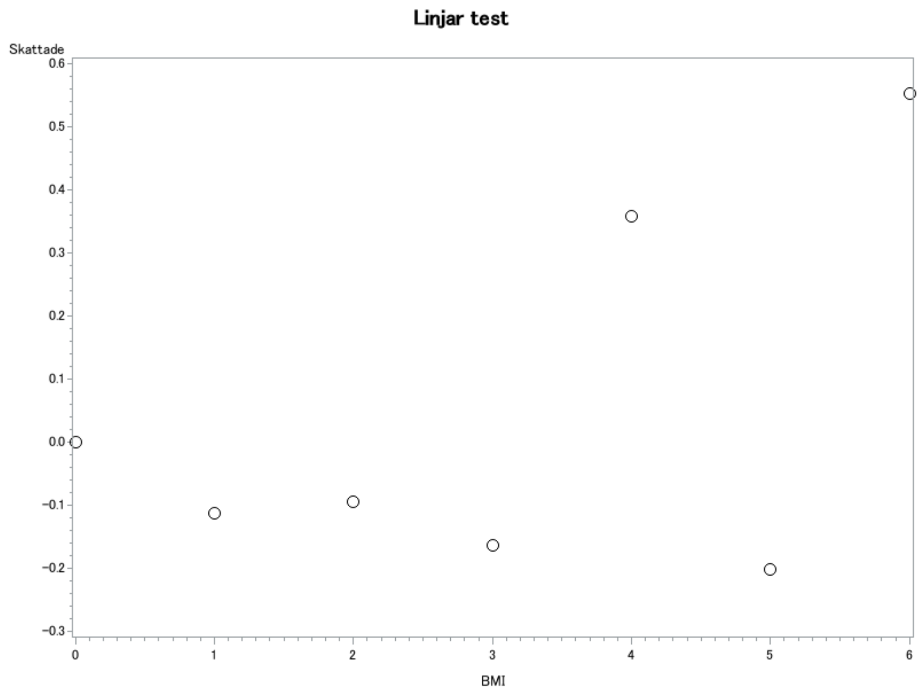
För att skapa en modell börjar vi med att undersöka en multipel logistisk regressionsmodell som innehåller alla förklarande variabler som vi kom fram till i sektionen 4.1. Vi gör detta för att skapa en första uppfattning utav de förklarande variablerna. Modellens estimerade effekter går att finna i Appendix 1. Från denna modell finner vi att variablerna Höst_träning2, BMI och Kön verkar vara variabler som har en signifikant inverkan på responsvariabeln. Däremot finner vi att variabeln Födelsemånad har ett p-värde på 0.50 vilket betyder att den inte har en signifikant inverkan på om en ansökande blir antagen.

Efter detta har gjorts undersöker vi data vidare genom att använda de tre variabelselektions-procedurerna som angavs i sektion 3.3 på samma förklarande variabler som ovan. Den på förhand angivna signifikansnivån sätts till 0.1 i backward elimination, forward selection och stepwise regression. Vi får från alla tre procedurerna samma modell. Modellen som fås innehåller de förklarande variablerna Höst_träning2, BMI och Kön.

Modellen innehåller två stycken kontinuerliga förklarande variabler och vi behöver undersöka om dessa har linjära skattade värden. Om variablerna inte har det behöver de antingen transformeras om så de blir linjära eller så delar man in variablerna i kategorier och därmed fås en kategorisk variabel istället. För att undersöka detta delar vi in våra två kontinuerliga variabler i kategorier och skapar sedan en modell med Höst_träning2, BMI och Kön där Höst_träning2 och BMI nu är kategoriska. De skattade effekterna för de två variablerna Höst_träning2 och BMI finner vi i Figur 2 och 3 och där ses att de inte är linjära. Vi väljer att göra om dessa två variabler till kategoriska variabler. BMI delas in i fyra stycken kategorier, underviktig, normalviktig, överviktig och fetma [10]. Höst_träning2 delas in i sju stycken kategorier, 0 träningar per vecka, 2 träningar per vecka, 3 träningar per vecka, 4 träningar per vecka, 5 träningar per vecka och 6 eller fler träningar per vecka.



Figur 2: Plot över Höst_träning2 skattade effekter.



Figur 3: Plot över BMI skattade effekter. BMI delas in i under 18, [18-20[, [20-22[, [22-24[, [24-26[, [26-28[och över 28.

När detta görs får vi inte längre någon signifikant effekt på BMI och vi väljer därför att plocka bort BMI. Vi får nu enbart en modell med de förklarande variablerna Kön och Höst_träning2. Vi döper denna modell till Modell 1 och modellen går att finna i Tabell 5.

