Lunds Universitet
Statistiska institutionen

Statistik 61-80 p
Vt 2007
Handledare: Björn Holmquist

Utvärdering av arbetsmarknadsprojektet Enter - en logistisk regression av bakgrundsvariabler.

Mattias Sellin

Abstract

This thesis purpose is to determine an explanation of the efficiency in a labour-market related project. “Enter” is an EU-financed project with a focus on long-term unemployed workers in the district Härlanda, Gothenburg. Starting in September 2004, Enter offers individually based coaching and job guidance to totally 110 young persons with problems entering the labour market. The goal is to provide these people courage to find either full or part time employment or to start study.

To detect in which cases the project is successful, or “efficient”, a logistic regression model is used. The model consists of the dependent binary variable efficiency, which are participators that have succeeded finding either a fulltime job or education within less than 6 months or not. The independent variables contain individual data, in categories relevant on the dependent.

The outcome shows three main variables for under current conditions predicting whether a person is going to find an employment within 6 months or not; impediment of work, the time in dependency of governmental subsidies and housing problems.
Housing problems shows a crucial effect for finding a job or not, which is confirmed by the project management. Impediment of work shows a large negative effect for problems such as drug abuse, psychical problems and neurone physic problems.

Finally, a long time outside the labour market is more likely to make it even longer. Time in dependency of governmental subsidies is the only significant independent variable for clients without impediment of work or housing problems. The logistic regression model estimates that for clients with less than 6 months unemployment (and no other documented problems) the chance of finding a job in efficient time is 52,7 %, but the probability shrinks down to less than 1 % for participants with 10 months of unemployment or more.

Innehållsförteckning sida

1 Inledning 1

1.1 Introduktion 1

1.2 Syfte och problemformulering 2

1.3 Bakgrund 3

2 Data **7**

3 Teori **11**

3.1 Val av statistisk metod 11

3.2 Logistisk Regression 11

3.3 Genomförande 13

4 Resultat **14**

5 Sammanfattning **17**

6 Referenslista **19**

7 Appendix **20**

Bilaga 1: Abstract

Inledning

1.1 Introduktion

Enter är sedan september 2004 ett arbetsmarknadspolitiskt projekt inom Enheten för Arbete och Integration i Göteborgs Stad, baserat i stadsdelen Härlanda. Projektets syfte är att stödja integrering av långtidsarbetslösa och långtidssjukskrivna ungdomar i åldern 18-30 år till sysselsättning i antingen studier eller arbetsliv. Insatserna är dynamiska och förväntas vara individanpassade med målsättningen att främja deltagarnas möjligheter till en egen försörjning. Projektet pågår under tre år och ämnar inkludera sammanlagt cirka 110 personer.

Efter två års verksamhet har projektets ansvariga beslutat genomföra en utvärdering av de hittills genomförda åtgärderna. Flera delrapporter har tidigare presenterats, men innehåller lite information och jämförelser av data på individnivå. Det behövs en utvärdering som inte bara visar om man uppnått de satta målen, utan även för vilka grupper av individer projektet är effektivt eller ej.

I samarbete med projektansvarig Jenny Forsberg har under de senaste månaderna relevanta data och frågeställningar tagits fram för att kunna göra en god bedömning av verksamheten.

1.2 Syfte och problemformulering

På grund av den individorienterade modellen man valt finns ingen entydig modell för hur olika grupper skall hanteras, då både insatser och klienternas bakgrund varierar. Uppsatsens målsättning är därför att analysera projektets individer och se vilka som, efter att ha deltagit i Enter, har bäst förutsättningar att ta sig in i arbetslivet.

Tillsammans med projektansvarig har jag valt att använda den tid klienten varit inskriven i projektet tills denne funnit en annan försörjande sysselsättning (och därmed lämnat projektet) som mått för hur väl projektet lyckats. Om klients deltagande är effektivt eller ej kan antas vara beroende av en eller flera individorienterade variabler. Dessa variabler har omsorgsfullt valts ut och konstruerats med utgångspunkter i både teori och erfarenheter projektets anställda samlat på sig.

Även om måttet för effektivitet är tid före utträde ur projekt, vilket är en kontinuerlig variabel, kommer analysen ske med en binär oberoende variabel. I arbetsmarknadsprojekt och på socialkontor används främst tidsintervall för att beskriva de faser klienterna befinner sig i. Den inledande fasen brukar räknas som cirka 6 månader. Har individen efter den tiden inte hamnat i sysselsättning betyder det ofta att mer omfattande insatser krävs. Klienter som deltagit i projektet mindre än 1 månad exkluderas, då dessa inte haft samma tidsmässiga förutsättningar att finna sysselsättning som övriga. Den del av klienterna som tillfälligt försvunnit ur projektet (beroende på inaktivitet eller att man inte deltagit i den bestämda tidsplanen) räknas här efter den tid de faktiskt varit inskrivna.

Tabell 1: Kodning av variabeln ”Effektiv (Benägenheten att finna sysselsättning inom 6 månader i projektet)”

	Tid i projektet			
Har funnit sysselsättning t.o.m. 31/9 2006	Mindre än en månad	1-6 månader	Mer än 6 månader	Totalt
Ja	Exkluderas (0/ej registrerade)	Effektiv (37)	Ej Effektiv (19)	**56**
Nej	Exkluderas (5)	Ej effektiv (5)	Ej effektiv (26)	**36**
Total	**5**	**42**	**45**	**92**

Materialet är dessutom relativt litet (88 observationer efter exkludering) och ger därför säkrare skattningar med en binär respons än med en kontinuerlig variabel.

Då även de flesta förklarande variablerna är antingen är binär- eller kategorivariabler, kommer en logistisk regressionsmodell användas. Det möjliggör en skattning av projektets effektivitet med grund i relevanta oberoende variabler, vilket kan visa förklaringen till varför vissa klienter stannar i projektet längre tid och har svårare att gå vidare till sysselsättning.

