Stockholms universitet

Statistiska modeller för naturvetares månadslön

Viktor Johansson

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2019:19
Matematisk statistik
Juni 2019

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Matematisk statistik
Stockholms universitet
Kandidatuppsats **2019:19**
http://www.math.su.se/matstat

Statistiska modeller för naturvetares månadslön

Viktor Johansson*

Juni 2019

Sammanfattning

Denna studie har för syfte att finna en statistisk modell för månadslönen för personer som nyligen antingen tagit examen från eller gjort examensarbete vid Naturvetenskapliga fakulteten vid Stockholms universitet. De förklarande variablerna i denna modell är personens kön, studiernas ämne, examensnivå, tid förlöpt sedan senast uttagna examen, samt huruvida personen utförde sitt examensarbete hos en extern uppdragsgivare.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige.
E-post: viktor.johanson@gmail.com. Handledare: Tom Britton, Martin Sköld.

Förord

Detta arbete utgör ett examensarbete i matematisk statistik om 15 högskolepoäng på kandidatnivå vid Matematiska institutionen på Stockholms universitet.

Jag vill tacka docent Cecilia Kullberg på Zoologiska institutionen, Stockholms universitet, som gett mig materialet bakom Efter studierna VI.

Till sist, ett stort tack till mina handledare Martin Sköld och Tom Britton.

Innehåll

1 Inledning **5**

2 Tidigare studier **5**
2.1 Card (1999) 5
2.2 Witting (2014) 5

3 Beskrivning av data **6**
3.1 Bortfall 7

4 Teori **7**
4.1 Multipel linjär regression 7
4.2 Kumulativ logistisk regression 7
4.3 Tak- och golveffekter 9

5 Genomförande och resultat **9**
5.0.1 Variabelselektion 9
5.1 Multipel linjär regression 10
5.1.1 Känslighetsanalys 14
5.1.2 Modelldiagnostik 14
5.2 Kumulativ logistisk regression 16
5.2.1 Modelldiagnostik 18
5.3 Jämförelse mellan modellerna 18

6 Slutsats **18**
6.0.1 Examensarbete 22
6.0.2 Examensår 22
6.0.3 Högsta examen 23
6.0.4 Ämne 23

1 Inledning

Med jämna mellanrum producerar Naturvetenskaplig fakultet vid Stockholms universitet en rapport vid namn "Efter studierna" baserad på en enkät för fakultetens alumni. Rapporten "Efter studierna" innehåller bl.a. information om hur de före detta studenterna ser på sin utbildning och vad de har gjort efter studierna.

I detta examensarbete undersöks huruvida det är möjligt att dra ytterligare slutsatser från enkätsvaren än de som gjorts i rapporten från Naturvetenskaplig fakultet. I synnerhet presenteras i föreliggande arbete statistiska modeller med bruttolön som responsvariabel och ett antal förklarande variabler: studieämne, examensnivå, uppdragsgivare för examensarbete, kön samt hur lång tid som förlöpt sedan senaste examen. Modeller av två olika typer kommer att presenteras och jämföras ur lämplighetssynpunkt: kumulativ logistisk regression och multipel linjär regression.

2 Tidigare studier

Åtskilliga studier har gjorts för att finna samband mellan inkomst och utbildningsnivå och kön. Av dessa betraktar en del livslön men det finns också gott om studier som fokuserar på månadslön eller årslön vid en given tidpunkt.

2.1 Card (1999)

En omfattande artikel av David Card vid University of California at Berkeley går under namnet "The Causal Effect of Education on Earnings" (1999). Den sammanfattar bl.a. resultat från ett antal tidigare studier, däribland studier av enäggstvillingar där tvillingarna läst olika utbildningar. Cards studie är mycket omfattande och grundlig, men beaktar däremot inte vilket ämne testsubjekten studerat. Card betraktar inkomst som en kontinuerlig variabel och av allt att döma är hans studie baserad på fall där man haft data över exakt inkomst, vilket inte är fallet i vår studie. Således har det inte funnits något behov för Card att tillämpa någon form av logistisk regression.

2.2 Witting (2014)

Vid Stockholms universitets matematiska institution har tidigare gjorts en studie (även den en kandidatuppsats) baserad på data från "Efter studierna". Wittings arbete tar dock sikte på en helt annan enkätfråga, nämligen "Skulle du välja samma utbildning igen?". Denna fråga fanns med i enkäten till "Efter studierna V" men togs bort inför "Efter studierna VI".

3 Beskrivning av data

Data kommer från undersökningen som ledde till rapporten "Efter studierna". Materialet erhölls i form av ett Excel-dokument med en rad för varje svarande och en kolumn för varje enkätfrågesvar. Många svar var fritt formulerade i text av de svarande, men de flesta frågor var flervalsfrågor. Rapporten ni nu läser beaktar endast svar på flervalsfrågor, dvs. kategoridata.

Denna studie begränsar sig till deltagarnas svar på sex artiklar: responsvariabeln månadslön och de förklarande variablerna kön, typ av examen (avser den högsta examen som uttagits), huvudämne på examen (avser den högsta examen som uttagits), examensår för senaste examen, samt var examens-arbetet utförts (på högskola/universitet eller med extern uppdragsgivare).

Responsvariabeln är här månadslön brutto, angiven på en skala av nio intervall, där de sju intervallen i mitten omfattar ett spann på 4000 SEK vardera medan de yttersta intervallen är öppna. Den underliggande variabeln kan betraktas som kontinuerlig, eftersom månadslönen i princip kan anta vilket icke-negativt värde som helst. Uppgiften om månadslön har insamlats i form av svar på frågan "Din nuvarande bruttolön per månad?" med svarsalternativen

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
>46 000

där de svarande fått kryssa i den ruta som motsvarar aktuell månadslön. Enkätsvaren från 2015 utgörs av svar från 887 personer. Efter att ha tagit bort de svar som saknar uppgift för någon av de här studerade frågorna, återstod 425 svar.

Det vore olämpligt om någon av de förklarande variablerna i hög grad skulle vara kraftigt korrelerad med andra förklarande variabler, dvs. om värdet hos en variabel kan förutsägas i hög grad utifrån värdet på de andra. Det vore ett tecken på att någon förklarande variabel är överflödig. I synnerhet misstänktes att alumni med masterexamen skulle vara markant överrepresenterade hos gruppen som gjort examensarbete hos extern uppdragsgivare. Denna farhåga visade sig omotiverad. Bland alumni med fil kand som högsta examen år 2015 hade 15.1 % gjort sitt examensarbete hos extern uppdragsgivare, medan motsvarande andel för alumni med master som högsta examen var 16.5 %.