Tabell 5: Tabell över Modell 1 estimerade effekter, standard error samt p-värde.

Variable		$\hat{\beta}$	SE	p-värde
Intercepte		0.249	0.129	0.053
Kön	Man	0	-	-
	Kvinna	0.196	0.069	0.005
Höst_träning2	0	0	-	-
	1	0.542	0.484	0.263
	2	0.184	0.236	0.435
	3	-0.294	0.169	0.081
	4	-0.555	0.161	0.001
	5	-0.518	0.186	0.006
	≥ 6	-0.118	0.237	0.005

4.2.2 Modell 2

Vi vill undersöka om det finns några samspelseffekter som inverkar på chansen för att bli antagen. Det finns totalt 11 stycken förklarande variabler vilket leder till att det existerar ett stort antal möjliga samspelseffekter. Vi bedömer att 8 stycken av dem är intressanta att undersöka.

En ungdom som är född tidigare på året än en annan är mer fysisk utvecklad. Detta kan leda till att en person som är född tidigt på året inte behöver utföra samma träningsmängd för att uppnå liknande fysiska nivå som en som är född senare. Detta resonemang leder till att vi vill undersöka om det existerar ett samspel mellan variablerna Höst_träning1 och Födelsemånad, Höst_träning2 och Födelsemånad, Antal_träningsår och Födelsemånad samt Skada och Födelsemånad.

Det skulle kunna vara så att vissa år är det lättare att bli antagen som kvinna och andra år som man. Därför undersöker vi om det existerar en samspelseffekt mellan variablerna Kön och Antagningsår.

Ansökande som tränat friidrott i flera år än andra lär vara mer tekniska och kan besitta en bättre fysisk förmåga. Hypotesen är att en som tränat en längre tid kommer klara av att komma tillbaka efter en längre periods uppehåll. Detta eftersom den redan besitter bättre grunder. Grundat på detta undersöker vi om det existerar en samspelseffekt mellan variablerna Antal_träningsår och Skada.

En personlig tränare kan ge sin aktiv bättre individanpassad träning än en grupptränare och är troligen närmare sin aktiv. Därför bör en personlig tränare enklare upptäcka en misstänkt skada på den aktive tidigare. Detta gör att vi tror att en aktiv som har en personlig tränare kommer prestera bättre än en aktiv med en grupptränare efter att den aktiva haft ett längre uppehåll p.g.a. skada eller sjukdom. Därför undersöks det om det existerar en samspelseffekt mellan variablerna Skada och PT.

Vi börjar med de förklarande variablerna i Modell 1 och lägger till en samspelseffekt i taget samt dess huvudeffekter för att undersöka om den har en signifikant effekt på responsvariabeln. Genom att göra detta finner vi att två utav de åtta möjliga samspelseffekterna har signifikant effekt. De två samspelseffekter som har signifikant effekt är Kön och Antagningsår samt Skada och PT. Modellen som innehåller dessa samspel och de förklarande variablerna Höst_träning2, Skada, PT, Antagningsår och Kön döper vi till Modell 2 och modellens skattade effekter går att finna i Tabell 6.

Tabell 6: Tabell över Modell 2 estimerade effekter, standard error samt p-värde.

Variable		$\hat{\beta}$	SE	p-värde
Intercepte		0.247	0.143	0.083
Kön	Man	0	-	-
	Kvinna	0.215	0.072	0.003
Höst_träning2	0	0	-	-
	1	0.633	0.493	0.199
	2	0.169	0.245	0.490
	3	-0.329	0.178	0.066
	4	-0.601	0.171	0.000
	5	-0.499	0.198	0.012
	≥ 6	-0.077	0.251	0.7581
PT	Grupptränare	0	-	-
	Personlig tränare	-0.055	0.077	0.475
Skada	Frisk	0	-	-
	Skadad	0.113	0.076	0.137
Antagningsår	2012	0	-	-
	2013	-0.105	0.128	0.413
	2014	-0.116	0.125	0.355
	2015	-0.059	0.121	0.625
Kön*Antagningsår	Man*2012	0	-	-
	Man*2013	0	-	-
	Man*2014	0	-	-
	Man*2015	0	-	-
	Kvinna*2012	0	-	-
	Kvinna*2013	0.059	0.126	0.641
	Kvinna*2014	0.212	0.124	0.088
	Kvinna*2015	-0.265	0.121	0.028
Skada*PT	Frisk*Personlig tränare	0	-	-
	Frisk*Grupptränare	0	-	-
	Skadad*Grupptränare	0	-	-
	Skadad*Personlig tränare	0.151	0.076	0.048

