



Stockholms
universitet

Skulle du välja samma utbildning igen? - en jämförande studie av alumner i matematik och matematisk statistik

Angelica Witting

Kandidatuppsats 2014:8
Matematisk statistik
Juni 2014

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Skulle du välja samma utbildning igen? - en jämförande studie av alumner i matematik och matematisk statistik

Angelica Witting*

Juni 2014

Sammanfattning

Syftet med denna uppsats var att analysera hur nöjda alumner i matematik och matematisk statistik är med sin utbildning mot bakgrund av deras huvudämne, kön och nuvarande lön, samt mot bakgrund av deras inställning till olika moment i sin utbildning och nuvarande sysselsättning. Analysen genomfördes med hjälp av resultat från en enkät som Naturvetenskapliga fakulteten gjort angående utbildning och arbetsmarknad. I uppsatsen lades störst fokus på att analysera frågan om alumnerna skulle välja samma utbildning igen med hänsyn till deras inriktning på utbildningen, deras kön och deras lön i ett antal olika kumulativa logit modeller. Slutligen jämfördes sambanden mellan denna fråga och ett antal övriga valda frågor från enkäten för respektive inriktning. Resultatet visade att kön och lön har betydelse för huruvida alumnerna skulle välja samma utbildning igen medan inriktning visade sig inte ha någon betydelse alls. De övriga valda frågorna visade inget intressant resultat.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.
E-post: witting.angelica@gmail.com. Handledare: Martin Sköld.

Abstract

The purpose of this paper was to analyze how satisfied alumni in mathematics and mathematical statistics are with their education in view of their main subject, gender and current wage, and also in view of their opinion to different elements in their education and current employment. The analysis was conducted by using the results from a survey that the Faculty of Science made regarding education and labor market. The greatest focus in the paper was analyzing the question whether the alumni would choose the same education again with respect to their specialization of their education, their gender and their wage in a number of cumulative logit models. Finally, the relationships between this question and a number of other selected questions from the survey got compared for each specialization. The results showed that gender and wage are important in whether alumni would choose the same education again while specialization was found to have no significance. The other selected questions showed no interesting results at all.

Förord

Denna uppsats utgör ett självständigt arbete omfattande 15 hp som leder till en kandidatexamen i matematisk statistik på Stockholms universitet. Jag vill främst tacka min handledare Martin Sköld för vägledning och hjälp under arbetets gång. Jag vill även rikta ett tack till Naturvetenskapliga fakulteten på Stockholms universitet för deras bidrag med datamaterial till arbetet.

Innehåll

1	Introduktion	5
2	Material och metod	6
2.1	Enkät	6
2.2	Data	6
2.3	Datahantering och kodning	7
3	Statistiska metoder	8
3.1	Kontingenstabell	8
3.2	Oberoendetest för tvåvägstabeller	9
3.3	Wald-chi två test	10
3.4	Kumulativ logit modell	10
3.5	AIC	12
3.6	Kendalls tau	12
4	Resultat	14
4.1	Statistisk analys av inriktning, kön och lön	16
4.1.1	Kumulativa logit modeller med en förklarandevariabel	17
4.1.2	Kumulativa logit modeller med två förklarandevariabler	18
4.1.3	Kumulativa logit modeller med alla förklarandevariabler	20
4.1.4	Val av modell	23
4.2	Statistisk analys av valda frågor	24
5	Diskussion och slutsats	26
A	Val av frågor	28

1 Introduktion

Syftet med denna uppsats är att främst undersöka om det råder skillnad mellan matematik och matematisk statistik alumner mot bakgrund av hur nöjda de är med deras val av utbildning. Det finns även intresse i att undersöka om alumnernas kön, lön och syn på olika moment i deras utbildning och nuvarande sysselsättning har någon påverkan på hur nöjda de är med sitt val av utbildning. Analysen har gjorts med hjälp av resultat från en enkät som Naturvetenskapliga fakulteten genomfört. Den fråga från enkäten som främst kommer analyseras är "*Skulle du välja samma utbildning igen?*". I första hand ligger intresset i att ta reda på det finns samband mellan alumnernas inställning till denna fråga med hänsyn till den inriktning på utbildningen de har gått, och det kön och den lön de har. Det finns även intresse i att se om det finns något samband mellan denna fråga och ett antal övriga valda frågor från enkäten som berör alumnernas utbildning och nuvarande arbete.

När fördelningen av svar på denna fråga undersöktes visade det sig större delen av de svarande skulle välja samma utbildning igen eller troligen skulle välja samma utbildning igen. Vad är det då som får alumner som läst matematik eller matematisk statistik att vilja välja samma utbildning igen? Antagligen ligger det framförallt i intresset. De flesta som väljer en inriktning med matematik eller matematisk statistik har förmodligen det som intresse och det intresset består nog även efter avslutad utbildning. Om det inte finns något intresse för matematik väljer man sällan någon inriktning med det. En annan teori om varför många av alumnerna i enkäten skulle välja samma utbildning igen är för att de helt enkelt är nöjda med sitt arbete de fått via utbildningen. Dessutom finns det efterfrågan på arbetsmarknaden idag för studenter som läst någon matematisk inriktning vilket medför att många får arbete. Får alumnerna ett bra arbete efter utbildningen som innebär att de får arbeta inom sitt intresseområde är de också nöjda med sitt val av utbildning.

2 Material och metod

2.1 Enkät

Enkäten som uppsatsen bygger på heter *Efter studierna-naturvetare i arbetslivet* och utförs av Naturvetenskapliga fakulteten på Stockholms universitet. Det är en enkät om utbildning och arbetsmarknad för naturvetare som läst vid Stockholms universitet. Denna enkät vars data gav tillhanda skickades ut till de personer som under åren 2006 - 2009 tagit ut examen eller utfört examensarbete vid Naturvetenskapliga fakulteten. Enkäten skickades ut till totalt 1413 personer. Personer vars adress saknades, var felaktig eller som hade adress utomlands uteslöts. Av dessa 1413 personer var det 880 som svarade. Enkäten omfattas av 20 stycken alternativfrågor. I stora drag beskriver enkäten hur synen är på sin utbildning i förhållande till sin nuvarande sysselsättning.