3.1 Bortfall

Inför rapporten "Efter studierna VI" sände man ut brev till 2 034 personer. Av dessa kom 49 brev i retur. Av de återstående 1 985 personerna svarade 890 på enkäten. De 1095 personer som fått brevet men inte svarat på enkäten kontaktades två månader efter att enkäten stängde och ombads svara på en kraftigt förkortad variant av enkäten. På denna svarade 121 personer, vars svar jämfördes med svaren från personerna som svarat på den ursprungliga enkäten. Av dessa 121 uppgav 39 % en månadslön på upp till 26 000 kr (att jämföra med 30 % av de ursprungliga svarande), 53 % en månadslön mellan 26 001-38 000 kr (att jämföra med 60 %) samt 9 % en månadslön över 38 000 kr (10 % för övriga svarande). Ett χ^2-test som presenteras i "Efter studierna VI" visade en teststatistika på 3.4 på 2 frihetsgrader, vilket motsvarar p-värdet 0.2, dvs. ingen signifikant skillnad mellan de ursprungliga svarande och de 121 personerna i bortfallsanalysgruppen. Huruvida denna lilla grupp på 121 personer är representativ för de 53.8 % av de kontaktade som inte svarade på den ursprungliga enkäten kan dock inte avgöras utifrån befintliga uppgifter.

4 Teori

4.1 Multipel linjär regression

Formen för en multipel linjär regressionsmodell är

$$Y_i = \mu_i + \varepsilon_i = \alpha + \boldsymbol{\beta}^T \boldsymbol{X}_i + \varepsilon_i, E[\varepsilon_i] = 0, Var(\varepsilon_i) = \sigma^2, 1 \leq i \leq n,$$

där n betecknar stickprovsstorleken. Vidare antas alla ε_i statistiskt oberoende av varandra och av $\boldsymbol{X}$.

4.2 Kumulativ logistisk regression

Denna modelltyp behandlar responsvariabeln som en variabel på ordinalskala, och involverar inga antaganden om "avstånden" mellan olika nivåer på responsvariabeln. De förutsätts alltså inte vara lika stora.

För varje subjekt i den data som fanns tillgänglig för denna studie var respons-variabelns värde (månadslön) något av nio ordnade, icke-överlappande intervall. Om man vill bygga en statistisk modell för detta utan att göra antaganden om hur mycket personerna i den "lägsta" (<18 000 SEK) respektive den "högsta" (>46 000 SEK) kategorin tjänar, kan man betrakta responsvariabeln som kategorisk.

Låt Y vara responsvariabeln med J kategorier. Vi definierar

$$pi_j(\boldsymbol{x}) := P(Y = j|\boldsymbol{x})$$

dvs. $pi_j(\boldsymbol{x})$ är sannolikheten att responsvariabeln antar kategori j givet att de förklarande variablerna har värde $\boldsymbol{x}$.

Då gäller att den kumulativa sannolikheten är (Agresti (2013), sid. 301)

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

Den logistiska funktionen, betecknad *logit*, definieras genom

$$logit\,[x] := log\frac{x}{1-x}.$$

(Här och i resten av rapporten betecknar *log* den naturliga logaritmen.)

Följaktligen gäller att (Agresti (2013) sid. 301)

$$logit\,[P(Y \leq j|\boldsymbol{x})] = log\frac{P(Y \leq j|\boldsymbol{x})}{1 - P(Y \leq j|\boldsymbol{x})} = log\frac{\pi_1(\boldsymbol{x}) + ... + \pi_j(\boldsymbol{x})}{\pi_{j+1}(\boldsymbol{x}) + ... + \pi_J(\boldsymbol{x})}, j = 1, ..., J{-}1.$$

Den form av kumulativ logistisk regressionsmodell som använts här benämns "proportional odds model" (McCullagh (1980)). Detta syftar på att denna modell antar att oddset för ett värde $\leq j$ (för godtycklig nivå j) på responsvariabeln då prediktorn $\boldsymbol{x}$ har värde $\boldsymbol{x_1}$ är $exp\left[\boldsymbol{\beta}^T(\boldsymbol{x_1} - \boldsymbol{x_2})\right]$ gånger oddset för ett värde $\leq j$ då prediktorn har värde $\boldsymbol{x_2}$, och att denna proportionalitetskonstant alltså är oberoende av vilken nivå j man väljer. I denna modell finns ett intercept för varje kumulativ logit (dvs. J-1 stycken), och en enda effektparameter för varje prediktorvariabel. Det är alltså inte möjligt att med denna modell återge ett mönster där t.ex. doktorsexamen har stor betydelse för lönen bland personer med hög lön men liten betydelse för lönen bland personer med låg lön.

Denna s.k. "proportional odds model" har formen (Agresti (2013) sid. 301)

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

medan den mer generella kumulativa logistiska regressionsmodellen har formen

$$logit\,[P(Y \leq j|\boldsymbol{x})] = \alpha_j + \boldsymbol{\beta_j}^T\boldsymbol{x}, j = 1, ..., J-1.$$

Skillnaden är alltså att den mer generella modellen har en separat parametervektor för varje kumulativ logit. Detta leder till en modell som har totalt $J + (J-1) * p$ parametrar, där p är antalet förklarande variabler, medan "proportional odds"-modellen endast har $J + p$ parametrar.

4.3 Tak- och golveffekter

I detta sammanhang syftar tak- och golveffekter (eng. ceiling and floor effects) på konsekvenser av valet (vid linjär regression) av de s.k. scores som väljs för den högsta respektive lägsta kategorin hos den ordinala responsvariabeln. Om detta val görs olämpligt kan resultaten bli missvisande. Detta problem undviks förstås helt och hållet genom att betrakta responsvariabel som ordinal istället.

5 Genomförande och resultat

Det finns ett antal olika typer av statistiska modeller som kan vara aktuella för denna typ av undersökning, och valet är långt från självklart. En möjlighet, som har fördelen att man får en modell som är lätt att tolka, är att konvertera responsvariabeln till lön som kontinuerlig variabel, så att varje löneintervall motsvarar ett fixt värde hos denna kontinuerliga variabel, och därefter anpassa en multipel linjär regressionsmodell.

5.0.1 Variabelselektion

De förklarande variabler som valdes var

- Kön (Man, Kvinna, "Annat") (nominell variabel)
- Högsta examen uttagen (Ingen, Fil kand, Fil mag, Master, Fil lic, Fil dr, "Annat högre") (nominell)
- Ämne (Biogeovetenskap, Biologi, Datalogi, Fysik, Geografi, Geologi, Geovetenskap, Kemi, Marinbiologi, Matematik, Matematisk statistik, Meteorologi, Miljö- och hälsoskydd, Miljövetenskap, Molekylärbiologi, Naturgeografi, Nutrition, Sjukhusfysik, Socioekologisk resiliens, Utbildningsvetenskap) (nominell)
- Examensarbete (utfört på högskola/universitet eller med extern uppdragsgivare) (nominell)
- Tid som förlupit sedan examen (angiven i hela år) (kontinuerlig)

För de nominella variablerna valdes basnivåerna Man, Fil kand, Matematisk statistik respektive Högskola/universitet.