4.3 Val av modell

Vi skall nu välja en modell. Detta kommer vi göra genom att studera de två modellernas AIC-värden. Däremot stöter vi på problem eftersom vårt data innehåller saknade värden. En modell med mindre antal värden generellt ger ett högre AIC-värde och därför går ej modeller med olika många observationer att jämföra. Vi löser detta genom att radera samma observationer från Modell 1 som raderas från Modell 2. Vi kan se i Tabell 7 att våra modeller besitter liknande AIC-värden och vi väljer därför Modell 1 som vår modell eftersom vi anser att denna är betydligt enklare än Modell 2. Modell 1 har även ett lägre AIC-värde än Modell 2.

Tabell 7: Tabell över AIC-värdena på Modell 1 och Modell 2 samt antalet observationer i modellerna när AIC beräknas.

Modell	AIC-värde	Antal observationer
Modell 1	1196.021	871
Modell 2	1196.432	871

4.4 Beskrivning av vald modell

I sektionen 4.3 valde vi Modell 1. Vi skall i denna sektion analysera de förklarande variablernas inverkan på responsvariabeln, hur väl modellen passar data samt undersöka modellens prediktiva förmåga.

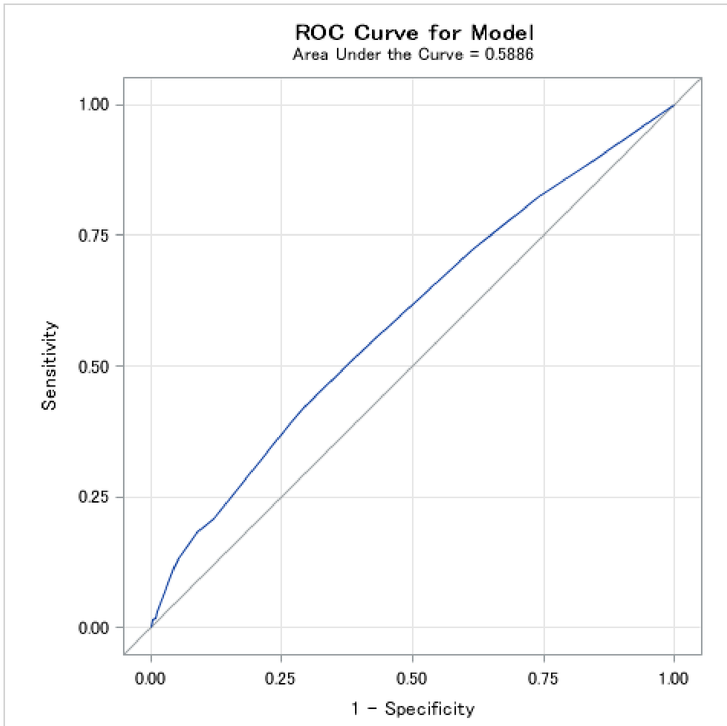
I Tabell 8 ser vi att oddset att bli antagen om du är man multipliceras med en faktor på 1.22, dvs kvinnor har en högre sannolikhet att bli antagna än män. Vi kan även utläsa från Tabell 8 att oddset att bli antagen ökar om en ansökande tränar 1-2 träningspass per vecka under perioden oktober till december, två år innan den ansökande börjar på gymnasiet. Oddset för att bli antagen sjunker däremot under samma träningsperiod om den ansökande tränar 3 eller fler pass per vecka. Om vi studerar konfidensintervallen i Tabell 8 ses att vi inte kan säga att odds-kvoten för att träna 1,2,3 och 6 eller fler pass per vecka är statistisk säkerhetsställa då de innehåller värdet 1.

Tabell 8: Tabell över Modell 1 odds kvoter och deras konfidensintervall.

Variabel		Odds kvot	Konfidensintervall
Kön	Kvinna	1.22	$(1.06, 1.39)$
Höst_träning2	1	1.72	$(0.667, 4.433)$
	2	1.20	$(0.757, 1.91)$
	3	0.745	$(0.536, 1.04)$
	4	0.574	$(0.419, 0.788)$
	5	0.596	$(0.414, 0.859)$
	≥ 6	0.889	$(0.558, 1.41)$

Hur väl Modell 1 passar data undersöker vi genom att studera Hosmer-Lemeshow test för denna modell. P-värdet för Hosmer-Lemeshow test är 0.989 och detta betyder att Modell 1 passar data bra. För Modell 2 gjordes även Hosmer-Lemeshow test och även denna modell passar data väl med ett p-värde på 0.515.