2.2 Data

Det var totalt 39 alumner från utbildningen i matematik som besvarade enkäten. Denna andel utgjordes av 25 stycken män och 14 stycken kvinnor. Från utbildningen i matematisk statistik var det 56 alumner som svarade. Andelen män och kvinnor av dessa var 26 respektive 30 stycken. Enkäten är uppbyggd så att frågorna 1-11 besvarades av samtliga från matematik och matematisk statistik, frågorna 12-20 som bara gällde de som haft något arbete i 6 månader eller längre sedan avslutad utbildning, var det totalt 30 stycken som svarade från matematik och 53 från matematisk statistik. (Av dessa var det 10 kvinnor och 20 män som svarade från matematik, och 29 kvinnor och 24 män som svarade från matematisk statistik). Dessa andelar utgör totalen av de medverkande i dessa frågor men med en del undantag. I flera av dessa frågor saknades svar från vissa medverkande. Det kan bero på att de svarande medvetet valt att inte svara på dessa frågor eller helt enkelt missat att kryssa i svar. *Fråga 3* i enkäten, "Skulle du välja samma utbildning igen?", är som tidigare nämnt den centrala frågan i denna uppsats och har fem svarsalternativ: *Ja, definitivt, Ja, troligen densamma, Nej, troligen en annan, Nej, definitivt en annan* och *Vet ej*. I första hand kommer fråga 3 jämföras mot den inriktningen, det kön och den lön respektive svarande har. *Lön* representerar svar på fråga 20 i enkäten, "*Din nuvarande bruttolön per månad?*" Den frågan består av 9 kategorier: (<18 000), (18 000-22 000), (22 001-26 000), (26 001-30 000), (30 001-34 000), (34 001-38 000), (38 001-42 000), (42 001-46 000) och (>46 000). Datan för de övriga valda frågorna redovisas i Appendix.

2.3 Datahantering och kodning

Resultatet av svaren för matematik- och matematisk statistik alumnerna till enkäten gavs tillhanda i form av en Excel-fil som Naturvetenskapliga fakulteten skickade via mail. Filen bestod av kolumner för varje alumns: kön, examen, examensår, ämne och svar på respektive frågor som enkäten bestod av. På det sättet gavs tillgång till det datamaterial som behövdes för den statistiska analysen. Vid datahanteringen har alumnernas svar sammanställts i tabeller på samma sätt som den tillhanda Excel-filen men med endast de väsentliga kolumnerna. Dessutom har ämne valts att kallas för inriktning istället. Samtlig kodning och statistisk analys har skett med hjälp av dataprogrammet SAS.

3 Statistiska metoder

Enkäten är utformad så att varje fråga består av ett flertal svarsalternativ, indelade i kategorier. Det handlar alltså om *kategoridata*. Många kategoriska variabler består endast av två kategorier. Sådana variabler kallas *binära variabler*. När en kategorisk variabel har fler än två kategorier så kallas det antingen *nominalvariabel* som mäts på en *nominalskala* eller *ordinalvariabel* som mäts på en *ordinalskala*. En nominalvariabel har kategorisering utan någon naturlig ordning medan en ordinalvariabel har kategorisering i form av ordnade kategorier. I övrigt finns variabler som mäts på *intervallskala*, det vill säga *intervallvariabler*. Det innebär att det finns en numerisk skillnad mellan två variabler. För de flesta sådana variabler är det också möjligt att jämföra två värden med deras kvot i vilket fall variabeln även kallas för *kvotvariabel*. Beroende på vilken slags variabel datan i ett datamaterial består av bör man välja lämplig statistisk metod. Frågorna i enkäten är utformade så att de går att rangordna och därför är det ordinaldata som analyseras i denna uppsats. Det kommer i denna del ges en beskrivning av de statistiska metoder som används vid den statistiska analysen i denna uppsats. Först ska två begrepp förklaras som används inom hypotesprövning och som utgöra stor del vid den statistiska analysen.

- Med *signifikansnivån* eller *ferisken*, α , för ett test menas

$$\alpha := P(\text{att förkasta } H_0 | \text{att } H_0 \text{ är sann}).$$

Det vill säga sannolikheten att förkasta nollhypotesen givet att nollhypotesen är uppfylld. I denna uppsats kommer $\alpha=0.05$ användas och därför kommer allting över denna nivå förkastas.

- Ett tests *P-värde* definieras

$$P := P(\text{att få ett minst lika extremt utfall som det observerade} | \text{att } H_0 \text{ är sann}).$$

H_0 förkastas om $P \leq \alpha$, det vill säga om P-värdet är mindre än den önskade felrisken.

3.1 Kontingenstabell

Låt X och Y motsvara två kategoriska variabler, X med I kategorier och Y med J kategorier. En tvåvägs kontingenstabell för X och Y ser ut på följande sätt

	1	2	...	J	Σ
1	n_{11}	n_{12}	\cdots	n_{1J}	n_{1+}
2	n_{21}	n_{22}	\cdots	n_{2J}	n_{2+}
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
I	n_{I1}	n_{I2}	\cdots	n_{IJ}	n_{I+}
Σ	n_{+1}	n_{+2}	\cdots	n_{+J}	n_{++}

där n_{ij} betecknar antalet observationer där X befinner sig i kategori i och Y i kategori j , n_{++} betecknar summan av antalet observationer, n_{i+} och n_{+j} betecknar totala antalet observationer där X befinner sig i i respektive Y i j .

3.2 Oberoendetest för tvåvägstabeller

Antag samma kontingenstabell som ovan. För att testa oberoende mellan variablerna X och Y sätts följande nollhypotes

$$H_0 : \pi_{ij} = \pi_{i+}\pi_{+j}, \quad i = 1, \dots, (I - 1), \quad j = 1, \dots, (J - 1).$$

Där π_{ij} är sannolikheten att X hamnar i i och Y i j , π_{i+} och π_{+j} är sannolikheten att X hamnar i i respektive Y i j .