Att kontrollera för kön verkar rimligt med hänsyn bl. a. till att somliga ämnen är dominerade av det ena eller det andra könet, och det finns risk att ämnesparametrarna skulle påverkas förvrängas av eventuella "köns-effekter" om modellen inte hade någon parameter för kön.

Svarande som angivit "Annat" som ämne utesluts för att inte störa övriga parameterskattningar, eftersom det inte går att kontrollera för ämne

när man inte kan avgöra vilken effekt ämnet i fråga har på lönen (det finns sannolikt flera olika ämnen i "Annat").

För att undersöka huruvida det är rimligt att ha tid sedan examen som kontinuerlig prediktor anpassas först en modell med tid sedan examen som nominell variabel. Koefficienterna för de olika nivåerna på denna visar att effekten kan betraktas som linjär, dvs. skillnaden i förväntad lön mellan ett och två år är ungefär densamma som mellan två och tre år osv.

En automatisk variabelselektionsprocedur (baserad på jämförelse av AIC) körs för respektive modelltyp, med upp till femte ordningens samspel tagna i beaktande. Det finns visserligen modeller med samspelstermer (både för linjär regression och kumulativ logistisk regression) som ger lägre AIC än den slutgiltigt valda modellen, men de modeller som erhölls på dessa sätt måste bedömas som olämpliga. Följande exempel illustrerar varför: den "bästa" linjära regressionsmodellen (enligt AIC) inkluderar de förklarande variablerna ämne, tid sedan examen, högsta examen, examensarbetets uppdragsgivare samt samspel mellan ämne och högsta examen. Enligt denna modell kan den som har kandidatexamen i utbildningsvetenskap räkna med en månadslön på 48 500 kr, medan en kandidat i matematisk statistik får hålla tillgodo med 23 200 kr i månaden. Att detta har föga med verkligheten att göra anar nog de flesta. Vid närmare undersökning finner man att det endast finns en person som angivit utbildningsvetenskap (lärarutbildning) som ämne tillsammans med kandidatexamen som högsta examen. Vederbörande uppger att han tjänstgör som "IT-chef". Detta är ett typiskt fall av överanpassning: modellen har fler parametrar än vad som är befogat med tanke på stickprovets storlek, och ger sämre prediktioner än en enklare modell utan samspel.

Det går att göra modeller som förklarar variationen i datamaterialet betydligt bättre genom att inkludera sysselsättning (såsom t.ex. arbetssökande, studerande, doktorand, fast anställning eller tillfällig anställning) bland de förklarande variablerna. Att vi inte väljer att göra så beror på att vi tror att vissa utbildningar tenderar att leda till relevant arbete i högre grad än andra, och vi vill att detta ska återspeglas i parameterskattningarna för dessa ämnen. Det är av ringa intresse (för vår studie) att en utbildning leder till hög lön givet att man lyckas få arbete, om det är praktiskt taget omöjligt att få arbetet i fråga. Man kan säga att utbildningsvalet påverkar lönen (bland annat) via sysselsättningsgraden. Att då ha sysselsättningsgraden som förklarande variabel vore då i viss utsträckning som att använda responsvariabeln för att förklara sig själv.

5.1 Multipel linjär regression

Man kan fråga sig varför man inte helt enkelt bett de svarande att skriva sin månadslön på en blank rad i enkätformuläret. Det går ju alltid att i efterhand placera sådana svar i intervall, medan det omvända förfarandet

inte låter sig göras. Enkätteknik är föremål för ett omfattande akademiskt studium, och man har funnit (Yan et. al. (2010) sid. 3) att det är betydligt lättare att få någon att ange sin lön genom att kryssa i en ruta än genom att skriva ett belopp. Det finns alltså skäl att tro att man skulle ha fått betydligt färre svar, om man bett de svarande att ange sin lön "exakt".

Det finns förstås olika sätt att omvandla data på denna "intervall-form" till kontinuerlig skala. Ofta ter det sig naturligt att försöka använda det aritmetiska medelvärdet för personerna i intervallen som approximation. Vi känner dock inte till detta medelvärde.

För de sju intervallen i mitten har intervallens mittpunkter valts som lön för alla svarande i respektive intervall, så att alla svarande som uppgivit lön "26 001 - 30 000" har antagits ha månadslönen 28 000 kr osv. För de minsta och största intervallen har olika möjligheter övervägts. Det lägsta intervallet (<18 000 SEK) inkluderar många arbetslösa personer, och sannolikt också många deltidsanställda. Det är mycket svårt att uttala sig om vilka siffror som egentligen ligger bakom uppgiften. Även det högsta intervallet (>46 000) är svårt att tilldela ett bestämt värde.

En ansats är att basera valet på officiell data över den empiriska fördelningen för månadsinkomst i Sverige. Det finns offentliga data från Skattemyndigheten för taxeringsår 2012, där man kan se alla percentiler för den svenska befolkningens månadsinkomst (inkomst av tjänst samt näringsverksamhet) inklusive aritmetiska medelvärde för varje percentil. Det vill säga: man kan t.ex. se att Sveriges befolknings bäst tjänande hundradel tjänade i genomsnitt 114 600 SEK/månad före skatt, och så vidare. Till följd av detta kan man approximativt beräkna den genomsnittliga månadsinkomsten för personer i den kvantil som tjänar över 46 000 SEK/månad, vilket visar sig vara 70 500 SEK (personer under 20 och över 65 års ålder exkluderas från kalkylen). Motsvarande beräkning för gruppen under 18 000 SEK ger en genomsnittlig månadsinkomst på 8 400 SEK. Det kan invändas att dessa uppgifter är missvisande, då de kommer från en population som inkluderar såväl förtidspensionärer som näringslivstoppar med lång arbetslivserfarenhet, vilka normalt inte förekommer i enkätmaterialet. Med andra ord: den empiriska fördelningen för den arbetsföra svenska befolkningens inkomster skiljer sig troligen alltför mycket från motsvarande fördelning för naturvetarna som här studeras. Därför förkastas denna ansats.

