



Lunds universitet  
Statistiska institutionen

# Effekter av kön, ålder och region på sjukpenningen i Sverige

-en variansanalys

Rikke Berner

Uppsats i statistik  
10 poäng  
Nivå 61-80 poäng  
Oktober 2006

Handledare: Mats Hagnell

## Abstract

According to TCO (The Swedish Central Organisation for Public Servants) the worst threat to the Swedish welfare is that so many people because of bad health and unemployment are not a part of the labour force. If Sweden could make people with health and unemployment problems half the size as today, the GDP would increase with more than five percent. In real money this means 110 billion Swedish kronor a year. According to SCB (the Swedish Central Bureau of Statistics) the difference between women and men's absence owing to illness should be investigated more closely in order to get a better understanding of the problem. A good way to study the absence owing to illness is to analyse the share of people with sick-benefit.

The purpose of this essay is to analyse whether the sick-benefit differs due to the variation in the independent factors (age, gender, region and time). Furthermore, the interaction, if any, between the analysed factors will be studied and interpreted. To get a better understanding of how much every factor and interaction explains the differences in sick-benefit, the strength of the effect (eta square) is calculated. This measure together with significant differences is far more informative than just significant differences. This allows us to eliminate significant factors from our model because of the small contribution to the variation in the dependent factor.

We have chosen to call the dependent factor "*Proportion of the population with sick-benefit*" and the independent factors in this essay are age, gender, region and time. The age-factor is divided into nine different groups, the region-factor is divided into three groups and the years included in this study are 2000-2005. The reason why a time factor is chosen for analyse is to study whether the absent owing to illness has changed over time and if so, in what way.

A four-way analysis of variance with fixed effects is the best way to examine the differences in "*Proportion of the population with sick-benefit*". We start out by eliminating the insignificant interactions and then according to interaction plots and the size of eta squared we reduce the model even more. In the final model we find significant interactions between gender and age, between gender and region and between region and age. There is a tendency for women in the northern parts of Sweden to have a higher proportion of sick-benefit. If we look at the interaction between gender and age it shows that the difference between men and women is not the same for those in the age group 20-24 years and 60-64 years, as for those in the middle age groups. The differences in the middle age groups are greater and the biggest difference is seen in the age-group 35-39 years where the proportion with sick-benefit is twice as big for women. The interaction between age and region indicates that people between 35 and 39 years, in the northern parts of Sweden, receive sick-benefits to a larger extent than those between 40 and 44 years. This is not the case in the rest of Sweden.

Tendencies were shown for the "*Proportion of the population with sick-benefit*" to decrease after year 2002. This indicates that the Swedish goal of cutting the proportion of people absent owing to illness in half is going in the right direction.

# Innehållsförteckning

<i>Abstract</i> .....	2
<i>Innehållsförteckning</i> .....	3
<i>1 Inledning</i> .....	4
1.1 Bakgrund .....	4
1.2 Syfte.....	5
<i>2 Data</i> .....	6
2.1 Undersökningens datamaterial.....	6
2.2 Databehandling .....	6
2.3 Andra tänkbara variabler .....	7
<i>3 Metod</i> .....	8
3.1 Variansanalys med obalanserad data.....	8
3.2 Tolkning av samspel.....	9
3.3 Styrkan av effekterna .....	10
<i>4 Resultat</i> .....	15
4.1 Översikt över den beroende variabeln .....	15
4.2 Variansanalys .....	15
4.3 Samspelseffekter.....	17
4.4 Sammanfattande resultat .....	22
<i>5 Slutdiskussion</i> .....	25
<i>Sammanfattning</i> .....	26
<i>Litteraturförteckning</i> .....	27
<i>Bilaga 1</i> .....	28
Sveriges län .....	28
Åldersgrupper .....	28
<i>Bilaga 2 "Nya" modellen</i> .....	29
SAS-kommandon .....	29
Resultat .....	29
<i>Bilaga 3 Slutgiltiga modellen</i> .....	33
SAS-kommandon .....	33
Resultat .....	33