Vi skall nu undersöka hur väl Modell 1 predikterar genom att skapa en ROC kurva och beräkna arean under dennes graf. I Figur 4 ser vi att arean under grafen är 0.589 och vi kan utav detta dra slutsatsen att Modell 1 inte är en god prediktionsmodell. Vi gjorde samma sak för Modell 2 och denna modells värde på arean under grafen är 0.625. Modell 2 är som Modell 1 ingen god prediktionsmodell.



Figur 4: ROC kurva för vald modell, Modell 1.

5 Diskussion

När detta arbete påbörjades trodde vi att vilken månad en ansökande är född i skulle besitta en signifikant effekt men en sådan existerade inte (p-värde = 0.5). Vi trodde att de som är födda tidigt på året skulle ha en större chans till antagning än de som är födda senare. Att vi inte kunde påvisa någon signifikant effekt kan bero på att flertalet utav de ansökande redan tillhör toppnivån i Sverige i sina respektive grenar. Flertalet av de ansökande är födda under första halvåret vilket kanske betyder att de som är födda senare inte tillhör en så pass hög nivå att de anser det värt att ansöka, dvs. en utsållning har redan skett. Detta kan vara en anledning till varför ingen signifikant effekt kunde påvisas för variabeln Födelsemånad.

En kvinna som ansöker har större sannolikhet att bli antagen än en man. Det är fler kvinnor än män som ansöker, 523 kvinnor mot 377 män. Detta skulle kunna bero på att flera kvinnor tränar och tävlar i friidrott och därav är konkurrensen hårdare på dam sidan om höga placeringar på tävlingar. Att konkurrensen är högre gör att kvinnor behöver prestera på en bättre resultatnivå för topp placeringar på tävlingar än männen.

Variabeln Höst_träning2 fick en signifikant effekt vilket är rimligt då det är under denna period personer som håller på med friidrott lägger sin uppbyggnadsträning. Att det blev just Höst_träning2 och inte Höst_träning1 är inte heller underligt då det är under oktober-december två år innan gymnasiet som representerar uppbyggnadsperioden för resultaten de ansökande ansöker med. Däremot känns det underligt att sannolikheten att bli antagen minskade om den ansökande tränade 3 eller fler träningspass per vecka mot 0 träningspass per vecka. Detta skulle kunna förklaras av att folk som tränar mer blir skadade p.g.a. överansträngning. Det skulle även kunna vara så att de som tränar 1-2 gånger per vecka även tränar andra idrotter och att detta gör att de får en bredare träning vilket leder till färre överansträngnings skador. Därför vore det intressant att i andra undersökningar studera om det ökar chansen till antagning att träna andra idrotter. De underliga resultatet skulle även kunna bero på att de som ansöker själva angivit hur många träningar de utfört per vecka. Detta kan ha lett till feltolkningar bland de ansökande, dvs. en del kan ha angett antalet friidrottsträningar medans andra angivit alla träningar de utfört under den angivna perioden. Vi bör även nämna att flertalet av de skattade effekterna hos variabeln Höst_träning2 inte besatt en signifikant effekt vilket kan bidra till det underliga resultatet.

I fortsatta undersökningar vore det intressant att studera om de ansökandes grengruppsval har en inverkan på antagningen. Vilken gren en ansökande huvudsakligen utför kan besitta ett samspel med den ansökandes BMI, då det inom framförallt kula är en fördel att vara lång och tung medan det inom höjdhopp är det en fördel att vara lång och lätt.

Vi kunde välja Modell 2 istället för Modell 1 om vi inte hade eftersträvat en så enkel modell som möjligt. Modell 2 visar som Modell 1 på att oddset för att bli antaget minskar om den ansökande är man eller om den ansökande tränar 3 eller fler pass per vecka i perioden oktober till december, två år innan förväntad gymnasiestart jämfört med 0 pass per vecka i samma period. Oddset för att bli antagen ökar i Modell 2 som i Modell 1 om den ansökande tränar 1-2 pass per vecka i perioden oktober till december, två år innan förväntad gymnasiestart jämfört med 0 pass per vecka i samma period. Däremot existerar i denna modell samspelseffekter samt samspelseffekternas huvudeffekter.