Enligt Clementi & Gallegati (2008) gäller att inkomsten X för en slumpmässigt vald person har en blandad sannolikhetsfördelning av typen

$$P(X \leq x) = \alpha F_1(x) + (1-\alpha)F_2(x), 0 < \alpha < 1,$$

där F_1 är lognormalfördelad och F_2 Pareto-fördelad. Man menar då att inkomsten för "höginkomsttagare" (ca 1-3 % av befolkningen) är Pareto-fördelad, medan inkomsten för resten (dvs. $\alpha * 100\%$) av befolkningen är lognormalfördelad. Denna observation avsåg Tyskland, Storbritannien och

USA, med olika parametrar för de olika länderna, och det verkar skäligt att anta att det även stämmer i Sverige. Inkomst och lön är inte synonyma begrepp, men de tenderar att praktiskt taget sammanfalla för majoriteten av befolkningen. En förenkling som verkar någorlunda rimlig är att anta att de svarande på enkäten inte har de allra högsta lönerna bland Sveriges befolkning (detta med hänsyn till att de nyligen tagit examen eller gjort ett examensarbete), och att deras löner beskrivs väl av en lognormalfördelning. Därefter antar man att de som angivit löner i de icke-extrema intervallen tjänar exakt den lön som är mitt i respektive angivna intervall. Sedan kan man betrakta de observerade lönerna som ett "censurerat" stickprov från en lognormalfördelning, där vi, för lönerna under 18 000 SEK respektive över 46 000 SEK endast vet antalet observationer, och inget närmare om deras storlek. Låt n_j vara antalet observationer i lönekategori j. Då ges likelihoodfunktionen av

$$L(\mu, \sigma^2) = F_1(18000)^{n_1} * f_2(20000)^{n_2} * ... * (1 - F_2(46000))^{n_9}.$$

Denna maximeras numeriskt med Newton-Raphson-metoden och ger upphov till ML-skattningarna $\mu = 10.33116$ och $\sigma = 0.2727496$ för väntevärde och standardavvikelse på logaritmisk skala, vilket innebär att lönen har väntevärdet $exp(10.33116 + 0.5 * 0.2727496^2) = 31836$ SEK och standardavvikelsen $\sqrt{exp(2 * 10.33116 + \sigma^2) * (exp0.2727496^2) - 1)} = 8847.307$ SEK.

Därefter simuleras 10^8 st. slumptal från denna lognormalfördelning i R. Det aritmetiska medelvärdet av de simulerade slumptal som är mindre än 18 000 respektive större än 46 000 beräknas till 16 302.01 och 52 182.24. Vi räknar därför från och med nu (i de linjära regressionsmodellerna) med att alla som tjänar mindre än 18 000 SEK tjänar exakt 16 302.01 SEK, och att alla som tjänar mer än 46 000 SEK tjänar exakt 52 182.24 SEK.

Parameterskattningarna för den multipla linjära regressionsmodellen anges i tabell 1.

Tabell 1: Parameterskattningar för multipel linjär regressionsmodell

	Värde	Medelfel	t-värde	p-värde
(Intercept)	27322.880	1855.229	14.727	0
Kön: Man	0			
Ämne: Biologi	-7582.583	1589.819	-4.769	0
Ämne: Fysik	-7981.817	1914.507	-4.169	0
Ämne: Geovetenskap	-5579.567	1623.761	-3.436	0.001
Ämne: Kemi	-9227.638	1775.755	-5.196	0
Ämne: Matematik	-6964.439	2266.547	-3.073	0.002
Ämne: Miljövetenskap	-5654.192	1840.535	-3.072	0.002
Ämne: Nutrition	-8504.396	2117.747	-4.016	0
Ämne: Sjukhusfysik	-4205.794	2412.837	-1.743	0.082
Ämne: Utbildningsvetenskap	-125.215	2007.528	-0.062	0.950
Ämne: Matematisk statistik	0			
Kön: Annat	-3768.632	2712.604	-1.389	0.166
Kön: Kvinna	-248.020	730.487	-0.340	0.734
Högsta examen: Fil lic/Fil dr	6620.047	1988.587	3.329	0.001
Högsta examen: Ingen examen uttagen	-267.712	4154.742	-0.064	0.949
Högsta examen: Master	2843.224	997.037	2.852	0.005
Tid sedan examen	1861.298	326.719	5.697	0
Högsta examen: Fil kand	0			
Examensarbete: Extern uppdragsgivare	3229.551	1046.205	3.087	0.002
Examensarbete: Högskola/universitet	0			

5.1.1 Känslighetsanalys

Koefficientskattningen för kvinnligt kön påverkas mycket av vilken månadslön man tillskriver den bäst tjänande gruppen: ett högt värde ger kraftigt negativ parameter för kvinnor. Däremot påverkas detta parametervärde knappast alls när man ändrar värdet för den sämst tjänande gruppens lön. Detta återspeglar det faktum att kvinnor varken är överrepresenterade eller underrepresenterade i denna grupp.

För övrigt observerades följande:

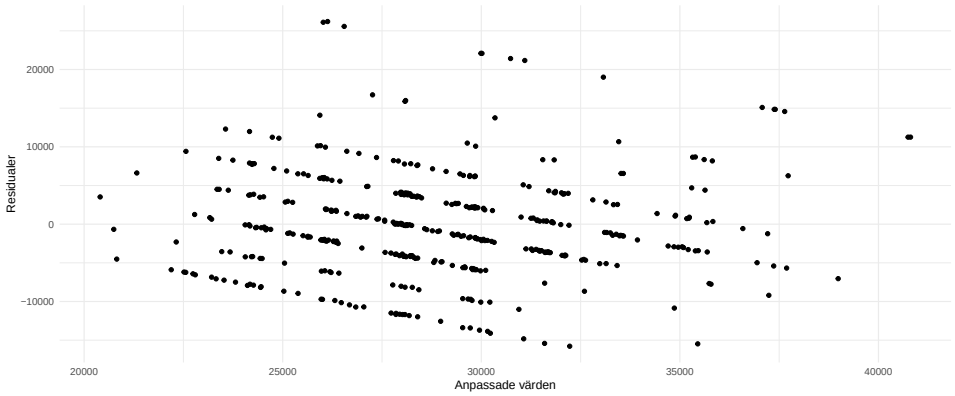
- I den högsta lönekategorin återfinns 3.7 % av männen men endast 1.9 % av kvinnorna.
- I den lägsta lönekategorin återfinns 10.5 % av männen och 9.3 % av kvinnorna.
- Vissa ämnens parameterskattningar var känsliga för de angivna värdena för de högsta respektive lägsta lönerna. Detta var särskilt påtagligt för fysik, som hade relativt många observationer i den lägsta lönekategorin och därför fick en kraftigt negativ effektparameter när den lägsta lönen angavs som mycket låg.
- Parameterskattningarna för högsta examen påverkades obetydligt av värdena som angavs för de högsta och lägsta lönerna.

5.1.2 Modelldiagnostik

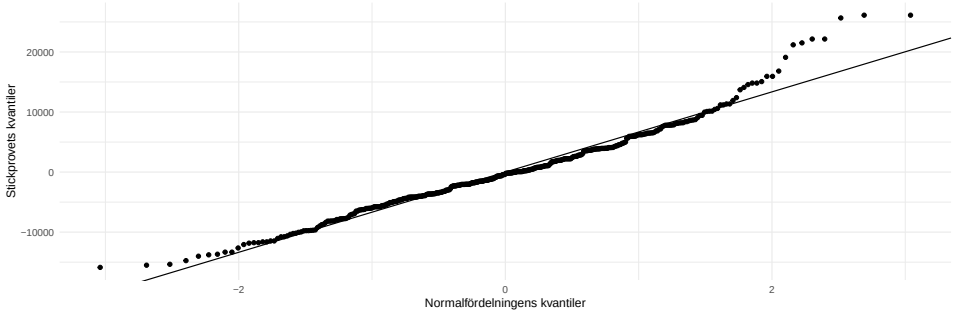
Denna modell har en diskret responsvariabel som endast antar nio olika värden. Det linjära uttrycket i modellens parametrar däremot kan anta ett stort antal olika värden (genom olika kombinationer av nivåer hos de förklarande variablerna). Detta innebär att residualerna, för varje anpassat värde, endast kan anta något av nio olika värden. Naturligtvis är detta inte ett idealiskt fall för linjär regression.