Under H_0 är de skattade förväntade frekvenserna

$$\begin{aligned} \hat{\mu}_{ij} &= n\hat{\pi}_{i+}\hat{\pi}_{+j} \\ &= n \frac{n_{i+}}{n} \frac{n_{+j}}{n} \\ &= \frac{n_{i+}n_{+j}}{n}. \end{aligned}$$

Då går det antingen att använda *Pearsons chi-två statistika*

$$X^2 = \sum_{i,j} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \stackrel{H_0}{\sim} \chi_{(I-1)(J-1)}^2$$

eller *Likelihood-kvot chi-två statistika*

$$G^2 = 2 \sum_{i,j} n_{ij} \log \frac{n_{ij}}{\hat{\mu}_{ij}} \stackrel{H_0}{\sim} \chi_{(I-1)(J-1)}^2$$

där $(I - 1)(J - 1)$ är antalet frihetsgrader.

3.3 Wald-chi två test

Wald chi-två statistikan används för att testa hypotesen att minst en av förklarandevariablernas koefficienter i en modell inte är lika med noll.

Statistikan beräknas olika beroende på om förklarandevariabeln är kontinuerlig eller kategorisk. Antag fallet där den i :te variabeln är kontinuerlig. Wald chi-två statistikan blir då

$$W_i = \frac{\hat{\beta}_i^2}{\hat{\sigma}_{\hat{\beta}_i}^2}$$

där $\hat{\beta}_i$ är Maximum Likelihood skattningen för i :te variabeln och $\hat{\sigma}_{\hat{\beta}_i}^2$ är variansen för $\hat{\beta}_i$.

Antag fallet där den i :te variabeln är kategorisk. Då blir Wald chi-två statistikan istället

$$W_i = \hat{\beta}_i' \mathbf{C}^{-1} \hat{\beta}_i$$

där $\hat{\beta}_i$ är en vektor av Maximum Likelihood skattningarna associerade med $m-1$ parametrar, och \mathbf{C} är den asymptotiska kovariansmatrisen för $\hat{\beta}_i$. Den asymptotiska fördelningen för Wald statistikan är då chi-två fördelad med frihetsgraderna lika med antalet skattade parametrar.

3.4 Kumulativ logit modell

Kumulativ logit modell är den populäraste modellen för ordinal data. Denna modell använder kumulativa sannolikheter upp till en viss gräns, och gör därigenom ett urval av ordinala kategorier binära till denna gräns. Fördelen med ordinal logit är att den ger en förbättring i statistisk styrka jämfört med om enskilda logit modeller körs för olika utfall. I fall separata logit modeller skulle analyseras skulle det bli att en modell kördes för varje enskild kategori hos responsvariabeln. Dessutom hindrar ordinal logit en från att behöva godtyckligt omvandla en ordinal variabel till en binär variabel.

Låt responsen vara $Y = 1, 2, \dots, J$, där ordningen är given, och de associerade sannolikheterna är $\pi_1, \pi_2, \dots, \pi_J$. Den kumulativa sannolikheten av en respons mindre än eller lika med j är då

$$P(Y \leq j | \mathbf{x}) = \pi_1(\mathbf{x}) + \dots + \pi_j(\mathbf{x}), \quad j = 1, \dots, J.$$

Då definieras de *kumulativa logitarna* som

$$\begin{aligned}\text{logit}[P(Y \leq j|\mathbf{x})] &= \log \frac{P(Y \leq j|\mathbf{x})}{1 - P(Y \leq j|\mathbf{x})} \\ &= \log \frac{\pi_1(\mathbf{x}) + \dots + \pi_j(\mathbf{x})}{\pi_{j+1}(\mathbf{x}) + \dots + \pi_J(\mathbf{x})}, \quad j = 1, \dots, J - 1.\end{aligned}$$

Denna formel beskriver log-oddset för två kumulativa sannolikheter. Det vill säga, den mäter hur rimlig responsen är att vara i kategori j eller lägre mot att vara i en kategori högre än j .

En modell för $\text{logit}P[Y \leq j]$ ensam är en vanlig logistisk modell för en binär respons i vilket kategorierna 1 till j formar ett utfall och kategorierna $j+1$ till J formar den andra. En modell som samtidigt använder alla $(J-1)$ kumulativa logitarna i en sammanförd modell är

$$\text{logit}P[Y \leq j|\mathbf{x}] = \alpha_j + \beta^T \mathbf{x}, \quad j = 1, \dots, J - 1. \quad (1)$$

Eftersom $P(Y \leq j|\mathbf{x})$ ökar i j för ett fixt \mathbf{x} och logiten är en ökande funktion av $P(Y \leq j|\mathbf{x})$ ökar också (α_j) i j . Detta innebär att varje kumulativ logit har sitt egna intercept medan (1) antar samma effekter β för varje logit.

Den kumulativa logit modellen (1) uppfyller

$$\begin{aligned}& \text{logit}[P(Y \leq j|\mathbf{x}_1)] - \text{logit}[P(Y \leq j|\mathbf{x}_2)] \\ &= \log \frac{P(Y \leq j|\mathbf{x}_1)/P(Y > j|\mathbf{x}_1)}{P(Y \leq j|\mathbf{x}_2)/P(Y > j|\mathbf{x}_2)} = \beta^T (\mathbf{x}_1 - \mathbf{x}_2).\end{aligned}$$

En oddskvot av kumulativa sannolikheter kallas en *kumulativ oddskvot*. Oddset av att sätta responsen $\leq j$ då $\mathbf{x} = \mathbf{x}_1$ är $\exp[\beta^T (\mathbf{x}_1 - \mathbf{x}_2)]$ gånger oddset då $\mathbf{x} = \mathbf{x}_2$. Den log kumulativa oddskvoten är proportionell mot längden mellan \mathbf{x}_1 och \mathbf{x}_2 . Samma proportionella konstant gäller för varje logit. På grund av denna egenskap kallas (1) ofta för en *odds proportionell modell*.

I SAS testas automatiskt *Proportionalitet av odds antagandet* vid val av kumulativ logit modell. Det vill säga SAS testar om koefficienterna är samma för varje logit, samtidigt för alla förklarande variabler. Om då den valda modellen inte uppfyller antagandet om odds proportionalitet kan man gå vidare med två alternativ. Den första är *Generaliserad logit modell*. Denna modell antar att responsvariabel är nominal istället för ordinal och kör en full uppsättning parametrar för varje generaliserad logit. Det andra alternativet är *Partiell proportionell odds modell*. Denna modell antar att responsen

fortfarande är ordinal men tillåts anta att odds proportionalitetsantagande uppfylls för vissa förklarande variabler men inte för andra. I denna uppsats kommer för enkelheten ändå vidare analys ske med en odds proportionell modell även om antagandet om odds proportionalitet inte är uppfyllt.