Det praktiska målet med uppsatsen är att finna några nyckelvariabler som visar vilka egenskaper och erfarenheter som gör att man effektivt lyckas vidarebefordra klienter till självförsörjning, och när man inte lyckas. Dessa variabler kan sedan användas som ett underlag för vidare studier och insatser inom området.

1.3 Bakgrund

Ungdomsarbetslösheten har under de senaste åren varit ett stort problem och genererat en allmän politisk diskussion om olika metoder för att bekämpa den. Bland ungdomar i åldern 25 eller yngre är arbetslösheten mer dubbelt så hög som för resten av befolkningen i arbetsför ålder (27,5 % april 2006, enligt SCB:s hemsida).

Enter är ett samverkansprojekt med arbetsmarknadspolitisk inriktning som delvis finansieras via EU:s strukturfonder, inom insatsområdet ”Ökad anställbarhet och företagaranda”. Med goda erfarenheter från tidigare, liknande projekt i stadsdelen har man valt att genomföra Enter som ett samarbete mellan stadsdelsförvaltning, försäkringskassa och arbetsförmedling i Härlanda stadsdel. Man har identifierat unga vuxnas svårigheter att ta sig in på arbetsmarknaden som ett viktigt förbättringsområde. Dessa problem, främst i fråga om studieval, yrkesval och vuxenlivet i allmänhet har i stadsdelen blivit allt mer påtagliga. Härlanda stadsdelsförvaltning initierade därför det treåriga projektet med syfte att ge dessa ungdomar tillgång till individuellt utformade åtgärder, inriktade mot arbete och studier. (Ek, Lind, 2005, s. 5-8)

Planen är att inom loppet av 3 år driva projektet för sammanlagt cirka 110 ungdomar i åldrarna 20-30, vilka på grund av funktionshinder eller nedsatt arbetsförhet/arbetsförmåga inte på egen hand kunnat genomföra en delaktighet i arbetsliv eller studier. Detta innefattar långtidsarbetslösa och personer med psykiska eller intellektuella funktionshinder. Målsättningen är att 65-70 % av de deltagande skall gå ifrån passivitet till aktivet, från en

bidragsunderstödd tillvaro till självförsörjning (Axelsson, Norman, 2005, s. 2-3, Växtkraft Mål 3, 2004, s. 2).

Ett indirekt mål är att genom projektet också knyta kontakter mellan stadsdelsförvaltning, försäkringskassa, arbetsförmedling, arbetsgivare och psykvården i stadsdelen. Även här syftar man till en mer resultatorienterad styrning. (Axelsson, Norman, 2005, s. 3)

Enter bygger sin verksamhet på ett problemlösande förhållningssätt, där man utgår från att klienterna själva är kapabla att ta ansvar för sin situation. Genom lösningsfokuserade samtal med kunniga coacher skall klienten långsamt få insikt om sina resurser och medvetandegöras om möjligheterna till att använda dem. Man försöker att genom en liten insats skapa en grund för människor att åstadkomma betydande förändringar i sin tillvaro. Lösningsfokus är idag en ofta använd metod inom myndigheter, kommuner och privata organisationer (Ek, Lind, 2005, s 11, Växtkraft Mål 3, 2004, s. 3).

Den primära verksamheten och kontakten med projektet klienterna har är samtal med coacherna som sker regelbundet; oftast åtminstone en gång i veckan. Vad som framkommit i samtalen, i kombination med information från arbetsträning eller praktik, ger coachen ett större sammanhang och gör det lättare att fokusera resurser. Vilka verktyg som sedan skall användas avgörs i samråd av coachen och deltagaren. Vidare förväntas man, utifrån individens möjligheter och kunskap, att genom praktik, introduktion och integrering på arbetsplatser kunna föra klienten till självförsörjning. En stor bredd av möjligheter skall tillhandahållas. (Ek, Lind, 2005, s 11-13,)

De flesta momenten i projekt Enters verksamhet är arbetsmarknadsrelaterade. Bland annat används intressetester för att guida osäkra klienter vidare till rätt områden. För anställningsfasen genomförs arbete med ansökningshandlingar, simulerande jobbinterjuver och arbetsträning, där deltagaren får vänja sig vid att passa tider jobba och samarbeta med andra och visa uthållighet. Handledningen sköts av kvalificerade jobbcoacher som skall ge deltagarna ökad erfarenhet och kompetens (ibid., s 12).

Arbetsplatsförlagd praktik är ett annat integrerande moment, som sker självständigt eller i kombination med övriga åtgärder. Klienten får då prova på ett nytt yrke och se om det passar

denne; alternativt kan praktiken ses om en slags provanställning där, om arbetsgivaren är nöjd med den anställde, jobb kan erbjudas (ibid., s 12).

Utöver direkt studievägledning ämnar man även inom ramen för projekt erbjuda olika vuxenutbildningar i samarbete med Folkuniversitetet i Göteborg. Det handlar då främst om mycket grundläggande utbildning; såsom datautbildning och truckförarkort.
Vidare anordnas återkommande förläsningar och seminarier, med fokus på arbetsmarknaden. (ibid., s 13).

Klienter

De fyra huvudsakliga uppdrag som Enter får från respektive instans är: stöd i jobbsökande, studievägledning, praktikplats eller rehabilitering mot arbetsmarknad. Målgruppen är långtidsarbetslösa och sjukskrivna ungdomar i ålder 18-30 år som på grund av funktionshinder eller nedsatt arbetsförmåga inte klarar integrering i arbetsliv och studier på egen hand. Större delen (cirka 75 %) av deltagarna är arbetslösa med försörjningsstöd, medan övriga är intellektuellt eller psykiskt funktionshindrade. Klienterna remitteras från respektive instans (socialkontor, arbetsförmedling eller försäkringskassa) som finner projektet lämpligt. Väl inskrivna görs sedan en bedömning av vilken åtgärd som passar individen bäst (Växtkraft Mål 3, 2004, s. 2).