Antagandet om residualernas homoskedasticitet är svårt att bedöma grafiskt eftersom nästan alla anpassade värden fanns i mitten av intervallet. Residualerna delades in i fyra grupper baserat på storleken på motsvarande anpassade värden. Stickprovsstandardavvikelsen (enhet SEK) hos dessa fyra grupper var 7 344.406, 7 671.938, 6 940.059 och 6 030.978, dvs. en viss heteroskedasticitet råder.

En Q-Q-plot mot normalfördelningens kvantiler avslöjar att antagandet om normalfördelade residualer inte är uppfyllt. I allmänhet gäller att de största anpassade värdena inte är tillräckligt stora, och att de minsta inte är tillräckligt små. Shapiro-Wilks normalitetstest ger ett p-värde på $5.5 * 10^{-8}$, dvs. mycket starkt bevis mot hypotesen att residualerna är normalfördelade. Faktum är att residualerna inte ens ser ut att ha en symmetrisk fördelning.



Figur 1: Residualer mot anpassade värden



Figur 2: Residualer, Q-Q-plot

Förklaringsgraden R^2 är låg (0.20) vilket får tolkas som att det finns en hel del av variationen i personernas lön som inte kan hänföras till de valda förklarande variablerna.

I den angivna multipla linjära regressionsmodellen är tid sedan examen en kontinuerlig förklarande variabel (som dock i praktiken endast antar några få olika värden) medan övriga är dummyvariabler. Inga samspelseffekter finns i modellen. För kön i, ämne j, högsta examen k (betraktas här som nominell), examensarbete gjort på plats l och tid x år förlöpt sedan examen, gäller alltså att väntevärdet för månadslön är

$\mu_{ijklx} = \alpha + \beta_{1i} + \beta_{2j} + \beta_{3k} + \beta_{4l} + \beta_5 x.$

5.2 Kumulativ logistisk regression

Parameterskattningarna för den erhållna kumulativa logistiska regressionsmodellen presenteras i tabell 2.

Tabell 2: Parameterskattningar för kumulativ logistisk regressionsmodell

	Värde	Medelfel	t-värde	p-värde	Oddskvot
Kön: Man	0				1
Kön: Annat	-0.954	0.710	-1.344	0.179	0.385
Kön: Kvinna	-0.001	0.188	-0.005	0.996	0.999
Ämne: Matematisk statistik	0				1
Ämne: Biologi	-1.904	0.452	-4.209	0	0.149
Ämne: Fysik	-1.933	0.549	-3.522	0	0.145
Ämne: Geovetenskap	-1.132	0.452	-2.507	0.012	0.322
Ämne: Kemi	-2.369	0.501	-4.725	0	0.094
Ämne: Matematik	-1.715	0.638	-2.687	0.007	0.180
Ämne: Miljövetenskap	-1.174	0.497	-2.364	0.018	0.309
Ämne: Nutrition	-2.020	0.569	-3.552	0	0.133
Ämne: Sjukhusfysik	-0.587	0.633	-0.928	0.354	0.556
Ämne: Utbildningsvetenskap	0.285	0.539	0.528	0.597	1.330
Examensarbete: Högskola/universitet	0				1
Examensarbete: Extern uppdragsgivare	0.775	0.275	2.824	0.005	2.171
Tid sedan examen	0.543	0.087	6.213	0	1.721
Högsta examen: Fil kand	0				1
Högsta examen: Fil lic/Fil dr	1.893	0.535	3.539	0	6.639
Högsta examen: Ingen examen uttagen	0.297	1.152	0.258	0.797	1.346
Högsta examen: Master	0.723	0.266	2.715	0.007	2.061
<18 000|18 000 - 22 000	-1.801	0.513	-3.509	0	0.165
18 000 - 22 000|22 001 - 26 000	-1.145	0.506	-2.261	0.024	0.318
22 001 - 26 000|26 001 - 30 000	-0.044	0.506	-0.087	0.931	0.957
26 001 - 30 000|30 001 - 34 000	1.142	0.511	2.236	0.025	3.132
30 001 - 34 000|34 001 - 38 000	2.491	0.522	4.774	0	12.070
34 001 - 38 000|38 001 - 42 000	3.668	0.539	6.810	0	39.174
38 001 - 42 000|42 001 - 46 000	4.045	0.547	7.393	0	57.095
42 001 - 46 000|>46 000	4.654	0.569	8.185	0	104.999

5.2.1 Modelldiagnostik

För att testa huruvida antagandet om proportionella odds håller, anpassades en mer generell modell med en separat uppsättning effektparametrar för varje logit. Loglikelihood för modellerna var -748.2816 (df=23) respektive -690.7472 (df=128) vilket ger en likelihoodkvot-statistika på 115.0687 vid 128-23=105 frihetsgrader, motsvarande ett p-värde på 0.2359375. Detta tolkas som att det inte finns skäl att förkasta antagandet om proportionella odds.

5.3 Jämförelse mellan modellerna

Den kumulativa logistiska regressionsmodellen antyder ungefär samma saker som den multipla linjära regressionsmodellen: de saker som är bra för lönen i den ena modellen är bra för lönen i den andra också, med ett par undantag. Till exempel: den logistiska regressionsmodellen antyder att det är aningen bättre att vara kvinna än att vara man, medan den linjära regressionsmodellen säger motsatsen. Vad gäller de större effekterna är de båda modellerna dock samstämmiga.

6 Slutsats

De huvudsakliga slutsatserna av denna undersökning är följande:

- En manlig fil kand i matematisk statistik som skrivit examensarbete på institutionen och tagit examen för ett år sedan förväntas tjäna ca. 29 200 SEK/månad enligt den linjära regressionsmodellen och mellan 26 001 och 30 000 SEK/månad enligt den kumulativa logistiska regressionsmodellen (baserat på vilken lönekategori som är mest sannolik enligt modellen).
- Om man istället för endast en Fil kand-examen har masterexamen, förväntas man tjäna 32027.41 respektive 30 001 - 34 000 SEK/månad.
- Har man dessutom gjort examensarbete hos extern uppdragsgivare förväntas man tjäna 35 300 respektive 30 001 - 34 000 SEK/månad.
- När det dessutom har gått fyra år sedan examen, förväntas man tjäna 41 000 respektive >46 000 SEK/månad, oavsett om man är kvinna eller man. Är man däremot av "annat" kön är 37 000 respektive 34 001 - 38 000 SEK/månad de predikterade lönerna.