I Modell 2 är oddset att bli antagen om du varit skadad eller sjuk högre än oddset att bli antagen om du inte varit skadad eller sjuk. Detta känns underligt men skulle möjligen kunna bero på att en person som fortsätter satsa efter en längre skada besitter en högre motivation, och därmed även skulle kunna få bättre resultat än en person som inte varit skadad. Eftersom de är unga individer så läker en skada snabbare och därför syns inte negativa effekter som för vuxna individer som fått en skada. Chansen att bli antagen är högre för en ansökande som har en grupptränare än en ansökande som har en personlig tränare. För samspelseffekterna i Modell 2 ses att kvinnor har en lägre sannolikhet att bli antagna än män året 2015 och att ansökanden som har en personlig tränare och är skadade har en högre sannolikhet att bli antagna än en ansökande som är skadad och har en grupptränare.

Vi valde i denna rapport att använda oss utav signifikansnivån 0.1 när vi använde de tre variabelselektions-procedurerna som beskrivs i sektion 3.4 och när vi sedan i skapandet av Modell 2 studerade vardera förklarande variabelns effekt själva. En annorlunda modell skulle kunna tagits fram om vi istället använt oss utav en annan signifikansnivå.

Vår responsvariabel Asökande kan visa på att personer som blivit antagna till friidrottsdelen inte blivit antagna p.g.a. för låga betyg. Detta gör att vi har ansökande som borde stå som antagna men istället är registrerade som nekade vilket kan leda till felaktiga slutsatser om de förklarande variablerna. Vi vet inte heller om det existerar personer som tackat nej till sin plats och därav är registrerade som icke antagna när de istället borde vara registrerade som antagna. Detta leder också till att felaktiga slutsatser om de förklarande variablerna kan dras. Vi har dock fått uppgifter om att antalet personer som blivit nekade p.g.a. betyg är lågt och att även ytterst få personer tackar nej till sina platser.

Vi har antagit att ansökande som angett att de tränar 6 eller mer pass tränar 6 pass per vecka i uträkningarna i Tabell 1. Detta kan leda till att medelvärdena är aningen missvisande. Det är möjligt att andra antaganden som gjordes i sektion 2.1 kan leda till missvisande resultat och om andra antaganden gjorts kunde annorlunda resultat uppnåtts.

Referenser

[1] Richard Öhrvall Lotta Persson. Januaribarn mer framgångsrika. http://www.scb.se/sv_/Hitta-statistik/Artiklar/Januaribarn-mer-framgangsrika/, 11 2016.

[2] Friidrott gymnasiestudier. http://www.friidrott.se/utbildning/plugga/gymnasium/intro.aspx, 11 2016.

[3] Riksfriidrottsgymnasium. https://en.wikipedia.org/wiki/Chronic_electrode_implant, 11 2016.

[4] Nationellt godkänd friidrottsutbildning. http://www.friidrott.se/utbildning/plugga/gymnasium/niufig/intro.aspx, 11 2016.

[5] Sas learning module. missing data in sas. http://www.ats.ucla.edu/stat/sas/modules/missing.htm, 11 2016.

[6] Alan Agresti. *Categorical Data Analysis.* WILEY, third edition edition, 2013.

[7] Rolf Sundberg. *Lineära Statistiska Modeller.* Stockholms Universitet, 2014.

[8] Stanley Lemeshow David W. Hosmer. *Applied Logistic Regression.* WILEY, second edition edition, 2000.

[9] Ida Hed Myrberg. Using logistic regression to predict the risk of depression among parents who have lost a child through suicide.

[10] Så bedömer du din vikt. http://www.1177.se/Stockholm/Tema/Halsa/Livsstil---att-andra-en-vana/Sa-bedomer-du-din-vikt/, 11 2016.

Appendix

Appendix 1

Tabell 9: Tabell över estimerade effekter, standard error samt p-värde för modellen som innehåller alla huvudeffekter.

Variabel		$\hat{\beta}$	SE	p-värde
Kön	Man	0	-	-
	Kvinna	0.274	0.094	0.003
Höst_träning2		-0.174	0.076	0.023
Höst_träning1		0.073	0.075	0.328
Gymnasium	RIG	0	-	-
	NIU	-0.008	0.108	0.939
	Båda	-0.079	0.108	0.502
Födelsemånad		-0.015	0.022	0.500
PT	Gruppträanare	0	-	-
	Personlig tränare	-0.096	0.076	0.208
Skada	Frisk	0	-	-
	Skadad	0.075	0.075	0.312
Antagningsår	2012	0	-	-
	2013	-0.079	0.126	0.532
	2014	-0.056	0.124	0.654
	2015	-0.095	0.121	0.433
Antal_träningsår		-0.034	0.028	0.224
Längd		0.006	0.012	0.586
BMI		0.063	0.036	0.082