Avgränsning och övrig forskning

Den bedömning av klienterna som sker innefattar endast den försörjningsorienterade process som sker inom ramen för projektet. Det innebär att studien utgår från att samtliga jobb- eller studierelaterade möjligheter klienterna fått under perioden är en effekt av Enters jobbcoachande. Antagandet är delvis felaktigt, då det ju finns en chans att sysselsättning kan genereras på annat sätt än genom samtalen med coacherna, men innebär ändå en nödvändig och verklighetsriktig förenkling. Flera studier visar att just långtidsarbetslösa har sämre chanser än övriga att finna ett jobb (Ek, Lind, 2005, s. 5-6).Studien tar därför inte heller hänsyn till någon form av kontrollgrupp, det vill säga: vad som hade hänt om klienterna inte vart inskrivna i Enter (Hemström - Martinson, 2002, s. 2-6,).

2. Data

Individdata som används har i samråd med projektansvarig sammanställts och kodats på ett lämpligt sätt. Då informationen om klienterna enligt personuppgiftslagen är konfidentiell har den i datamaterialet anonymiserats och ersatts med löpnummer**.** Materialet består dels av enkätsvar, vilka de projektintagna svarat på vid den första träffen; såsom ålder, kön och körkort. Dels har vi i Socialstyrelsens databaser funnit andra kvalitativa data, såsom längd på arbetslöshet, bidragsberoende och utbildningsnivå. Det finns också en rad variabler vilka antingen är bedömningar av jobbcoacherna eller information som vidarebefordrats från Arbetsförmedling eller Socialkontor. Detta innefattar uppdrag, boendeproblematik och eventuella psykiska problem. Därtill finns data för själva projektets verksamhet, såsom handlingsplan, in- och utskrivningsdatum och anledningen till utskrift. Det analyserade materialet består av 88 observationer och 10 variabler. Nedan redovisas samtliga inkluderade variabler:

Tabell 2: Inkluderade variabler

Variabel	**Typ av variabel**	**Möjliga utfall**	**Antal möjliga utfall**
P: Effektiv (Benägenheten att finna sysselsättning inom 6 månader i projektet)	Binär	1. Ja (6 månader eller mindre i projektet och funnit sysselsättning) 0. Nej (Mer än 6 månader i projektet)	2
β_1: Kön	Binär	Man, kvinna	2
β_2: Uppdrag till Enter	Kategorivariabel	1. Stöd i jobbsök/ studievägledning 2. Praktikplats 3. Rehabilitering mot arbetsmarknad	3
β_3: Arbetshinder	Binär	1. Ja (Psykiska problem, missbruk, dyslexi, (INF)) 0. Nej (försörjningsstöd/ej känt hinder).	2
β_4:Utbildning	Kategorivariabel (ordinalskala)	1. Högskoleexamen. 2. Gymnasieexamen. 3. Grundskola med slutbetyg. 4. Grundskola utan slutbetyg.	4
β_5: Ålder vid inskrivning	Kategorivariabel	1. 24 år eller yngre, 0. 25 år eller äldre	2
β_6:Bostadsproblematik	Binär	1. Ja 0. Nej	2
β_7: Körkort	Binär	1. Ja 0. Nej	2
β_8: Arbetslöshet innan projektets början (innan inträde i Enter)	Kategorivariabel (ordinalskala)	1. Mindre än 6 månader 2. 6-9 månader 3. 10 mån – 2 år 4. Över 2 år	4
β_9: Antal månader i bidragsberoende (innan inträde i Enter)	Kategorivariabel (ordinalskala)	1. Mindre än 6 månader 2. 6-9 månader 3. 10 mån – 2 år 4. Över 2 år	4

Förklaring av variabler

P: Effektiv (Benägenheten att finna sysselsättning inom 6 månader i projektet). Innehåller de binära utfall av effektiviteten, som den beskrivs i avsnitt 1.2.

För att förenkla den logistiska regressionsmodellen är ordinalskalade variabler kodade i samma ordning som utfallet på den beroende variabeln. Det innebär att det utfall som mest är troligt att bidra till sysselsättning rankas högst, näst mest jobbgenererande som tvåa, och så vidare. Rangordningen har skett i samarbete med projektansvarig, utifrån de erfarenheter man har från klienternas jobbsökande. Variabeln Uppdrag till Enter, som är en nominell kategorivariabel kodas i den logistiska regressionen som en rad av binära utfall.
De binära variablerna kodas dock alltid ”1” om individen har den angivna egenskapen.

β_1: *Kön* kodas här, utan någon speciell värdering, som 1 för man och 0 för kvinna.

β_2*: Uppdrag till Enter*. Klienterna klassificeras här utifrån närmst möjlighet till att få ett jobb. Det gör att den som endast behöver ett stöd i arbetssökande eller val av studier ligger bäst till. Praktikplats eller rehabilitering tar längre tid, i och med att klienten då genom dessa åtgärder först behöver införlivas i arbetslivet, innan det är möjligt att gå vidare mot en eventuell anställning eller utbildning.

β_3*: Arbetshinder* visar om det finns någon dokumenterad begränsning i klientens arbetsförmåga. Cirka 60 % av klienterna har inga kända hinder, och kan därför antas ha störst möjlighet att finna en sysselsättning. Arbetshinder kräver i olika grader specialiserade och anpassade anställningar och arbetsuppgifter, vilket tar lång tid att finna. Diagnosticerade psykiska och missbruks- problem påverkar chanserna att finna sysselsättning negativt. Deltagare med neuropsykologiska funktionshinder (INF), såsom dyslexi, har enligt projektets ansvariga sämst förutsättningar att få jobb. Dessa är dock väldigt få (endast 7 stycken) och därför grupperas samtliga arbetshinder tillsammans som en binär variabel.