De två modellerna indikerar på det stora hela samma effekter av olika nivåer på prediktorerna, men den kumulativa logistiska modellen lyckas

bättre med att prediktera de högsta och de lägsta lönerna. Den linjära regressionsmodellen tenderar att prediktera värden för nära mitten av löneintervallet för de personer som de facto har de högsta och lägsta lönerna.

Ett antal parameterskattningar i den linjära regressionsmodellen har låga p-värden enligt datorutskrifterna, men eftersom modellantagandena är så till den grad illa uppfyllda kan man inte av detta dra slutsatser om effekternas statistiska signifikans.

Viktigt att minnas är att den största delen av variationen i månadslön hos den undersökta poulationen inte förklaras genom de variabler vi undersökt här. Det måste också poängteras att data är insamlade år 2015. Arbetsmarknaden kan naturligtvis ha förändrats sedan dess.

Det kan verka märkligt att effekten av att inte ta ut examen inte är sämre än vad modellerna indikerar. Det bör då observeras att samtliga svarande i denna enkät har inkluderats för att de antingen gjort ett examensarbete eller tagit ut en examen. Personer som avbrutit sina studier utan att vare sig ha tagit ut examen eller skrivit ett examensarbete är alltså inte inkluderade i studien. Det finns också svarande bland de som inte tagit ut examen, som uppger att de erbjudits en för utbildningen relevant anställning innan de hunnit fullborda studierna. Detta har naturligtvis betydelse för lönen.

Det finns utrymme för att utveckla studien. Vad gäller den linjära regressionsmodellen finns t. ex. bootstrapmetoder för att beräkna konfidensintervall för parameterskattningarna, som lämpar sig för fall då de ordinarie parametriska metoderna inte är tillämpliga. Vad gäller den kumulativa logistiska regressionsmodellen, kan modelldiagnostiken kompletteras med analys av s.k. “surrogate residuals” (Greenwell et. al. 2018).

Referenser

[1] Agresti, A. (1994) *Analysis of Ordinal Categorical Data*, John Wiley & Sons, Inc. http://www.ctan.org/tex-archive/info/lshort/english/lshort.pdf

[2] Agresti, A. (2013) *Categorical Data Analysis, Third Edition*, John Wiley & Sons, Inc.

[3] Alm, S. E. & Britton, T. (2008) *Stokastik: Sannolikhetsteori och statistikteori med tillämpningar*, Liber AB

[4] Clementi, F. & Gallegati, M. (2005) *Pareto's Law of Income Distribution: Evidence for Germany, the United Kingdom, and the United States.* Utgör sid. 3-14 i Chatterjee, A., Yarlagadda, S., & Chakrabarti B. K. (2005) *Econophysics of Wealth Distributions*, Springer-Verlag Italia

[5] Greenwell, B. M., McCarthy, A. J., Boehmke, B. C., Liu, D. (2018) *Residuals and Diagnostics for Binary and Ordinal Regression Models.* The R Journal Vol. 10/1, juli 2018

[6] Kullberg, C. & Högberg, C. (2016) *Efter Studierna, Naturvetare i arbetslivet VI*, Naturvetenskapliga fakulteten, Stockholms universitet https://pp-prod-admin.it.su.se/polopoly_fs/1.274044.1457364772!/menu/standard/file/EfterstudiernaVI_webb.pdf

[7] Hlavac, M. (2018) *stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2.*, Journal of the Royal Statistical Society. Series B (Methodological) Vol. 42, No. 2 (1980), sid. 109-142 https://CRAN.R-project.org/package=stargazer

[8] McCullagh, P. (1980) *Regression Models for Ordinal Data*, Journal of the Royal Statistical Society. Series B (Methodological) Vol. 42, No. 2 (1980), sid. 109-142

[9] R Core Team (2018) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Wien, Österrike https://www.R-project.org/

[10] Thompson, L. A. (2009) *R (and S-PLUS) Manual to Accompany Agresti's Categorical Data Analysis (2002)*, gratis manual tillhandahållen av Alan Agresti

[11] Hlavac, M. (2018) *stargazer: Well-Formatted Regression and Summary Statistics Tables. R package version 5.2.2.*, Journal of the Royal Statistical Society. Series B (Methodological) Vol. 42, No. 2 (1980), sid. 109-142 https://CRAN.R-project.org/package=stargazer

[12] Venables, W. N. & Ripley, B. D. (2002), Fourth Edition. *Modern Applied Statistics with S*, Springer, New York, USA

[13] Walker, A. (2018) *openxlsx: Read, Write and Edit XLSX Files. R package version 4.1.0.*, https://CRAN.R-project.org/package=openxlsx

[14] Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis* Springer-Verlag New York https://ggplot2.tidyverse.org

[15] Witting, A. (2014) *Skulle du välja samma utbildning igen? - en jämförande studie av alumner i matematik och matematisk statistik*, Kandidatuppsats 2014:8, juni 2014, Stockholms universitet

[16] Yan, T., Curtin, R., Jans, M. (2010) *Trends in Income Nonresponse Over Two Decades*, Journal of Official Statistics, Vol. 26, No. 1, 2010, sid. 145?164

Appendix

Detaljer om databehandling

Det fanns många enkätsvar som innehöll all relevant information för denna studie, men som var i behov av någon form av korrigering för att kunna användas. 310 svarande saknade antingen uppgift om huvudämne i högsta examen eller löneuppgift, och uteslöts därför. I de fall då kandidatexamen tagits ut efter en högre examen, har "tid sedan examen" räknats från året då senaste kandidatexamen tagits ut.

6.0.1 Examensarbete

Många svarade "Annat" på frågan om var de gjort sitt examensarbete, men angav intill sitt svar något som bedömdes passa in i någon av kategorierna "Högskola/universitet" eller "Extern uppdragsgivare". Följande svar har således kategoriserats som "Högskola/universitet":

Prince Henry's Institute for Medical Research, Melbourne, Australien, Kandidatarbete: universitet i Kanada. Masterarbete: SU, Oxford, Storbritannien, Ett samarbete med SU och Org Kemi KTH, men examen föll på SU!, University of Pennsylvania, MEB på KI, Dalarnas Högskola , karolinska institutet, Samarbete Sthlm Universitet/KTH, Stockholms univ. och Rhodes univ. Sydafrika, Karolinska institutet, University of Utah, Salt Lake City, USA, Karolinska Sjukhuset, Var på Tokyo universitet men examinerades av su, SU samt SU & KTH (magister & master), Kandidat på KTH, Master på SU, Karolinska Institutet, NOVUM på Karolinska Institutet som samarbetar med SU på nutritionsutbildningen., Samarbete mellan SU och UNM i Albuquerque (New Mexico), utomlands, universitet.