3.5 AIC

Med hjälp av ett mått på en modells anpassning kan jämförelse ske mellan olika modeller. AIC (*Akaike information criterion*) bedömer hur nära de anpassade värdena hos en modell tenderar att vara gentemot modellens sanna medelvärden. AIC räknar ut detta mått i termer av ett särskilt förväntat värde. Detta mått väljer den modell som minimerar

$$\text{AIC} = -2(\text{maximerad log likelihood} - \text{antalet parametrar i modellen}).$$

3.6 Kendalls tau

Kendalls tau eller som det även kallas *Kendalls rank korrelation* är ett icke-parametriskt mått som mäter styrkan av beroende mellan två variabler. Det utförs genom rangordning av data, så att varje variabel enskilt värderas och numreras, 1 för det lägsta värdet, 2 för näst lägsta värdet, och så vidare. Följande formel används för att räkna ut Kendalls tau

$$\hat{\tau} = \frac{C - D}{\frac{1}{2}n(n - 1)}.$$

Där C står för antalet par i datan som är *concordant* och D står för antalet par som är *discordant*. Ett par är concordant om datan som rankas högre på ena variabeln också rankas högre på den andra variabeln. Ett par är discordant om datan som rankas högre på ena variabeln rankas lägre på den andra variabeln. Precis som med andra mått på korrelation så antar Kendalls tau värden i intervallet $-1 \leq \hat{\tau} \leq 1$, en positiv korrelation innebär att rangordningen av båda variablerna ökar tillsammans medan en negativ korrelation innebär att om rangordningen av en variabel ökar så minskar den andra.

Det finns flera olika sorters mått av Kendalls tau. SAS räknar vanligtvis med *Kendalls Tau-b* och därför kommer det vara det korrelationsmättet som analyseras i denna uppsats. Till skillnad från vanliga Kendalls tau tar Kendalls tau-b även hänsyn till oavgjorda par. Antag en $I \times J$ tabell som beskrivits i kontingenstabell-avsnittet och sätt:

t_i = antalet matchade värden i den i :te gruppen för variabel X

u_j = antalet matchade värden i den j :te gruppen variabel Y

$n_0 = n(n - 1)/2$

$$n_1 = \sum_i t_i(t_i - 1)/2$$

$$n_2 = \sum_j u_j(u_j - 1)/2$$

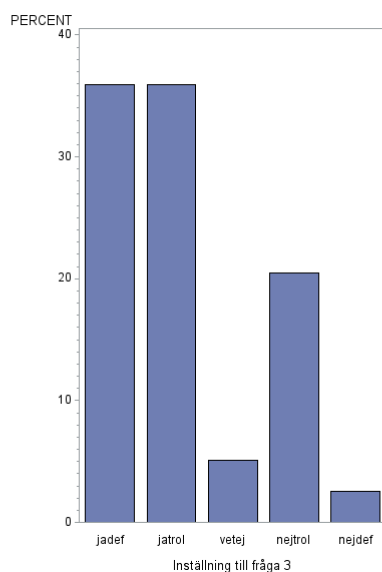
Följande formel används då för att beräkna Kendalls tau-b

$$\hat{\tau}_B = \frac{C - D}{\sqrt{(n_0 - n_1)(n_0 - n_2)}}.$$

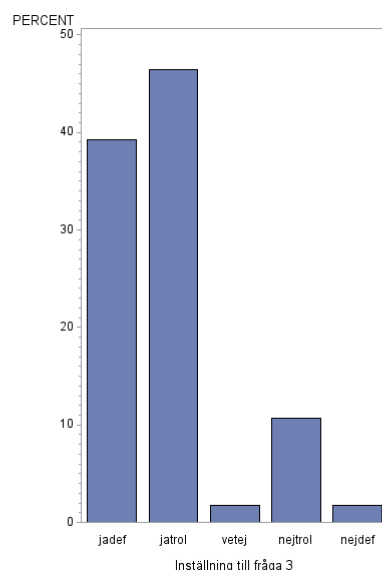
4 Resultat

I denna del redovisas resultatet av den statistiska analysen. I första hand kommer sambandet mellan de svarandes inriktning, kön och lön och inställning till fråga 3 undersökas genom utförandet av kumulativa logit modeller som analyseras i SAS. Det kommer även vara en del i slutet där korrelationen mellan fråga 3 och de övriga valda frågorna för matematik respektive matematisk statistik undersöks. Med hänsyn till de kumulativa logit modellerna kommer först varje förklarandevariabel enskilt ställas mot fråga 3. Det kommer även testas att ställa två förklarandevariabel i taget mot fråga 3 för att se om det finns någon relation. Sedan kommer samtliga förklarandevariabler ställas mot fråga 3 för analys. Det ska testas både med och utan samspel i modellerna för att se hur variablerna reagerar. Slutligen blir valet den modell som har bäst anpassning.

För att undersöka skillnaden hur alumner svarat i respektive inriktning granskas först den procentuella fördelningen av svar för fråga 3 för matematik och matematisk statistik enskilt. Resultatet ses i stapeldiagrammen nedan.



Fördelning av svar på frågan “*Skulle du välja samma utbildning igen?*” för matematik.



Fördelning av svar på frågan “*Skulle du välja samma utbildning igen?*” för matematisk statistik

Det skiljer sig inte nämnbart mellan matematik och matematisk statistik i fördelning. Det behöver därför undersökas på något annat sätt om det skiljer sig mellan inriktningarna och fråga 3. Det kan göras genom utförandet av chi-två test i SAS. Nedan redovisas Pearsons- och Likelihood-kvot chi-två statistikor respektive P-värde för dessa, för test av oberoende mellan fråga 3 och inriktning. För att se om det även skiljer sig mellan män och kvinnor med hänsyn till hur nöjda de är med sin utbildning redovisas också samma tester fast mellan kön och fråga 3.