β_4*:Utbildning.* Ju högre utbildning individen har, desto större är förutsättningarna att få ett jobb.

β_5*:Ålder vid inskrivning.* Variabeln visar de två primära grupperna, som använts för klassificering av ålder i projektet. Kodningen utgår från att personer under 25 år har lättare att finna sysselsättning än de som är äldre.

β_6*: Bostadsproblematik* En stor del (30 %) av klienterna har problem med boende. Antingen har man ingen hemadress, är tillfälligt inneboende eller bor fortfarande hos sina föräldrar. Dessa har därför sämre förutsättningar att förbättra sin situation än övriga.

β_7*:Körkort*. Kan antas främja chanserna att få jobb.

β_8*: Tid i arbetslöshet (innan inträde i Enter),* β_9*: Tid i bidragsberoende (innan inträde i Enter)*. Dessa variabler beskriver intervall av bidragsberoende och arbetslöshet generellt som används i projektet. Som tidigare nämnts används faser och intervall som begrepp i arbetsmarknadsprojekt, snarare än exakta tider. Båda variabler har kodats så att de utgår från att ju kortare tid en klient spenderat utan egen försörjning, desto troligare är det att denne kommer hitta sysselsättning.

Remitteringsinstans visar om klienten vidarebefordrats till projektet från antingen Försäkringskassan, Enheten för ekonomiskt bistånd eller Enheten för stöd till funktionshindrade. Variabeln exkluderas, då endast 7 av de i analysen inkluderade klienterna var remitterade från någon annan instans än Försäkringskassan.

3. Teori

3.1 Val av statistisk metod

Studien ämnar undersöka en rad olika variablers förhållande till ett binärt utfall; om vistelsen i projektet varit effektiv eller inte. De vanligaste statistiska metoderna för att testa detta är diskriminantanalys och logistisk regression. Valet av metod utgår från artikeln ”*Comparasion of Logistic Regression and Linear Discriminant analysis”* (Pohar et. al., 2004), vilken beskriver en rad simulationer som gjorts för att jämföra metoderna. I de flesta hänseenden är metoderna för detta material likvärdiga, men simulationerna visar att logistisk regression ger bättre prediktion än diskriminantanalysen, då man använder sig av relativt få kategorier i de oberoende variablerna. De 12 oberoende variabler i materialet har i genomsnitt 2.7 kategorier och artikeln visar, då antalet kategorier är 3, bättre resultat för i 78 % av fallen gentemot diskriminantanalysen. (Pohar et. al., 2004, s 154-156) Därför används logistisk regression, som förklaras närmare nedan:

3.2 Logistisk Regression

Logistisk regression möjliggör en skattning av variabel med ett diskret utfall. Den logistiska regressionen antar inte ett linjärt förhållande mellan den beroende och de oberoende variablerna och kan hantera icke-linjära effekter även när exponentiella och polynominala effekter inte är explicit inkluderade i ekvationen som oberoende variabler. En logistisk regression förutsätter inte någon speciell fördelning, men då den beroende variabeln P kan, då utfallet här är binärt, antas vara binominalfördelad.

P är alltså här sannolikheten för ett av två möjliga utfall för händelser, baserade på en icke-linjär funktion med den bästa kombinationen av prediktorer:

$$P = \frac{e^u}{1 + e^u} \qquad (3.1.1)$$

Där P är den skattade sannolikheten att fall *i (*1,2…,n) är i en av kategorierna och *u* är den linjära regressionsekvationen:

$$u = A + \beta_1 X_1 + \beta_2 X_2 \ldots \beta_k X_k \qquad (3.1.2)$$

med konstanten A, koefficienterna β, och prediktorerna X_j för *k* prediktorer (1,2,…,*k*).

Den linjära regressionsekvationen skapar logaritmer av oddsen, även kallade logits:

$$\ln\left(\frac{P}{1-P}\right) = \beta_0 + \sum \beta_j X_{ij} \qquad (3.1.3)$$

Det innebär alltså att den linjära regressionsekvationen är logaritmen av sannolikheten att ingå i en grupp dividerat med sannolikheten att ingå i den andra.

För att tolka logit-värdet kan det till en så kallad oddskvot genom att använda konverteringen e^{β_k}. Oddskvoten visar förändringen i oddset att tillhöra endera effektivtets-utfall för varje enhet den oberoende variabeln består av. Nedan syns oddskvoten mellan β_k =1 och β_k =0:

$$e^{\beta_k}(Oddskvot) = \frac{\dfrac{P(Effektiv|\beta_k = 1)}{P\left(Ej \quad Effektiv|\beta_k = 1\right)}}{\dfrac{P(Effektiv|\beta_k = 0)}{P\left(Ej \quad Effektiv|\beta_k = 0\right)}} \qquad (3.1.4)$$

Ett värde större än 1 indikerar en ökning av oddset att hamna i ”Effektiv” i förhållande till β_k =0, medan värden mindre än 1 innebär en reducerat odds. Värden mycket nära 1 visar att utfallet i prediktorn inte har någon större effekt på den beroende variabeln.

För att skatta den beroende variabeln i förhållande till en logit-variabel används vid logistisk regression en maximum likelihood-skattning. Skattningen utgår från oberoende kontinuerliga, diskreta eller kategoriserade variabler men där den beroende variabeln måste vara kategoriserad. Därigenom skattar den logistiska regressionen sannolikheten för att observationer ska klassificeras i respektive kategori . Målet med maximum likelihood-skattningen är att finna de bästa linjära kombinationerna av prediktorer för att förklara P.

Låt ett utfall i den beroende variabeln vara Y_i. Maximum likelihood maximerar då den logaritmerade likelihood funktionen LL enligt:

$$LL = \sum_{i=1}^{N} \left[Y_i \ln(\hat{Y}_i) + (1 - Y_i)\ln(1 - \hat{Y}_i) \right] \qquad (3.1.5)$$

vilket visar hur troligt det är att den beroende variabeln kan predikteras från de oberoende variablerna. Det sker genom en iterativ process, där maximum likelihood-algoritmen med

grund i residualer successivt estimerar koefficienterna tills konvergens nås. Det vill säga, det läge då LL inte signifikant förändras.