# 1 Inledning

Enligt Tjänstemännens Centralorganisation, TCO, (2003) finns det i dagsläget drygt 5,2 miljoner personer i så kallad arbetsför ålder i Sverige. Med detta menas personer som är mellan 20 och 64 år gamla. Av dessa går endast tre miljoner till jobbet en vanlig dag. Det finns naturligtvis många anledningar till att ungefär 40 procent av arbetskraften befinner sig på annat håll, och några av orsakerna är "bra" orsaker. Exempel på "bra" egenskaper är studier, barnledighet och semester. "Bra" egenskaper är de därför att de i längden höjer vår välfärd. Det finns däremot alldeles för många fall av "dåliga" orsaker till att man inte befinner sig i arbetskraften. Sådana orsaker är sjukdom och arbetslöshet. Förutom att dessa "dåliga" egenskaper leder till personligt lidande så är de det största hotet mot den svenska välfärden. I samhällsdebatten brukar det annars hävdas att det är en åldrande befolkning i kombination med för få födda som utgör det värsta hotet mot vår välfärd. Men enligt TCO är detta alltså inte sant. Det är med andra ord viktigt att vi får ner dagens höga sjukskrivnings- och förtidspensioneringstal. Tidigare undersökningar har visat på att en halvering av ohälsotalen skulle innebära att bruttonationalprodukten stiger med mer än fem procent. Detta skulle alltså betyda att en halv miljon fler människor arbetar och i reda pengar 110 miljarder kronor om året.

I början av 1990-talet ökade arbetslösheten fyrfaldigt i Sverige. Så gott som var tionde person förlorade sitt arbete och det dröjde ett par år innan upp emot hälften av dessa människor hittat tillbaka in på arbetsmarknaden. Denna omfattande rörelse in, ut och inom arbetsmarknaden påverkade inte bara den svenska befolkningens försörjningssituation utan även arbetsplatserna anpassade sig till den nya situationen. Arbetskraven ökade och mycket av den gamla tryggheten gick förlorad. Troligtvis har dessa miljöombyten på arbetsplatserna mycket att göra med att folk sjukskriver sig mer och till följd har vi i Sverige under de senaste åren sett en påfallande ökning av Sveriges ohälsotal. (SCB 2004:3).

Enligt Statistiska Centralbyrån, SCB, (2004:3) har de senaste årens ökade ohälsotal fört med sig en omfattande debatt, till största del till följd av de ökade kostnader ohälsotalen fört med sig. I början av år 2003 påbörjade SCB ett stort projekt vars ändamål var att samordna statistiken som beskriver sjukfrånvaro och ohälsa i Sverige. I projektets sammanfattande rapport (SCB 2004:3) beskriver man förhållanden som har analyserats men som bör studeras mera ingående. Ett av dessa förhållanden är skillnaden mellan männens och kvinnornas sjukskrivningar. Generellt har man sett att andelen sjukskrivningar är högre bland kvinnor än män. Detta gäller oavsett i vilken näringsgren eller sektor de arbetar i och oavsett vilken ålder de har. Fokus bör enligt SCB läggas på förklaringar av skillnader och likheter mellan kvinnors och mäns sjukskrivningar.

## 1.1 Bakgrund

Enligt försäkringskassan (2005) har utbetalningarna från socialförsäkringen ökat sedan slutet av 1990-talet. Detta beror till största del på att fler och fler människor är sjukskrivna och att sjukskrivningsperioderna tenderar att bli allt längre. En person som är sjukskriven får sjukpenning eller rehabiliteringspenning under hela veckan eller del av veckan, beroende på hur mycket han eller hon bedöms orkar arbeta samt hur mycket han eller hon arbetar i vanliga fall. Sjukpenningen och rehabiliteringspenningen ersätter alltså en del av inkomstförlusten orsakat av sjukdom. Av utbetalningarna från sjukförsäkringen, med andra ord utbetalningarna av sjukpenning, rehabiliteringspenning och närståendepenning, utgjorde sjukpenningen 83 och 81 procent under åren 2003 respektive 2004. Rehabiliteringspenningen utgjorde däremot "bara" 7

respektive 6 procent av de totala utbetalningarna samma år. En bra variabel att analysera sjukskrivningarna efter bör alltså vara sjukpenningen.