Statistics for Table of install3 by inrikt			
Statistic	DF	Value	Prob
Chi-Square	4	3.0525	0.5491
Likelihood Ratio Chi-Square	4	3.0167	0.5550

Statistics for Table of install3 by kon			
Statistic	DF	Value	Prob
Chi-Square	4	5.6690	0.2253
Likelihood Ratio Chi-Square	4	6.4512	0.1679

Testen för inriktning mot fråga 3 är båda osignifikanta. Testen för kön mot fråga 3 är också osignifikanta, men har dock betydligt lägre P-värden än de för inriktning. Sammantaget kan det från dessa tester konstateras att det inte skiljer sig mellan inriktningarna respektive mellan könen och fråga 3.

4.1 Statistisk analys av inriktning, kön och lön

För att se om det finns någon relation mellan alumnernas inställning till hur nöjda de är sitt val av utbildning och deras respektive inriktning, kön och lön, görs en analys med en kumulativ logit modell. För detta krävs en responsvariabel som är ordinal. Svartalternativen till fråga 3 är ordinala på det sättet att de går att rangordna. Det sätts: Y = Fråga 3 (med kategorier 0=Nej, definitivt en annan, 1=Nej, troligen en annan, 2=Vet ej, 3=Ja, troligen densamma och 4=Ja, definitivt), x_1 = Inriktning (med kategorier 0=matematik och 1=matematisk statistik), x_2 =Kön (med kategorier 0=kvinna och 1=man) och x_3 =lön, som är en kontinuerlig variabel.

Lön består som tidigare nämnt av 9 kategorier. Dessa kategorier omfattas av stora intervall och dessutom skiljer det sig stort mellan kategorierna. Det gör att lön inte blir lika lätt att tolka resultatet av, framförallt som kontinuerlig variabel. Det som kan göras då för att underlätta analysen är att *standardisera* lön. Det sker genom att först ta mittpunkten för varje lönekategori, på det sättet fås ett värde för lön i varje kategori istället för ett intervall. Följande värden skapas då för lön: 20 000, 24 000, 28 000, 32 000, 36 000, 40 000, 44 000 och 46 000. Följande formel ger då en standardisering av variabeln lön.

$$zlön = \frac{(lön - \mu)}{\sigma}$$

där μ är väntevärdet och σ är standardavvikelsen för lön, och $zlön$ är den nya variabeln för lön som skapas. Denna typ av standardisering utförs med hjälp av SAS vilket resulterar i att lön blir på en lämpligare form för analysen.

Först undersöks fördelningen av svar för fråga 3 mot inriktning respektive kön. Resultat ses i frekvenstabellerna nedan. Eftersom vissa celler är tomma eller har extremt små värden kan resultatet påverkas och modellen kan bli svår att köra. Det är en viktig åtanke för vidare analys.

Fråga 3	Inriktning		
	Matem.	Matstat.	Totalt
Nej def. inte	1	1	2
Nej trol. inte	6	6	12
Vet ej	2	1	3
Ja trol.	11	24	35
Ja def.	10	20	30
Totalt	30	52	82

Fråga 3	Kön		
	Kvinna	Man	Totalt
Nej def. inte	2	0	2
Nej trol. inte	8	4	12
Vet ej	2	1	3
Ja trol.	14	21	35
Ja def.	12	18	30
Totalt	38	44	82

För att lätt kunna särskilja de olika modellerna i följande analys sätts: I=inriktning, K=kön och L=lön.

4.1.1 Kumulativa logit modeller med en förklarandevariabel

Först ligger intresset i att undersöka om det finns någon relation om bara en av förklarandevariablerna i taget inkluderas i en kumulativ logit modell med fråga 3 som responsvariabel. Nedan redovisas resultatet av parameterskattningarna från SAS för respektive modell.

Modell (I): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1 x_1$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x1	1	0.4549	0.4276	1.1320	0.2873

Modell (K): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1 x_2$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x2	1	0.7373	0.4190	3.0969	0.0784

Modell (L): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1 x_3$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x3	1	0.3625	0.2111	2.9475	0.0860

Det som redovisas i samtliga tabeller är parameter, frihetsgrader, skattning (koefficient), standardavvikelse, Wald chi-två statistika och P-värde (för sannolikheten att observera ett större värde av Wald chi-två statistikan under nollhypotesen). Det kan tilläggas att referensparametern är satt till den lägsta kategorin för inriktning och kön, det vill säga "0". Frihetsgraderna blir 1 för samtliga skattningar eftersom det bara är en parameterskattning för varje variabel. Det kan konstateras vid analys av modellerna ovan att det inte är någon av förklarandevariablerna inriktning, kön och lön som är signifikant för sin respektive modell.

4.1.2 Kumulativa logit modeller med två förklarandevariabler

Sedan är intresset att undersöka vad som händer om inriktning, kön och lön inkluderas två i taget i en kumulativ logit modell. Först undersöks utan samspel mellan variablerna och sedan med samspel. För respektive modell erhålles följande parameterskattningar från SAS.

Modell (I, K): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1 x_1 + \beta_2 x_2$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x1	1	0.6021	0.4406	1.8673	0.1718
x2	1	0.8368	0.4308	3.7742	0.0520

Modell (I, L) $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1 x_1 + \beta_2 x_3$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x1	1	0.2419	0.4528	0.2854	0.5932
x3	1	0.3254	0.2229	2.1313	0.1443

Modell (K, L): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1x_2 + \beta_2x_3$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x2	1	0.7132	0.4214	2.8646	0.0905
x3	1	0.3502	0.2128	2.7082	0.0998

Modell (IK): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1x_1 + \beta_2x_2 + \beta_3x_1x_2$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x1	1	1.4031	0.6963	4.0609	0.0439
x2	1	1.7240	0.7402	5.4253	0.0198
x1*x2	1	-1.3281	0.8993	2.1809	0.1397

Modell (IL): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1x_1 + \beta_2x_3 + \beta_3x_1x_3$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x1	1	0.3223	0.4629	0.4848	0.4863
x3	1	0.0986	0.3773	0.0683	0.7939
x1*x3	1	0.3211	0.4660	0.4746	0.4909