I den logistiska regressionsmodell som används här skattas endast den beroende variabelns effekt av interaktionen med de förklarande variablerna. Även om det finns möjlighet att skatta interagerande mellan prediktorerna är det här inte intressant, då målet är att finna en förenkling av sambanden, snarare än en matematiskt exakt skattning.

Signifikanstest

För att testa den logistiska regressionsmodellen används flera tester för att utvärdera såväl enskilda parametrar som den fullständiga modellen. Här används log likelihood-kvoten och Wald-test.

Log likelihoood-kvoten testar differensen mellan den genomförda modellen och en reducerad modell, med endast interceptet β_0. Om nollhypotesen, H_0, accepteras innebär att det inte finns någon signifikant skillnad modellerna emellan och att samtliga B_j är lika med 0. Om H_0 förkastas är den nya modellen signifikant bättre.

Kvoten mellan de två skattningarna logaritmeras och testas sedan enligt nedan:

$$2\left[LL(\beta_0, \beta_1, \ldots, \beta_k) - LL(\beta_0)\right] \qquad (3.1.6)$$

med fördelningen $\chi^2(k-1)$

Wald-testet används för att testa signifikansen för varje individuell koefficient (β_0, β_1, β_2 …, β_j). Det testar om effekten av en specifik logit är signifikant skild från 0 enligt:

$$\frac{\hat{\beta}_k^{\,2}}{Var(\beta_k)} \qquad (3.1.7)$$

med fördelningen $\chi^2(1)$

3.3 Genomförande

Syftet är att finna de mest relevanta variablerna och med hjälp av dessa finna den bästa förenklade modellen för att beskriva projekt Enters klienter. För att välja ut de variabler som skall ingå i den slutgiltiga modellen används framåt stegvis eliminering. Det tidigare nämnda

log likelihood-kvoten används för att bestämma vilka variabler som först skall läggas till den reducerade modellen($LL(\beta_0)$) och sedan vidare utökade modeller ($LL(\beta_0, \beta_j)$). För att undvika risken att exkludera variabler som ändå ger en god och relevant förklaring är kriteriet i den stegvisa elimineringen satt till signifikansnivån 30%.

Anpassningsförmågan hos en logistisk regressionsmodell utvärderas ofta med Hosmer-Lemeshows test. Proceduren delar upp datamaterialet i 10 segment och jämför sedan de faktiska värdena med dem som räknats fram i modellen. Tabellen testas sedan statistiskt med ett χ^2-test där ett signifikant värde (>0,05) innebär att det finns en så stor skillnad i materialet att modellen har en tillräckligt god anpassning. Detta material har dock för få observationer som krävs för att genomföra testet korrekt, då χ^2-tester bland annat kräver åtminstone 5 observationer i varje cell. Hosmer-Lemeshows test kommer ändå användas här som en indikator på anpassning, snarare än ett exakt test.

Dessa resultat jämförs sedan med den fullständiga modellen, Wald-tester och oddskvoter för att formulera en slutlig modell.

4. Resultat

Inledningsvis körs en logistisk regression med samtliga variabler. Resultatet syns nedan:

Tabell 3: Logistik regression med samtliga variabler

Parameter	*Koefficient*	*Medelfel*	*Wald-test* χ^2	*Wald-test p-värde*
Intercept	-4,401	1,15	14,71	0,0001
Kön	0,323	0,33	0,95	0,3301
Uppdrag (Rehabilitering)	0,007	0,65	0,00	0,9913
Uppdrag (Praktikplats)	-0,157	0,89	0,03	0,8605
Arbetshinder (Ja)	-0,951	0,42	5,03	0,0249
Bostadsproblematik (Ja)	-0,567	0,42	1,80	0,1802
Körkort (Ja)	-0,725	0,58	1,55	0,2130
Utbildning	-0,303	0,50	0,36	0,5464
Bidragsberoende	0,877	0,35	6,34	0,0118
Åldersgrupp (Över 25 år)	-0,134	0,92	0,02	0,8841
Arbetslöshet	1,065	0,46	5,33	0,0210

Tabell 4: Oddskvoter: logistik regression med samtliga variabler

Parameter	*Oddskvot*	*95 %-igt Wald konfidensintervall*	
Kön	1,910	0,519	7,020
Uppdrag (Rehabilitering)	0,867	0,067	11,188
Uppdrag (Praktikplats)	0,736	0,028	19,720
Arbetshinder (Ja)	0,149	0,028	0,787
Bostadsproblematik (Ja)	0,322	0,061	1,689
Körkort (Ja)	0,235	0,024	2,298
Utbildning	0,739	0,276	1,975
Tid i bidragsberoende	2,405	1,215	4,760
Åldersgrupp (Över 25 år)	0,874	0,144	5,319
Arbetslöshet	2,902	1,175	7,167

Log likelihood-testet visar att modellen ej är skild från 0 (λ= 52,5835, fg.=10, p.=0,0001). Flera av variablerna får riktigt dåliga Wald-värden, och det är tydligt att dessa behöver rensas ur modellen. På 5-% nivån ges signifikanta värden endast för ”bidragsberoende”, ”arbetshinder” och ”arbetslöshet”. Oddskvoterna är närheten av 1 för ”uppdrag”, ”utbildning” och ”åldersgrupp”, något som indikerar att dessa har en låg effekt på utfallen i P. För att definiera de mest relevanta variablerna används en stegvis eliminering:

Tabell 5: Logistik regression efter eliminering

Parameter	*Koefficient*	*Medelfel*	*Wald* χ^2	*Wald-test p-värde*
Intercept	-4,958	1,193	17,273	0,000
Arbetshinder	-0,987	0,403	6,005	0,014
Körkort	-0,780	0,409	3,644	0,056
Bostadsproblematik	-0,787	0,648	1,476	0,224
Tid i bidragsberoende	0,988	0,336	8,671	0,003
Arbetslöshet	1,028	0,442	5,417	0,020

Tabell 6: Oddskvoter: Logistik regression efter eliminering

Parameter	Oddskvot	95 %-igt Wald konfidensintervall	
Arbetshinder	0,139	0,029	0,674
Bostadsproblematik	0,210	0,042	1,043
Körkort	0,207	0,016	2,626
Tid i bidragsberoende	2,687	1,392	5,187
Arbetslöshet	2,796	1,176	6,647

Den stegvisa elemineringsprocessen lägger till 4 variabler, vilka signifikant förbättrar modellen. Det visar sig vara samma variabler som tidigare fått goda Wald-värden; ”arbetshinder”, ”arbetslöshet” och ”tid i bidragsberoende” samt även ”bostadsproblematik” och ”körkort”. Oddskvoterna visar stora positiva effekter för varje tidsintervall hos bidragsberoende och arbetslöshet. Bostadsproblematik och körkort får båda starkt negativa effekter, vilket är oväntat för den senare, som ju antogs ha en positiv inverkan.

Då koefficienterna och även oddskvoterna för arbetslöshet och bidragsberoende i princip är identiska (1,028 och 0,988 respektive 2,796 och 2,687) går det att misstänka multikolinjäritet. Variablerna beskriver överlappande stadier; personer som är arbetslösa är ofta bidragsberoende, åtminstone då man varit det i längre perioder är det nödvändigt för försörjning. Materialet visar sig innehålla individer i motsvarande grupp (klienter med exempelvis både bidragsberoende och arbetslöshet mindre än 6 månader) till 44 %. Andelen som befunnit sig något längre i arbetslöshet än bidragsberoende (exempel: bidragsberoende under 6 månader, arbetslös 6-10 månader) uppgår till 46%.

För att undvika multikolinjäritet och få en korrekt modell behöver därmed en av variabler tas bort. Arbetslöshet exkluderas därför ur modellen, då tid i bidragsberoende kan ses som en kausal effekt av bland annat arbetslöshet. Skillnaden är att bidragsberoende också innefattar även de fall av försörjningsstöd som inte är registrerade som arbetslöshet.

Den stegvisa processen görs alltså nu om utan ”Arbetslöshet”:

Tabell 7: Slutlig logistisk regressionsmodell

Parameter	*Koefficient*	*Medelfel*	*Wald* χ^2	*Wald-test p-värde*
Intercept	-3,694	0,789	21,930	0,000
Arbetshinder	-1,232	0,389	10,045	0,002
Bostadsproblematik	-0,920	0,391	5,542	0,019
Bidragsberoende	1,268	0,307	17,073	0,000

”Tid i bidragsberoende”, ”arbetshinder” och ”bostadsproblematik” får nu skattas nu ha större effekter, medan Körkort exkluderas. Samtliga variabler är signifikanta enligt Wald-test och ingen multikolinjäritet tycks nu förekomma. Hosmer-Lemeshows test visar ett icke-signifikant värde (p=0,0686) för modellen, vilket den även gjort för de båda föregående (p=0,3029 och p=0,1922). Inte oväntat är anpassningsförmågan sämre än tidigare modeller, då antalet variabler reducerats, men då det är en förenklad modell som söks indikerar resultaten den som fullgod. Denna slutliga modell skrivs enligt nedan:

$$P = \frac{e^{-3,694-1,232X_3-0,920X_6+1,268X_9}}{1+e^{-3,694-1,232X_3-0,920X_6+1,268X_9}}$$

Tabell 8: Oddskvoter för slutlig logistisk regressionsmodell

Parameter	Oddskvot	95 %-igt Wald konfidensintervall	
Arbetshinder (β_3)	0,085	0,019	0,391
Bostadsproblematik (β_6)	0,159	0,034	0,735
Bidragsberoende (β_9)	3,553	1,947	6,483

Tabell 9: Skattningar med slutlig logistisk regressionsmodell

Tid i bidragsberoende (X_9)	Bostadsproblematik (X_6)	Arbetshinder (X_3)	Effektiv (Benägenheten att finna sysselsättning inom 6 månader i projektet) (P)
Under 6 månader	Nej	Nej	0,527
6-10 månader	Nej	Nej	0,239
10 månader – 2 år	Nej	Nej	0,081
Över 2 år	Nej	Nej	0,024
Under 6 månader	Nej	Ja	0,246
6-10 månader	Nej	Ja	0,084
Under 6 månader	Ja	Nej	0,308
6-10 månader	Ja	Nej	0,111
10 månader – 2 år	Ja	Nej	0,034

5. Sammanfattning

Under de premisser data samlats in går det att urskilja tre signifikanta variabler som avgör huruvida personen kommer att finna arbete inom sex månader efter inträde i projekt Enter. Den slutgiltiga modellen innehåller variablerna arbetshinder, tid i bidragsberoende och bostadsproblem, vilka har en signifikant effekt på utfallet.

Arbetshinder ger oddskvoten 0,0805, vilket innebär att en person med exempelvis psykiska problem har ungefär en tolftedel av chansen att uppnå effektivitet i projektet jämfört med en klient utan några arbetshinder. Skattningen i tabell 9 visar att arbetshinder i med en kort tid i bidragsberoende endast ger 24,6 procents sannolikhet att lyckas hitta jobb inom 6 månader inom ramarna för Enter och en obefintlig chans (mindre än 1 %) vid längre bidragsberoende. Att variabeln är signifikant kan också tolkas som en verifiering på att den grupperingen av olika arbetshinder som gjorts är riktig och att dessa hinder faktiskt ger skillnad i anställbarhet.