## **1.2 Syfte**

När det kommer till skillnader mellan andelen män och kvinnor som får sjukpenning har man med avseende på tidigare undersökningar redan en klar bild av att kvinnor i större utsträckning tilldelas denna bidragstyp. Syftet med undersökningen i denna uppsats är att med hjälp av statistiska metoder se efter vart dessa skillnader finns och undersöka om det finns samspel med andra faktorer. Då Sverige sedan år 2002 haft tydliga mål att halvera ohälsotalen, ska vi även undersöka om andelen som får sjukpenning har ökat eller minskat med tiden. För att få en bättre förståelse för hur stor betydelse varje analyserad faktor har för skillnaden i tilldelad sjukpenning beräknas styrkan av effekten. Detta mått, tillsammans med signifikanta skillnader, ger oss en mera informativ bild av hur stor vikt man ska lägga på de signifikanta skillnaderna.

### **1.2.1 Avgränsningar**

Studien begränsas till att studera de variabler som försäkringskassan har offentliggjord på sin hemsida med hänsyn till sjukpenning. Tidpunkten för sjukpenningutbetalningarna är år 2000 till och med 2005. Undersökningen görs med hjälp av variansanalys.

## 2 Data

I detta avsnitt presenteras datamaterialet som undersökningen bygger på samt hur materialet har behandlats. Avsnittet avslutas med en diskussion om andra tänkbara variabler.

### 2.1 Undersökningens datamaterial

Undersökningen är baserad på Försäkringskassans statistik över utbetalda sjukpenningar från år 2000 till och med år 2005. Statistiken över sjukpenningar beskrivs i Försäkringskassans databas som ”*antal personer som har fått sjukpenning*” under ett år. Datan är i sin tur indelad efter region, ålder och kön. Regionerna utgörs av Sveriges 21 olika län (se bilaga 1). Åldersvariabeln är indelad i olika åldersgrupper. Den första åldersgruppen är 20-24 år och sista 60-64 år. Däremellan finns alltså sju åldersgrupper, alla med samma klassbredd (se bilaga 1).

För att få ett relativt mått på antal utbetalda sjukpenningar, som gör det möjligt att jämföra regionerna, beräknas ett mått som vi valt att kalla ”*Andel av befolkningen med sjukpenning*” fram. ”*Andel av befolkningen med sjukpenning*” fås genom att ta ”*antal personer som har fått sjukpenning*” under ett år och dividera med 12. Vi har nu ett genomsnittligt värde på antal utbetalda sjukpenningar per månad. Denna siffra divideras i sin tur med det totala antalet invånare för tillhörande kön, åldersgrupp och län. Till exempel så hade kvinnor i åldern 20-24 år, under år 2000, i Stockolms län, totalt 10984 stycken utbetalda sjukpenningar. Detta blir igenomsnitt 915,33 per månad. Antalet divideras med 54446, som alltså är det totala antalet kvinnor i åldern 20-24 som var bosatta i Stockholms län under år 2000. Slutligen får vi alltså fram att i genomsnitt 1,68 procent av kvinnorna fick sjukpenning varje månad. Samma beräkningar görs för alla län, åldersgrupper, kön och årtal. Sammanlagd blir det 2268 observationer.

### 2.2 Databehandling

För att datan ska kunna analyseras är det nödvändigt att koda om vissa variabler. I följande delavsnitt tar vi upp hur de olika förklarande variablerna på bästa sätt transformeras.

#### 2.2.1 Kön

Kön är en så kallad kvalitativ variabel. Detta innebär alltså att kön är en variabel med icke-numeriska värden. Oftast när man analyserar kön kodar man om variabeln till en 0/1-variabel. Detta betyder alltså att värdet på de individer som är av kvinnligt kön blir 1 och värdet på de individer som är av manligt kön kodas till 0.

#### 2.2.2 Region

Variabeln region har 21 olika nivåer (län). Inkluderas alla i analysen skulle vi troligtvis få problem med kollinearitet. Vi blir alltså tvungna att definiera om variabeln region på ett sådant sätt som ger oss mindre nivåer (regioner) att jämföra. SCB delar upp Sveriges 21 län i tre olika länsregioner (se bilaga 1). Använder vi oss i stället av dessa får vi följande variabler;

Länsregion 1 = 1

Länsregion 2 = 2

Länsregion 3 = 3

En annan anledning som stödjer vår nya regionsindelning är att de slutgiltiga resultaten blir mera lättolkade. Det är lättare att dra godtagbara slutsatser från en jämförelse av tre olika grupper än med 21 olika grupper.