Följande svar har kategoriserats som "Extern uppdragsgivare":

Huddinge kommun, miljötillsynsavdelningen, Länsstyrelsen Dalarna, På Stockholms universitet men för ett företag., Södertälje kommun, Trafikverket, kommun, Kommun, Innventia, Karolinska sjukhuset, På miljö- och teknikkonsultföretaget WSP (Stockholm), SCB, Smittskyddsinstitutet, Södra roslagens miljö o hälsoskyddsförbund, Företaget, Livsmedelsverket, Företag, En kommun, Naturhistoriska riksmuséet, IVL Svenska miljöinstitutet.

6.0.2 Examensår

I somliga fall hade den svarande inte angivit något svar alls på frågan om examensår, men däremot angivit examensåret i "Kommentar till examen". I sådana fall har det aktuella examensåret skrivits in under artikeln om examensår.

För de svarande som angett året för kandidatexamen, och svarat "Vet ej" på frågan om året för senaste examen, har kandidatexamen antagits vara den senaste examen.

6.0.3 Högsta examen

Nivåerna på variabeln "högsta examen" har hanterats på följande sätt:

- Alla som svarat "Annat fil kand" har tilldelats nivån "Fil kand".
- Filosofie licentiat och filosofie doktor har slagits samman.
- De svarande som ej tagit ut examen har fått en egen nivå på variabeln "högsta examen".
- Alla övriga (Fil mag, Master, Sjukhusfysiker, Sjukhusfysiker högre, Lärarexamen, Lärarexamen högre) har tilldelats nivån "Master".

6.0.4 Ämne

Den ursprungliga enkäten hade 27 svarsalternativ för ämne. Efter att ha exkluderat svarande som inte besvarat alla frågor som undersökts inom ramen för detta examensarbete, fanns åtskilliga ämnen som endast representerades av en handfull svarande. De ämnen som inte har en egen parameter i de modeller som presenteras nedan, har slagits samman med liknande ämnen på grundval av organisatoriska hänsyn. Sålunda har exempelvis marinbiologi och toxikologi slagits ihop med biologi eftersom ämnena hör till samma sektion på Stockholms universitet. Sorteringen har ägt rum på följande sätt:

- Toxikologi, marinbiologi, molekylärbiologi, hållbart företagande och socioekologisk resiliens har sorterats under "Biologi".
- Biogeovetenskap, geografi, geologi och naturgeografi har sorterats under "Geovetenskap".
- Biofysik har sorterats under "Kemi".
- Biomatematik och beräkningsbiologi har sorterats under "Matematik".
- Finansmatematik och försäkringsmatematik har sorterats under "Matematisk statistik".
- Miljö- och hälsoskydd har sorterats under "Miljövetenskap".
- Meteorologi och astronomi har sorterats under "Fysik".
- De sex svarande som läst datalogi eller beräkningsteknik kunde inte rimligen slås ihop med någon annan grupp och exkluderades därför från vidare undersökning.
- De svarande som angivit "Annat" som ämne, och inte kunnat placeras i någon passande kategori, har likaså exkluderats.

- En svarande uppgav att han har gått Psykologprogrammet och uppger ingen utbildning på Naturvetenskaplig fakultet. Denna observation uteslöts.

Välkommen till Naturvetenskapliga fakultetens alumnundersökning!

Vi vill göra våra utbildningar ännu bättre och utför därför en uppföljning av studenter som gjort examensarbete och/eller tagit ut examen hos oss. Våra resultat och därmed möjligheter att göra utbildningen bättre beror helt på att så många som möjligt svarar på enkäten, så vi är väldigt glada om du har tid att hjälpa oss. Enkäten är helt anonym och tar ca 10 minuter att svara på. Stort tack för din hjälp!

1. Kön

☐ Man	☐ Kvinna	☐ Annat

2. Hur många terminer har du läst vid Stockholms universitet?

☐ Mindre än 1 termin	☐ 1-3 terminer	☐ 4-6 terminer	☐ Mer än 6 terminer

3. Om du har bedrivit studier på kandidatnivå ber vi dig fylla i uppgifterna nedan

Typ av exemen på kandidatnivå

☐ Fil Kand	☐ Lärarexamen	☐ Sjukhusfysiker	☐ Annat	☐ Ingen exemen uttagen

Examensår på kandidatnivå

☐ 2011	☐ 2012	☐ 2013	☐ 2014	☐ Vet ej

4. Om du har bedrivit studier på högre nivå än kandidatnivå ber vi dig att fylla i uppgifterna nedan

Typ av examen

☐ Fil mag	☐ Master	☐ Fil lic	☐ Fil dr	☐ Lärarexemen	☐ Sjukhusfysiker	☐ Annat	☐ Ingen examen uttagen

Examensår på högre nivå än kandidatnivå:

☐ 2011	☐ 2012	☐ 2013	☐ 2014	☐ Vet ej

5. Vilket huvudämne/huvudområde har din utbildning på kandidatnivå?

☐ Astronomi	☐ Beräkningsteknik	☐ Biofysik	☐ Biogeo	☐ Biomatematik och beräkningsbiologi
☐ Biologi	☐ Datalogi	☐ Fysik	☐ Geografi	☐ Geologi
☐ Geovetenskap	☐ Kemi	☐ Lärarutbildning	☐ Marinbiologi	☐ Matematik
☐ Matematisk statistik	☐ Meteorologi	☐ Molekylärbiologi	☐ Nutrition	☐ Annat, vänligen ange vad_____

6. Vilket huvudämne/ huvudområde har din utbildning på högre nivå än kandidat (magister/master)?

☐ Astronomi	☐ Beräkningsteknik	☐ Biofysik	☐ Biogeo	☐ Biomatematik och beräkningsbiologi
☐ Biologi	☐ Datalogi	☐ Fysik	☐ Geografi	☐ Geologi
☐ Geovetenskap	☐ Kemi	☐ Lärarutbildning	☐ Marinbiologi	☐ Matematik
☐ Matematisk statistik	☐ Meteorologi	☐ Miljö- och hälso-skydd	☐ Molekylärbiologi	☐ Nutrition
☐ Sjukhusfysik	☐ Annat, vänligen ange vad_____			

7. Är du nöjd med din utbildning?

☐ Nej, definitivt inte	☐ Nej, inte riktigt	☐ Ja, ganska	☐ Ja, defivitivt	☐ Vet ej

8. Skulle du rekommendera din utbildning till andra?

☐ Ja	☐ Nej	☐ Vet inte

9. Hur bedömer du kvaliteten på din utbildning inom naturvetenskapliga fakulteten vid Stockholms universitet med tanke på:

Forskningsanknytning

☐ Mycket bristfällig	☐ Delvis bristfällig	☐ Bra	☐ Mycket bra

Arbets- och näringslivsanknytning

☐ Mycket bristfällig	☐ Delvis bristfällig	☐ Bra	☐ Mycket bra

10. Hur bedömer du den kompetens du erhållit i din utbildning vid naturvetenskapliga fakulteten vid Stockholms universitet med tanke på:

Kunskaper i...