Bostadsproblematik har en mindre effekt än arbetshinder, men förekomsten påverkar ändå resultatet tydligt; oddskvoten visar att en person med boendeproblem har en sjättedel så stor chans att klara målsättningarna som övriga deltagande. Detta har även av projektanställda tidigare påtalats som ett viktigt och avgörande problem. 30,8 % av klienterna med bostadsproblematik förväntas uppnå effektivitet, då dessa är utan arbetshinder och har haft en kort tid i bidragsberoende.

De utfall som i dessa variabler ger någon effekt innefattar en varsin grupp som representerar cirka 40 procent av materialet. Om projektdeltagare utan arbetshinder eller bostadsproblematik lyckas uppnå ”effektivitet” förklaras det i slutliga modellen endast med den tid de varit i bidragsberoende innan projektets start. Sannolikheten för dessa individer att hitta ett jobb är alltså större än de övriga, men avtar successivt med den tid som tillbringats i bidragsberoende. En person med mindre än 6 månaders bidragsberoende skattas enligt modellen ha 52,7 % chans att uppnå effektivitet inom projektet, medan grupperna ”10 månader – 2 år” och ”mer än 2 år” skattas till att mindre än 1 % chans (se tabell 9). Teorin om att lång tid utanför arbetsmarknaden gör det än svårare att hitta sysselsättning verifieras alltså här.

Även om inga nya samband åskådliggjorts går det att tydligt visa vilka variabler som har störst relevans. Den slutgiltiga logistiska regressionsmodellen visar de största effekterna och en förenkling av viktiga erfarenheter som bör tas i beaktande, då en person kommer till projektet. Modellen visar var framtida stödåtgärder kan tänkas ligga och förhoppningsvis bidrar uppsatsen på så sätt till en diskussion om Enter och andra liknande projekts klienter och insatser.

6. Referenslista

Tryckta källor

Hemström M, Martinson, S, 2002. *Att följa upp och utvärdera arbetsmarknadspolitiska program.* Institutet för arbetsmarknadspolitisk utvärdering: Uppsala.

Barbara G. Tabachnick, 2007, *Using Multivariate Statistics.* Allyn & Bacon: Boston.

Christensen, Ronald, 1997, *Log-linear models and logistic regression.* Springer: New York.

Pohar, M, Blas, M, Turk, S, 2004. *Comparison of Logistic Regression and Linear Discriminant Analysis: A Simulation Study.* Metodoloski zveski.

Offentligt tryck

Ek, E, Lind, B, 2005. *Enter – Delrapport 1: Perioden 2004-09 – 2005-09. Göteborgs stad: Göteborg.*

Bihari Axelsson, S, Norman, C, 2005. *Delutvärdering av EU-projekt Enter: Perioden 2005-01-01 – 2005-09-30.* Nordiska högskolan för folkhälsovetenskap: Göteborg.

Ursprunglig projektbeskrivning, 2004. *Växtkraft Mål 3. Göteborgs Stad (Härlanda)*

7. Appendix

Logistisk regression med samtliga variabler inkluderade

```
The SAS System                                                    17:03 Friday, March 16, 2007
101

The LOGISTIC Procedure

                     Model Information

Data Set                      WORK.ENTER
Response Variable             effektiv             effektiv
Number of Response Levels     2
Model                         binary logit
Optimization Technique        Fisher's scoring


Number of Observations Read          88
Number of Observations Used          87


          Response Profile

 Ordered                      Total
   Value     effektiv     Frequency

       1            1            37
       2            0            50

Probability modeled is effektiv=1.

NOTE: 1 observation was deleted due to missing values for the response or explanatory
variables.


      Class Level Information

                                Design
Class            Value     Variables

kon              1          1
                 0         -1

uppdrag          2          1      0
                 1          0      1
                 0         -1     -1

Arbetshinder     1          1
                 0         -1

Bostadsprob      1          1
                 0         -1

korkort          1          1
                 0         -1


The SAS System                                                    17:03 Friday, March 16, 2007
102

The LOGISTIC Procedure

                     Model Convergence Status

         Convergence criterion (GCONV=1E-8) satisfied.


         Model Fit Statistics

                              Intercept
              Intercept            and
Criterion          Only     Covariates

AIC             120.658         88.074
SC              123.124        115.199
-2 Log L        118.658         66.074
```

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	52.5835	10	<.0001
Score	42.3019	10	<.0001
Wald	25.3147	10	0.0048

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
kon	1	0.9484	0.3301
uppdrag	2	0.0335	0.9834
Arbetshinder	1	5.0284	0.0249
Bostadsprob	1	1.7962	0.1802
korkort	1	1.5506	0.2130
utbildning	1	0.3639	0.5464
bidrag	1	6.3442	0.0118
aldersgrupp	1	0.0213	0.8841
arbetsloshet	1	5.3303	0.0210

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.4013	1.1475	14.7117	0.0001
kon 1	1	0.3234	0.3321	0.9484	0.3301
uppdrag 2	1	0.00710	0.6498	0.0001	0.9913
uppdrag 1	1	-0.1565	0.8903	0.0309	0.8605
Arbetshinder 1	1	-0.9506	0.4239	5.0284	0.0249
Bostadsprob 1	1	-0.5668	0.4229	1.7962	0.1802
korkort 1	1	-0.7248	0.5820	1.5506	0.2130
utbildning	1	-0.3027	0.5017	0.3639	0.5464
bidrag	1	0.8775	0.3484	6.3442	0.0118
aldersgrupp	1	-0.1343	0.9213	0.0213	0.8841
arbetsloshet	1	1.0652	0.4614	5.3303	0.0210