### **2.2.3 Ålder**

Ålder är en så kallad ordinalskala. Det finns med andra ord en naturlig rangordning mellan nivåerna. Den första nivån, 20-24 år, är minst (yngst) och den sista nivån, 60-64 år, är störst (äldst). Vi kan välja en kodning där 1 = 20-24 år, 2 = 25-29 år och så vidare. En fördel med denna numeriska omkodning är att nivåerna kommer i rätt ordning i tabeller och diagram.

### **2.2.4 Årtal**

Eftersom årtal är en kvantitativ variabel behöver denna alltså inte kodas om. Vi har sex olika årtal, år 2000, 2001, 2002, 2003, 2004 och 2005.

## **2.3 Andra tänkbara variabler**

Syftet med detta delavsnitt är att ge läsaren tips om andra, mera ingående, intressanta analyser.

Begränsar man sig inte till att använda de variabler som Försäkringskassan offentliggjort på sin hemsida, beträffande sjukpenning, finns det naturligtvis en mängd andra intressanta variabler att titta på. Eftersom det inte var möjligt att hitta relevant data till faktorerna, som nämns nedan, var det däremot inte möjligt att genomföra godtagbara analyser.

Några för oss intressanta exempel vore storleken på företaget individen arbetar i, vilken sektor man är verksam i, vilka familjeförhållanden man har samt vilka motions-, mat- och rökvanor individen har. En annan intressant infallsvinkel vore att undersöka om det finns genetiska samband mellan sjukskrivningarna.

### 3 Metod

När man söker efter statistiska samband mellan en beroende variabel och kvalitativa förklaringsvariabler används med fördel variansanalys.

#### 3.1 Variansanalys med obalanserad data

Det är inte ovanligt att man har obalanserad data i sina experiment. I vårt fall har vi olika replikat i de olika länsregionerna. I länsregion 1 har vi fyra replikat, alltså fyra olika län (se bilaga 1), i länsregion 2 har vi 11 replikat och i länsregion 3 har vi sex olika replikat. Enligt Montgomery (2005) får man göra om formlerna som används vid balanserad variansanalys när man har så kallad obalanserad data. Med obalanserad data menas alltså att vi har olika replikat för olika grupper. Anta att vi har en tvåsidig variansanalysmodell med fixa effekter;

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$$

Där antal observationer i den  $ij$ :te cellen är  $n_{ij}$ . Dessutom är  $n_{i\cdot} = \sum_{j=1}^b n_{ij}$  antal observationer i den

$i$ :te raden (den  $i$ :te nivån hos faktor A) och  $n_{\cdot j} = \sum_{i=1}^a n_{ij}$  är antal observationer i den  $j$ :te kolumnen

(den  $j$ :te nivån hos faktor B). Detta innebär att  $n_{\cdot\cdot} = \sum_{i=1}^a \sum_{j=1}^b n_{ij}$  är det totala antalet observationer.

##### 3.1.1 Hypoteser och testfunktioner

Vi börjar alltid med att testa samspelseffekten. Finns det en signifikant samspelseffekt, med andra ord att vi förkastar nollhypotesen i testet

$$\begin{cases} H_0 : (\tau\beta)_{ij} = 0 \text{ för alla } ij \\ H_1 : (\tau\beta)_{ij} \neq 0 \text{ för minst ett par } ij \end{cases}$$

kan man genom en så kallad samspelsplott få en bild av och tolka samspellet mellan de två faktorerna. Samspelsplottar beskrivs vidare i delavsnitt 3.2. Finns det inget signifikant samspel är det möjligt att analysera de två faktorerna var för sig. Detta görs med följande test.

$$\begin{cases} H_0 : \tau_i = 0 \text{ för alla } i \\ H_1 : \tau_i \neq 0 \text{ för minst ett } i \end{cases}$$

$$\begin{cases} H_0 : \beta_j = 0 \text{ för alla } j \\ H_1 : \beta_j \neq 0 \text{ för minst ett } j \end{cases}$$

Testfunktionerna för de tre olika testen blir i tur och ordning;



$$F = \frac{MS_{AB}}{MS_E}, F = \frac{MS_A}{MS_E} \text{ och } F = \frac{MS_B}{MS_E}$$

Där alltså det första testet syftar till att testa samspelet, det andra till att testa skillnader hos faktor A's grupper och slutligen det sista testet ämnar analysera skillnader hos faktor B's grupper.