Modell (KL): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1x_2 + \beta_2x_3 + \beta_3x_2x_3$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x2	1	0.6831	0.4274	2.5550	0.1099
x3	1	1.0019	0.3610	7.7013	0.0055
x2*x3	1	-1.0201	0.4494	5.1514	0.0232

Det ses från modellerna ovan att alla modeller utan samspel är osignifikanta för samtliga sina förklarandevariabler. Av samspelsmodellerna är modell (IL) osignifikant för alla parametrar medan modell (IK) och modell (KL) har parametrar som visar signifikans. I modell (IK) är inriktning och kön signifikanta var för sig men samspelet inriktning*kön är osignifikant. Koefficienterna för inriktning och kön säger att den kumulativa sannolikheten (med början vid det lägsta svarsalternativet för fråga 3) är högre för matematisk statistik än för matematik respektive högre för män än för kvinnor. I modell (KL) är lön och samspelet kön*lön signifikanta. Löns skattning

säger att den kumulativa sannolikheten ökar då lön ökar. Samspelsfaktorn kön*lön säger att för män jämfört med kvinnor, minskar den kumulativa sannolikheten då lön ökar.

4.1.3 Kumulativa logit modeller med alla förklarandevariabler

I denna del är det meningen att analysera vad som händer då inriktning, kön och lön inkluderas tillsammans i en kumulativ logit modell. Först undersöka effekterna av en modell utan samspel och i resterande modeller undersöks effekterna mellan olika samspel. Nedan redovisas parameterskattningarna från SAS för respektive modell.

Modell (I, K, L): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1x_1 + \beta_2x_2 + \beta_3x_3$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x1	1	0.4034	0.4699	0.7368	0.3907
x2	1	0.7832	0.4352	3.2384	0.0719
x3	1	0.2869	0.2264	1.6058	0.2051

Modell (IK, L): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_1x_2$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x1	1	1.1493	0.7376	2.4278	0.1192
x2	1	1.5972	0.7550	4.4752	0.0344
x3	1	0.2494	0.2287	1.1896	0.2754
x1*x2	1	-1.2048	0.9114	1.7476	0.1862

Modell (KL, I): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_2x_3$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x1	1	0.2432	0.4787	0.2581	0.6115
x2	1	0.7286	0.4415	2.7234	0.0989
x3	1	0.9413	0.3827	6.0488	0.0139
x2*x3	1	-0.9863	0.4558	4.6836	0.0305

Modell (IL, K): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_1x_3$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x1	1	0.5456	0.4855	1.2630	0.2611
x2	1	0.8544	0.4443	3.6980	0.0545
x3	1	-0.0480	0.3876	0.0154	0.9014
x1*x3	1	0.4740	0.4778	0.9844	0.3211

Modell (IK, KL): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_1x_2 + \beta_5x_2x_3$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x1	1	0.5270	0.8215	0.4114	0.5212
x2	1	1.0286	0.8232	1.5611	0.2115
x3	1	0.8695	0.4219	4.2473	0.0393
x1*x2	1	-0.4319	1.0103	0.1828	0.6690
x2*x3	1	-0.8982	0.5039	3.1769	0.0747

Modell (IK, IL): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_1x_2 + \beta_5x_1x_3$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x1	1	1.5314	0.7940	3.7195	0.0538
x2	1	1.8936	0.8009	5.5906	0.0181
x3	1	-0.2273	0.4100	0.3074	0.5793
x1*x2	1	-1.5008	0.9527	2.4814	0.1152
x1*x3	1	0.6625	0.4970	1.7764	0.1826

Modell (IL, KL): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_1x_3 + \beta_5x_2x_3$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x1	1	0.4086	0.4925	0.6884	0.4067
x2	1	0.8089	0.4502	3.2282	0.0724
x3	1	0.5791	0.4806	1.4518	0.2282
x1*x3	1	0.5634	0.4806	1.3739	0.2411
x2*x3	1	-1.0400	0.4623	5.0609	0.0245

Modell (IK, KL, IL): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_1x_2 + \beta_5x_2x_3 + \beta_6x_1x_3$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x1	1	0.9346	0.8775	1.1346	0.2868
x2	1	1.3510	0.8705	2.4085	0.1207
x3	1	0.3932	0.5524	0.5066	0.4766
x1*x2	1	-0.7648	1.0508	0.5298	0.4667
x2*x3	1	-0.8964	0.5077	3.1178	0.0774
x1*x3	1	0.6576	0.5019	1.7168	0.1901

Modell (IKL): $\text{logit}[P(Y \leq j|\mathbf{x})] = \alpha_j + \beta_1x_1 + \beta_2x_2 + \beta_3x_3 + \beta_4x_1x_2 + \beta_5x_2x_3 + \beta_6x_1x_3 + \beta_7x_1x_2x_3$

Analysis of Maximum Likelihood Estimates					
Param.	DF	Estimate	S.E.	WaldChiSq	Pr > ChiSq
x1	1	0.8812	1.0001	0.7763	0.3783
x2	1	1.2873	1.0316	1.5572	0.2121
x3	1	0.4588	0.7961	0.3321	0.5644
x1*x2	1	-0.7076	1.1592	0.3726	0.5416
x2*x3	1	-0.9903	0.9353	1.1210	0.2897
x1*x3	1	0.5682	0.9347	0.3696	0.5432
x1*x2*x3	1	0.1285	1.1031	0.0136	0.9072

Det går lätt att se i modell (IK, L) och modell (IK, IL) att kön är den enda parameter som är signifikant med en ungefär lika stor koefficient som i modell (IK). I modell (KL, I) är variabeln lön och samspelsfaktorn kön*lön signifikanta. I modell (IK, KL) är lön den enda signifikanta parametern. I modell (IL, KL) är den enda signifikanta parametern kön*lön. Koefficienterna för

samtliga dessa parametrar har ungefär samma värde och därför fås samma resultat som för motsvarande parametrar i modell (KL). Det vill säga effekterna för lön respektive kön*lön är ungefär lika stora i modell (KL), modell (KL, I), modell (IK, KL) och modell (IL, KL). Resterande modeller i denna del är osignifikanta för samtliga sina förklarandevariabler/samspelsfaktorer.