Odds Ratio Estimates

Effect		Point Estimate	95% Wald Confidence Limits	
kon	1 vs 0	1.910	0.519	7.020
uppdrag	2 vs 0	0.867	0.067	11.188
uppdrag	1 vs 0	0.736	0.028	19.720
Arbetshinder	1 vs 0	0.149	0.028	0.787
Bostadsprob	1 vs 0	0.322	0.061	1.689
korkort	1 vs 0	0.235	0.024	2.298
utbildning		0.739	0.276	1.975
bidrag		2.405	1.215	4.760
aldersgrupp		0.874	0.144	5.319
arbetsloshet		2.902	1.175	7.167

Association of Predicted Probabilities and Observed Responses

Percent Concordant	90.6	Somers' D	0.815
Percent Discordant	9.1	Gamma	0.818
Percent Tied	0.3	Tau-a	0.403
Pairs	1850	c	0.908

Partition for the Hosmer and Lemeshow Test

Group	Total	effektiv = 1 Observed	effektiv = 1 Expected	effektiv = 0 Observed	effektiv = 0 Expected
1	9	1	0.12	8	8.88
2	9	0	0.42	9	8.58
3	9	0	0.70	9	8.30
4	9	1	1.29	8	7.71
5	9	3	2.57	6	6.43
6	9	5	4.28	4	4.72
7	9	5	6.23	4	2.77
8	9	8	7.35	1	1.65
9	9	8	8.27	1	0.73
10	6	6	5.77	0	0.23

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
9.4870	8	0.3029

Stegvis logistisk regression 1

The SAS System 17:03 Friday, March 16, 2007
94

The LOGISTIC Procedure

NOTE: No (additional) effects met the 0.3 significance level for entry into the model.

Summary of Forward Selection

Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq	Variable Label
1	arbetsloshet	1	1	27.6259	<.0001	arbetsloshet
2	bidrag	1	2	9.3893	0.0022	bidrag
3	Arbetshinder	1	3	7.5920	0.0059	Arbetshinder
4	Bostadsprob	1	4	3.4897	0.0618	Bostadsprob
5	korkort	1	5	1.4865	0.2228	korkort

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Arbetshinder	1	6.0050	0.0143
Bostadsprob	1	3.6438	0.0563
korkort	1	1.4763	0.2243
bidrag	1	8.6711	0.0032
arbetsloshet	1	5.4170	0.0199

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-4.9579	1.1929	17.2726	<.0001
Arbetshinder 1	1	-0.9867	0.4027	6.0050	0.0143
Bostadsprob 1	1	-0.7800	0.4086	3.6438	0.0563
korkort 1	1	-0.7874	0.6481	1.4763	0.2243
bidrag	1	0.9883	0.3356	8.6711	0.0032
arbetsloshet	1	1.0283	0.4418	5.4170	0.0199

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Arbetshinder 1 vs 0	0.139	0.029	0.674
Bostadsprob 1 vs 0	0.210	0.042	1.043
korkort 1 vs 0	0.207	0.016	2.626
bidrag	2.687	1.392	5.187
arbetsloshet	2.796	1.176	6.647

Association of Predicted Probabilities and Observed Responses

Percent Concordant	90.5	Somers' D	0.823
Percent Discordant	8.2	Gamma	0.834
Percent Tied	1.2	Tau-a	0.407
Pairs	1850	c	0.912

Partition for the Hosmer and Lemeshow Test

		effektiv = 1		effektiv = 0	
Group	Total	Observed	Expected	Observed	Expected
1	8	1	0.10	7	7.90
2	9	0	0.43	9	8.57
3	9	0	0.69	9	8.31
4	9	1	1.17	8	7.83
5	8	2	2.48	6	5.52
6	10	4	4.27	6	5.73
7	9	7	5.72	2	3.28
8	9	7	7.22	2	1.78
9	9	8	8.19	1	0.81
10	7	7	6.74	0	0.26

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
11.1702	8	0.1922

Stegvis logistisk regression 2

The SAS System 17:03 Friday, March 16, 2007
99

The LOGISTIC Procedure

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	120.658	79.788
SC	123.124	89.651
-2 Log L	118.658	71.788

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	46.8699	3	<.0001
Score	37.9317	3	<.0001
Wald	24.2103	3	<.0001

Residual Chi-Square Test

Chi-Square	DF	Pr > ChiSq
2.6574	6	0.8504

NOTE: No (additional) effects met the 0.3 significance level for entry into the model.

Summary of Forward Selection

Step	Effect Entered	DF	Number In	Score Chi-Square	Pr > ChiSq	Variable Label
1	bidrag	1	1	25.6426	<.0001	bidrag
2	Arbetshinder	1	2	11.3120	0.0008	Arbetshinder
3	Bostadsprob	1	3	6.0591	0.0138	Bostadsprob

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Arbetshinder	1	10.0452	0.0015
Bostadsprob	1	5.5416	0.0186
bidrag	1	17.0729	<.0001

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-3.6938	0.7888	21.9301	<.0001
Arbetshinder 1	1	-1.2316	0.3886	10.0452	0.0015
Bostadsprob 1	1	-0.9205	0.3910	5.5416	0.0186
bidrag	1	1.2678	0.3068	17.0729	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
Arbetshinder 1 vs 0	0.085	0.019	0.391
Bostadsprob 1 vs 0	0.159	0.034	0.735
bidrag	3.553	1.947	6.483

Association of Predicted Probabilities and Observed Responses

Percent Concordant	87.0	Somers' D	0.777
Percent Discordant	9.4	Gamma	0.806
Percent Tied	3.6	Tau-a	0.384
Pairs	1850	c	0.888

Partition for the Hosmer and Lemeshow Test

Group	Total	effektiv = 1 Observed	effektiv = 1 Expected	effektiv = 0 Observed	effektiv = 0 Expected
1	9	1	0.18	8	8.82
2	11	0	0.67	11	10.33
3	13	4	2.11	9	10.89
4	8	0	2.00	8	6.00
5	11	3	4.54	8	6.46
6	6	4	3.25	2	2.75
7	8	5	5.61	3	2.39
8	21	20	18.63	1	2.37

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	DF	Pr > ChiSq
11.7164	6	0.0686