Det som skiljer obalanserad data från balanserad data är sättet man beräknar fram kvadratsummorna på. Med obalanserad data fås kvadratsummorna enligt;

$$SS_A = \sum_{i=1}^a \left( \frac{y_{i..}^2}{n_i} \right) - \frac{y_{...}^2}{N}$$

$$SS_B = \sum_{j=1}^b \left( \frac{y_{.j.}^2}{n_j} \right) - \frac{y_{...}^2}{N}$$

$$SS_{AB} = \sum_{i=1}^a \sum_{j=1}^b \left( \frac{y_{ij.}^2}{n_{ij}} \right) - \sum_{i=1}^a \left( \frac{y_{i..}^2}{n_i} \right) - \sum_{j=1}^b \left( \frac{y_{.j.}^2}{n_j} \right) + \frac{y_{...}^2}{N}$$

$$SS_{TOT} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n y_{ijk}^2 - \frac{y_{...}^2}{N}$$

$$SS_E = SS_T - SS_A - SS_B - SS_{AB}$$

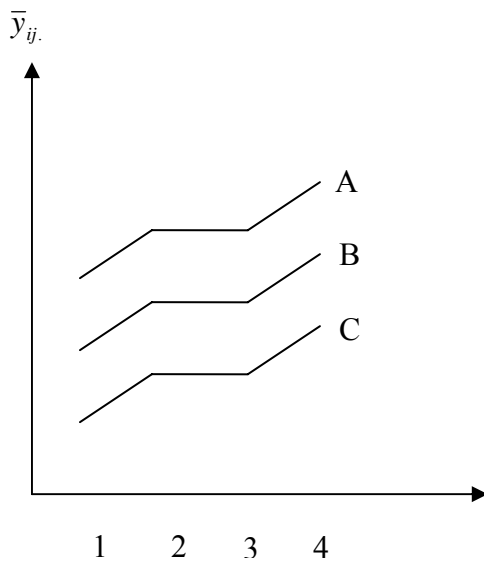
Med hjälp av kvadratsummorna tas medelkvadratsummorna fram, vilka alltså utgör nämnare och täljare i testfunktionerna.

De olika faktorerna i variansanalys har antingen fixa effekter eller stokastiska effekter. Med fixa effekter menas att nivåerna på en faktor inte är valda slumpmässigt, utan undersökningen täcker helt enkelt alla de nivåer som är intressanta. Säg istället att man har femton olika nivåer på en faktor och inte har råd (eller tid) att undersöka alla. Det är då möjligt att slumpmässigt välja ut, till exempel, 3 nivåer och analysera skillnader mellan dessa. I ett sådant fall är effekterna stokastiska. Testfunktionerna ovan är alla resultat av en analys där alla effekterna är fixa. Anledningen till att testfunktioner med stokastiska effekter inte presenteras är att det är irrelevant för denna undersökning, där ju alla effekterna är fixa. Inga slumpmässiga urval av nivåer har nämligen gjorts.