	Mycket bristfällig	Delvis bristfällig	Bra	Mycket bra	Vet ej
ditt huvudämne	☐	☐	☐	☐	☐
ämnets vetenskapliga grund	☐	☐	☐	☐	☐
laborativa tekniker	☐	☐	☐	☐	☐
fältstudier	☐	☐	☐	☐	☐
aktuella forskningsfrågor	☐	☐	☐	☐	☐

Färdigheter att...

	Mycket bristfällig	Delvis bristfällig	Bra	Mycket bra	Vet ej
hitta och värdera ny kunskap	☐	☐	☐	☐	☐
kritiskt tolka och diskuter problem-ställningar	☐	☐	☐	☐	☐
argumentera och övertyga	☐	☐	☐	☐	☐
självständigt identifiera och lösa problem	☐	☐	☐	☐	☐
följa med i kunskapsutvecklingen	☐	☐	☐	☐	☐
undervisa	☐	☐	☐	☐	☐
bearbeta statistiskt material	☐	☐	☐	☐	☐
göra muntliga presentationer	☐	☐	☐	☐	☐
göra skriftliga presentationer	☐	☐	☐	☐	☐
göra presentationer på engelska	☐	☐	☐	☐	☐
göra etiska bedömningar	☐	☐	☐	☐	☐
förklara för ickespecialister/lekmän	☐	☐	☐	☐	☐
arbeta i grupp	☐	☐	☐	☐	☐
projektledning	☐	☐	☐	☐	☐

11. Var det något i din utbildning som var särskilt värdefullt?

☐ Nej	☐ Ja. Utveckla gärna:_____

12. Var det något som du saknade i utbildningen?

☐ Nej	☐ Ja. Utveckla gärna:_____

13. Var utförde du ditt examensarbete?

☐ Inom Stockholms universitet	☐ Annat universitet/högskola	☐ Extern uppdragsgivare	☐ Annat. Ange var:_____

14. Vad är din nuvarande sysselsättning?

☐ Fast anställning	☐ Tillfällig anställning	☐ Egen firma	☐ Studerar
☐ Forskarutbildning	☐ Föräldraledig	☐ Arbetssökande	☐ Annat, ange vad:_____

Om forskarutbildning, vilken institution? _____

Om forskarutbildning, vilket universitet/högskola?_____

15. Hur fick du ditt FÖRSTA arbete efter avslutad utbildning?

☐ Tog själv kontakt	☐ Via annons	☐ Via arbetsförmedling	☐ Via exemensarbetet	☐ Via lärare	☐ Annat, ange vad:___

16. Ditt första arbete...

Vad för arbete var det?_____

Vilken arbetsgivare?_____

17. Har du haft något arbete som varat 6 månader eller längre sedan du avslutade din utbildning?

☐ Ja	☐ Nej	☐ Ja, jag går forskarutbildningen för närvarande

18. Hur fick du ditt nuvarande arbete?

☐ Tog själv kontakt	☐ Via annons	☐ Via arbetsförmedlingen	☐ Via examensarbetet	☐ Via lärare	☐ Annat sätt, ange hur:

19. Ditt nuvarande arbete...

Vilken arbetsgivare har du?_____

Vad arbetar du med?_____

20. Om du för närvarande är fast eller tillfälligt anställd, eller har egen firma, vilka är dina huvudsakliga arbetsuppgifter? Rangordna alternativen 1-3, där 1= det du gör största delen av tiden (en aktivitet/siffra).

	1	2	3
Administrativa uppgifter	☐	☐	☐
Arbetsledning	☐	☐	☐
Forskning och utveckling	☐	☐	☐
Kontroll	☐	☐	☐
Marknadsföring och försäljning	☐	☐	☐
Produktion	☐	☐	☐
Undervisning	☐	☐	☐
Utredning och planering	☐	☐	☐
Och/eller annat, ange vad:	☐	☐	☐

Om annat, ange vad:_____

21. I vilken utsträckning har du i ditt arbete under den senaste månaden haft arbetsuppgifter som ställt krav på:

Kunskaper i...

	Inte alls	Till viss del	Till stor del	Helt
ditt huvudämne	☐	☐	☐	☐
ämnets vetenskapliga grund	☐	☐	☐	☐
laborativa tekniker	☐	☐	☐	☐
fältstudier	☐	☐	☐	☐
aktuella forskningsfrågor	☐	☐	☐	☐

Färdigheter att...

	Inte alls	Till viss del	Till stor del	Helt
hitta och värdera ny kunskap	☐	☐	☐	☐
kritiskt tolka och diskutera problemställningar	☐	☐	☐	☐
argumentera och övertyga	☐	☐	☐	☐
självständigt identifiera och lösa problem	☐	☐	☐	☐
följa med i kunskapsutvecklingen	☐	☐	☐	☐
undervisa	☐	☐	☐	☐
bearbeta statistiskt material	☐	☐	☐	☐
göra muntliga presentationer	☐	☐	☐	☐
göra skriftliga presentationer	☐	☐	☐	☐
göra presentationer på engelska	☐	☐	☐	☐
göra etiska bedömningar	☐	☐	☐	☐
förklara för ickespecialister/lekmän	☐	☐	☐	☐
arbeta i grupp	☐	☐	☐	☐
projektledning	☐	☐	☐	☐

22. Hur viktiga för dina nuvarande arbetsuppgifter är de kunskaper och färdigheter som du förvärvat genom...

	Helt oviktiga	Delvis oviktiga	Viktiga	Mycket viktiga	Ej relevant då jag ej har erfarenhet av detta
Utbildning vid naturvetenskapliga fakulteten	☐	☐	☐	☐	
Poänggivande praktik	☐	☐	☐	☐	☐
Att varit aktiv i studentförening	☐	☐	☐	☐	☐
Arbetserfarenhet	☐	☐	☐	☐	☐

23. Har din utbildning relevans för ditt nuvarande arbete?

☐ Ingen relevans	☐ Viss relevans	☐ Relevans	☐ Hög relevans

24. Är det någon annan utbildning som du upplever som konkurrerande, eventuellt bättre, för att passa ditt arbete?

☐ Ja	☐ Nej	☐ Min utbildning vid naturvetenskapliga fakulteten är ej relevant för mitt nuvarande arbete

Om ja, vilken?_____

Om ja, på vilket sätt?_____

25. Din nuvarande bruttolön per månad?

☐ < 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	☐ > 46 000	

26. På vilket sätt skulle du vilja ha fortsatt kontakt med universitetet?

☐ Gå med i alumnnätverket	☐ Ta emot en praktikant	☐ Ta emot en student som skriver uppsats/ examensarbete	☐ Gästföreläsa	☐ Vara mentor	☐ Jag vill inte ha fortsatt kontakt	☐ Annat, ange vad:_____

Dina kontaktuppgifter (frivilligt)_____