4.1.4 Val av modell

Slutligen är intresset att analysera vilken av alla dessa modeller som är bäst anpassad och på så sätt välja den bästa modellen. Det kan göras genom att jämföra varje modells anpassningsmått. I nedanstående tabell redovisas respektive modell med sitt AIC-värde.

Modell	AIC
(I)	209.639
(K)	207.640
(L)	207.693
(I, K)	207.749
(I, L)	209.413
(K, L)	206.799
(IK)	207.527
(KL)	203.232
(IL)	210.966
(I, K, L)	208.050
(IK, L)	208.262
(KL, I)	204.971
(IL, K)	209.096
(IK, IL)	208.497
(IK, KL)	206.785
(IL, KL)	206.785
(IK, KL, IL)	207.068
(IKL)	209.054

Gemensamt för de modellerna med högst AIC, det vill säga de modeller med sämst anpassning, är att de innehåller något samspel med förklarandevariabeln inriktning. Det är två modeller som sticker ut med AIC-värden betydligt lägre än de andra, det är modell (KL) och modell (KL, I). Båda dessa innehåller det enda två-faktorsamspelet kön*lön. Den modellen med absolut lägsta AIC av dessa två är modell (KL).

4.2 Statistisk analys av valda frågor

I denna del av resultatet är intresset att se om alumnerna som besvarat enkäten skulle välja samma utbildning igen beroende på vad de svarat på övriga frågor i enkäten. Tanken är alltså att med hjälp av ett lämpligt mått undersöka om det finns något samband mellan fråga 3 och övriga valda frågor. Kendalls tau är ett bra korrelationsmått och lämplig för kategori-data. Det finns flera mått av Kendalls tau och eftersom SAS räknar med Kendalls tau-b är det som redovisas i tabellen nedan. I tabellen redovisas korrelationen för fråga 3 mot respektive övriga valda frågor för matematik (Matem.) och matematisk statistik (Mat. stat.).

Fråga	$\hat{\tau}$ för Matem.	$\hat{\tau}$ för Mat. stat.
4	0.38	0.15
5.6	0.28	0.003
5.7	-0.03	0.24
5.9	-0.10	0.08
5.10	0.22	0.07
5.13	-0.08	0.15
5.14	-0.22	0.05
5.18	-0.04	0.14
14.1	0.20	0.08
14.2	0.10	0.07
14.5	0.13	0.07
14.6	0.10	0.01
14.7	0.18	-0.01
14.8	0.04	0.14
14.9	0.02	0.12
14.10	-0.02	0.08
14.11	-0.16	0.12
14.13	-0.12	0.14
14.14	-0.06	0.16
14.15	-0.04	-0.01
14.17	-0.14	0.12
14.18	-0.03	-0.15
15.1	0.24	0.13
15.2	-0.13	-0.06
16	0.16	0.14

Det är inga av korrelationsvärdena som visar något starkt samband mellan fråga 3 och respektive övriga frågor. Både matematik och matematisk statistik har värden som ligger relativt nära noll. Det som läggs märke till är att det är betydligt fler negativa korrelationer hos matematik än hos matematisk statistik. Visserligen är några av dessa så pass små att de är i stort sett

noll. Samtidigt har matematik fler högre korrelationsvärden än matematisk statistik. Det innebär att det är större variation mellan korrelationsvärdena för matematik än för matematisk statistik. Ett värde som avviker är korrelationsvärdet för fråga 4 för matematik. Den har ett värde på 0.38 och är betydligt högre än övriga. Det kan dock vara en ren tillfällighet.

5 Diskussion och slutsats

Resultatet från analysen med de kumulativa logit modellerna visar dels att de modeller utan samspel inte har signifikans för någon förklarande variabel. Det visar även att inriktning är osignifikant i alla modeller utom den modellen med samspelet mellan inriktning och kön. Slutligen visar det att kön och lön är signifikanta i flera av modellerna med samspel. Köns signifikans stärks då inriktning inkluderas i modellen, och löns signifikans stärks då kön inkluderas i modellen. Kön och lön visar framförallt på signifikans i samspel tillsammans. Det stärks av att den bästa modellen är den med faktorsamspelet mellan kön och lön. Resultatet för de valda frågorna visar ingen av dem något särskilt samband och det går inte heller att se någon relation mellan inriktningarna.

Slutsatsen som kan dras från resultatet är att matematik- och matematisk statistik alumner är i stort sett lika nöjda med deras val av utbildning. Men det skiljer sig till viss del hur nöjda alumnerna är med sitt val av utbildning mot bakgrund av vilket kön och hur hög lön de har. På så sätt att män jämfört med kvinnor respektive lön som helhet förväntas ha en positiv effekt på hur nöjda alumnerna är med sitt val av utbildning. Även på så sätt att männen lön jämfört med kvinnornas lön förväntas ha en negativ effekt på hur nöjda alumnerna är med sitt val av utbildning. Det kan även dras slutsatsen att det inte finns något direkt samband mellan hur nöjda alumnerna är med sitt val av utbildning och hur de ser på olika moment i sin utbildning respektive sitt nuvarande arbete.

Vad beror då resultatet på? Att alumnerna som läst matematik respektive matematisk statistik är lika nöjda med sitt val av utbildning är inte orimligt med tanke på det som nämndes i introduktionen, att väljer en person ett matematiskt ämne är det ofta på grund av intresset och att möjligheten att få arbeta med det finns. Även att högre lön förväntas ge nöjdare alumner till sitt val av utbildning är inget att ifrågasätta. Resultatet som visar att män förväntas vara mer nöjda med sitt val av utbildning än kvinnor samt resultatet som visar att män jämfört med kvinnor förväntas vara mindre nöjda med sitt val av utbildning för en högre lön kan förklaras av att de var fler svarande för män än för kvinnor. Männen svar kan då bli av större påverkan, men det kan även vara så att det faktiskt skiljer sig mellan könen. Det kan också förklaras av den låga svarsfrekvensen hos de lägsta svarsalternativen i fråga 3. Det är möjligt att om fler personer hade besvarat enkäten, att dessa svarsalternativ hade fått en högre svarsfrekvens och en slutsats hade kunnat dras med större säkerhet. Men det är inte säkert att de hade utgjort någon skillnad. Eftersom större delen av alumnerna som besvarat enkäten är nöjda med sitt val av utbildning finns det ingenting som säger att inte andra som läst matematik eller matematisk statistik är det också.