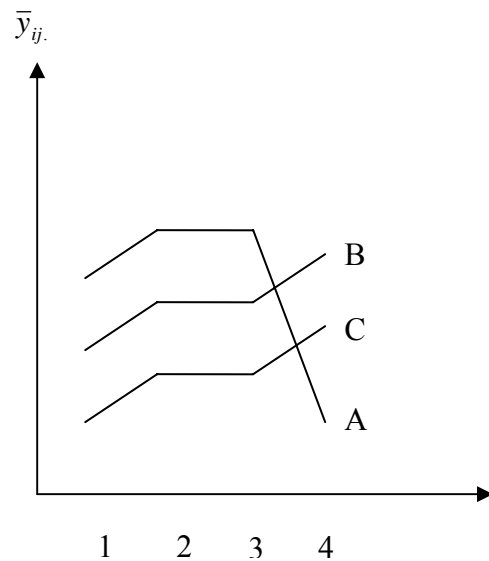
### 3.2 Tolkning av samspel

Samspelsplottar är figurer över interaktionen mellan två faktorer. En samspelsplott visar medelvärdena hos en faktors olika nivåer när den andra faktorns nivåer hålls konstant. Ett samspel mellan två faktorer betyder att reaktionen från en faktornivå beror på nivåerna hos den andra faktorn. Grafiskt ses detta när linjerna i samspelsplotten korsar varandra eller går isär. Parallella linjer indikerar däremot att inget samspel verkar finnas mellan faktorerna.

I figur 1 och 2 ser vi exempel på samspelsplottar. I den första figuren finns inga indikationer på samspel, medan vi i figur 2 tydligt kan se samspel.



**Figur 1:** Samspelsplot utan samspelseffekter



**Figur 2:** Samspelsplot med samspelseffekter

### 3.3 Styrkan av effekterna

När man i ett statistiskt experiment har funnit skillnader mellan olika gruppers medelvärden indikerar detta på att ett samband finns mellan den beroende och de oberoende variablerna. I variansanalys hör man, i detta sammanhang, ofta termen effekt. Effekten är en återspeglning av dessa skillnaden mellan populationernas medelvärden.

Enligt Hays (1974) är ett resultat som visar på både signifikanta skillnader mellan grupper och en relativ hög skattning av effekternas styrka mycket mera informativt än enbart de signifikanta skillnaderna.

För att på bästa sätt få en bild av vad som menas med effekternas styrka och hur de uppskattas, börjar vi med att förklara utifrån ensidig variansanalys för att sedan gå in på mera komplicerade modeller.

#### 3.3.1 Ensidig variansanalys

I en analys där man har en oberoende variabel med  $a$  olika behandlingar och där varje behandling utförs  $n$  gånger används följande modell:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, n \end{cases}$$

Effekten för behandling (grupp)  $i$  definieras enligt Hays (1974) som skillnaden mellan den  $i$ :te gruppens medelvärde  $\mu_i$  och populationens medelvärde  $\mu$ . Alltså;

$$\tau_i = \mu_i - \mu \text{ där } \hat{\tau}_i = \bar{x}_i - \bar{x}.$$

Om ingen effekt finns är  $\tau_i = 0$  för varje grupp  $i$ . Detta är ekvivalent med att skriva  $\mu_1 = \mu_2 = \dots = \mu_a = \mu$ .

Det är naturligtvis intressant att inte bara hitta skillnader mellan olika grupper, utan också få en idé om hur stor styrkan på sambandet mellan beroende och oberoende variabler dessa signifikanta skillnader representerar. Enligt Hays (1974) kan man för detta ändamål använda indexet  $\eta^2$  (eta i kvadrat), som kallas för ”*proportionen av variansen i Y som förklaras av variabeln X*”. När det kommer till variansanalys utökar man  $\eta^2$  så att den förklarar samma sak, fast nu med avseende på att den oberoende variabeln har olika grupper, eller nivåer, och alltså inte är kontinuerlig. Vi vill med andra ord att  $\eta^2$  representerar en procentuell förändring (minskning) i variationen hos Y som ges av att man vet att en observation tillhör en oberoende variabelkategori  $X_i$ , alltså  $X_1, X_2, \dots$  eller  $X_a$ .

Ett exakt mått på  $\eta^2$  är;

$$\eta^2 = \frac{\sigma_Y^2 - \sigma_\varepsilon^2}{\sigma_Y^2} = \frac{\sum_{i=1}^a \tau_i^2 p(X_i)}{\sigma_\varepsilon^2 + \sum_{i=1}^a \tau_i^2 p(X_i)}$$

$\eta^2$  symboliserar alltså hur mycket det betyder att veta till vilken grupp  $X_i$  en observation tillhör, när det kommer till förmågan att förutspå värdet på den beroende variabeln Y.