Referenser

- [1] Agresti, A. (2013). *Categorical Data Analysis*. 3:e uppl. New Jersey: John Wiley & Sons, Inc.
- [2] Alm, S.E., Britton, T. (2008). *Stokastik: Sannolikhets teori och statistikteori med tillämpningar*. 1:a uppl. Stockholm: Liber AB.
- [3] Ohlsson, E. (2005). *Kort handledning i SAS*. Stockholms universitet: Matematiska institutionen.
- [4] Nordström, P. (2012). *Efter studierna*. Hämtad 2014-05-27, från <http://www.science.su.se/utbildning/arbete-karri%C3%A4r/efter-studierna-1.35172>.
- [5] IBM (2011). *Wald Statistic (logistic regression algorithms)*. Hämtad 2014-05-27, från http://pic.dhe.ibm.com/infocenter/spssstat/v20r0m0/index.jsp?topic=%2Fcom.ibm.spss.statistics.help%2Falg_logistic_stepwise_wald.htm
- [6] The Pennsylvania State University (2014). *The Proportional-Odds Cumulative Logit Model*. Hämtad 2014-05-27, från <https://onlinecourses.science.psu.edu/stat504/node/176>
- [7] Wikipedia (2014). *Kendall tau rank correlation coefficient*. Hämtad 2014-05-27, från http://en.wikipedia.org/wiki/Kendall_tau_rank_correlation_coefficient
- [8] Idre UCLA (2014). *SAS FAQ How do I standardize variables in SAS?*. 2014-05-27, från <http://www.ats.ucla.edu/stat/sas/faq/standard.htm>

A Val av frågor

I denna del måste det bestämmas vilka frågor från enkäten som ska jämföras med fråga 3. När övriga frågor valdes togs det hänsyn till vilka som var lämpliga för det ändamålet. I första hand är frågor angående erhållen kunskap intressantast att jämföra mot bakgrund av om den svarande skulle välja samma utbildning igen. Frågor med öppna svar valdes bort eftersom svaren skiljer sig så mycket mellan personer och blir därför svåra att analysera. Fråga 13 som är en rangordningsfråga valdes att inte heller ha med. Efter övervägande valdes alltså frågorna 4, 5, 14, 15 och 16 att jämföra fråga 3 med. Fråga 4 består av två mindre frågor (delfrågor), där valdes endast den andra delfrågan. Fråga 5 och 14 är uppbyggda på samma sätt med 19 likadana delfrågor. Det som skiljer dem åt är att fråga 5 tar hänsyn till den svarandes utbildning medan fråga 14 tar hänsyn till det nuvarande arbetet. Frågor om fältstudier och laborativa tekniker som ingick i fråga 5 och 14 valdes bort eftersom både fältstudier och laborativa tekniker är en tolkningsfråga. Fältstudier är inget som tillämpas i varken matematik eller matematisk statistik och kan dessutom feltolkas. Laborativa tekniker kan både tolkas som exempelvis kemiska laborationer eller datalaborationer. Även frågan om att bearbeta statistiskt material togs bort eftersom statistiskt material används övervägande mer i matematisk statistik än i matematik. Det känns därför inte väsentligt att jämföra. Delfrågorna i fråga 5 där fler än 5 alumner svarat *Ej relevant* valdes att inte ta med. De flesta av delfrågorna i fråga 14 valdes att ta med, och de som valdes bort förutom de som nämnts ovan, var de som verkade mindre intressanta för undersökningen. Fråga 15 bestod av 4 delfrågor, där valdes bara de första två delfrågorna med. De valda frågorna med respektive svarsalternativ blev sammantaget:

- **Fråga 4.** *Hur bedömer du kvalitén på din utbildning inom naturvetenskapliga fakulteten i Stockholm med tanke på arbets- och näringslivsanknytning?*

Svarsalternativ: *Mycket bristfällig, Delvis bristfällig, Bra, Mycket bra.*

- **Fråga 5.** *Hur bedömer du den kompetens du erhållit i din utbildning vid Naturvetenskapliga fakulteten vid Stockholms universitet med tanke på färdigheter att...*

6. *hitta och värdera ny kunskap?*
7. *kritiskt tolka och diskutera problemställningar?*
9. *självständigt identifiera och lösa problem?*
10. *följa med i kunskapsutvecklingen?*
13. *göra muntliga presentationer?*
14. *göra skriftliga presentationer?*
18. *att arbeta i grupp?*

Svarsalternativ: *Mycket bristfällig, Delvis bristfällig, bra, Mycket bra, Ej relevant.*

- **Fråga 14.** *I vilken utsträckning har du i ditt arbete under den senaste månaden haft arbetsuppgifter som ställt krav...*

1. *på kunskaper i ditt huvudämne?*
2. *på kunskaper i ämnets vetenskapliga grund?*
5. *på kunskaper i aktuella forskningsfrågor?*
6. *färdigheter att hitta och värdera ny kunskap?*
7. *färdigheter att kritiskt tolka och diskutera problemställningar?*
8. *färdigheter att argumentera och övertyga?*
9. *färdigheter att självständigt identifiera och lösa problem?*
10. *färdigheter att följa med i kunskapsutveckling?*
11. *färdigheter att undervisa?*
13. *färdigheter att göra muntliga presentationer?*
14. *färdigheter att göra skriftliga presentationer?*
15. *färdigheter att göra presentationer på engelska?*
17. *färdigheter att förklara för ickespecialister/lekmän?*
18. *färdigheter att arbeta i grupp?*

Svarsalternativ: *Inte alls, Till viss del, Till stor del, Helt.*

- **Fråga 15.** *Hur viktiga för dina nuvarande arbetsuppgifter är de kunskaper och färdigheter som du förvärvat genom...*

1. *utbildning vid naturvetenskapliga fakulteten, SU?*
2. *arbetserfarenhet?*

Svarsalternativ: *Helt oviktiga, Delvis viktiga, Viktiga, Mycket viktiga.*

- **Fråga 16.** *Har din utbildning relevans för ditt nuvarande arbete?*

Svarsalternativ: *Ingen relevans, Viss relevans, Relevans, Hög relevans.*