I ett stickprov kan vi skatta  $\eta^2$  genom att använda oss av:

$$E(MSB) = \sigma_\varepsilon^2 + \frac{\sum_{i=1}^a n_i \tau_i^2}{a-1}$$

$$E(MSW) = \sigma_\varepsilon^2$$

Om  $p(X_i) = \frac{n_i}{N}$  kan man grovt skatta  $\eta^2$  enligt:

$$\hat{\eta}^2 = \frac{SSB - (a-1)MSW}{SST + MSW}$$

Formeln kan också skrivas om som:

$$\hat{\eta}^2 = \frac{F' - 1}{\frac{v_2 + 1}{v_1} + F'}$$

Där  $F' = \frac{MSB}{MSW}$  när  $H_0$  förkastas ( $H_0: \eta^2 = 0$ ) och  $v_1 =$  antal frihetsgrader i nämnaren och  $v_2 =$  antal frihetsgrader i täljaren.

När  $H_0$  förkastas kommer  $F'$  att följa en icke-central (noncentral) F-fördelning med parametrarna  $v_1$ ,  $v_2$  och  $\delta^2$ . Man kan uttrycka väntevärdet för denna  $F'$ -fördelning som:

$$E(F') = \left( \frac{v_2}{v_2 - 2} \right) \left( 1 + \frac{\delta^2}{v_1} \right)$$

Där

$$\delta^2 = \frac{\sum_{i=1}^a \tau_i^2 p(x_i)}{\sigma^2} = \frac{\sum_{i=1}^a n_i \tau_i^2}{N \sigma^2}$$

Detta innebär att vi kan skriva om skattningen av  $\eta^2$  till:

$$\hat{\eta}^2 = \frac{\delta^2}{N + \delta^2}$$

Om  $F'$  byts ut mot stickprovets  $\frac{MSB}{MSW}$  blir skattningen av  $\eta^2$ :

$$\hat{\eta}^2 = \frac{\frac{(v_2 - 2)F' - 1}{v_2}}{\frac{v_2 + 1}{v_1} + \frac{(v_2 - 2)F' - 1}{v_2}}$$

Det är viktigt att påpeka att indexet  $\eta^2$  inte alltid är helt tillfredsställande. Största anledningen till detta kommer sig av att när man skattar  $\eta^2$  med hjälp av en icke-central F-fördelning ( $F'$ -fördelningen) kan man ibland få negativa värden på  $\eta^2$ , vilket inte är möjligt i praktiken då  $\eta^2 \geq 0$ . Skulle vi få ett negativt värde sätts  $\eta^2$  således lika med noll.

Om  $\eta^2$ -värdet är lika med 0,089 tolkas det som att vi sänker osäkerheten i Y med 8,9 % om vi vet till vilken grupp en observation tillhör.

### 3.3.2 Tvåsidig variansanalys

Modellen för en tvåsidig variansanalys ser ut enligt följande;

$$y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}$$

Där alltså  $\tau_i = \mu_i - \mu$  och  $\beta_j = \mu_j - \mu$ . Termen  $(\tau\beta)_{ij}$  symboliserar samspelseffekten mellan de två oberoende variablerna. Enligt Winer (1971) definieras parametern  $\theta_\tau^2$  enligt;

$$\theta_\tau^2 = \frac{\sum_{i=1}^a \tau_i^2}{a}$$

På samma sätt fås;

$$\theta_\beta^2 = \frac{\sum_{j=1}^b \beta_j^2}{b} \text{ och } \theta_{\tau\beta}^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b (\tau\beta)_{ij}^2}{ab}$$

Eftersom varianskomponenterna definieras som;

$$\sigma_\tau^2 = \frac{\sum_{i=1}^a \tau_i^2}{a-1}, \sigma_\beta^2 = \frac{\sum_{j=1}^b \beta_j^2}{b-1} \text{ och } \sigma_{\tau\beta}^2 = \frac{\sum_{i=1}^a \sum_{j=1}^b (\tau\beta)_{ij}^2}{(a-1)(b-1)}$$

kan vi skriva om  $\theta_\tau^2$ ,  $\theta_\beta^2$  och  $\theta_{\tau\beta}^2$  till;

$$\theta_\tau^2 = \frac{a-1}{a} \sigma_\tau^2, \theta_\beta^2 = \frac{b-1}{b} \sigma_\beta^2 \text{ och } \theta_{\tau\beta}^2 = \frac{(a-1)(b-1)}{ab} \sigma_{\tau\beta}^2$$

Skattningarna av  $\theta_\tau^2$  blir;

$$\hat{\theta}_\tau^2 = \frac{(a-1)(MSA - MSE)}{npq} = \frac{(a-1)(F_A - 1)MSE}{npq}$$

På samma sätt blir;

$$\hat{\theta}_\beta^2 = \frac{(b-1)(F_B - 1)MSE}{npq} \text{ och } \hat{\theta}_{\tau\beta}^2 = \frac{(a-1)(b-1)(F_{AB} - 1)MSE}{npq}$$

För att sedan beräkna hur mycket variationen i den förklarande variabeln som förklaras av de olika faktorerna A, B och dess samspel AB beräknas  $\eta^2$ -värdena enligt;

$$\eta_\tau^2 = \frac{\theta_\tau^2}{\theta_\tau^2 + \theta_\beta^2 + \theta_{\tau\beta}^2 + \sigma_\varepsilon^2}$$

Skattningen ges av;

$$\hat{\eta}_{\tau}^2 = \frac{(a-1)(F_A - 1)}{(a-1)(F_A - 1) + (b-1)(F_B - 1) + (a-1)(b-1)(F_{AB} - 1) + npq}$$

På samma sätt blir alltså;

$$\hat{\eta}_{\beta}^2 = \frac{(b-1)(F_B - 1)}{(a-1)(F_A - 1) + (b-1)(F_B - 1) + (a-1)(b-1)(F_{AB} - 1) + npq}$$

och

$$\hat{\eta}_{\tau\beta}^2 = \frac{(a-1)(b-1)(F_{AB} - 1)}{(a-1)(F_A - 1) + (b-1)(F_B - 1) + (a-1)(b-1)(F_{AB} - 1) + npq}$$

I en modell av högre ordning, säg en tresidig variansanalysmodell eller en fyrsidig variansanalysmodell, kan  $\eta^2$ , för de olika faktorerna och samspelet, skattas analogt.

## 4 Resultat

I tidigare avsnitt nämndes att variansanalys var ett bra analysverktyg när man söker efter statistiska samband mellan en beroende variabel och kvalitativa förklaringsvariabler. Eftersom så är fallet i denna uppsats använder vi med fördel variansanalys. Den modell som bäst passar vår data är en fyrsidig variansanalys med fixa effekter. Modellen som följer av de olika databehandlingarna blir följande;

$$y_{ijklm} = \mu + \tau_i + \beta_j + \gamma_k + \delta_l + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\beta\gamma)_{ijk} + (\tau\delta)_{il} + (\beta\delta)_{jl} + (\tau\beta\delta)_{ijl} + (\gamma\delta)_{kl} + (\tau\gamma\delta)_{ikl} + (\beta\gamma\delta)_{jkl} + (\tau\beta\gamma\delta)_{ijkl} + \varepsilon_{ijklm}$$

$$\left. \begin{array}{l} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, c \\ l = 1, 2, \dots, d \\ m = 1, 2, \dots, n_h \end{array} \right\}$$

Där  $y_{ijklm}$  utgör de olika observationernas ”Andel av befolkningen med sjukpenning”.  $a = 9$  (nio olika åldersgrupper),  $b = 6$  (sex olika år),  $c = 2$  (de olika könen),  $d = 3$  (de tre olika länsregionerna) och eftersom vår data är obalanserad symboliserar  $n_h$  de olika replikaten i de olika länsregionerna. I bilaga 1 ser vi att i länsregion 1 har vi 4 replikat, vilket alltså innebär att  $n_1 = 4$ . Länsregion 2 som har hela 11 replikat gör att  $n_2 = 11$  och länsregion 3 med sina 6 replikat ger  $n_3 = 6$ .

### 4.1 Översikt över den beroende variabeln

Innan vi börjar med analysen är det intressant att skaffa sig en bild av hur ”Andel av befolkningen med sjukpenning” ser ut. I tabell 1 ges en sammanfattning av variabeln. Vi kan utläsa att i genomsnitt 5,8 procent av befolkningen mellan 20 och 64 år får sjukpenning. Vi ser också att skillnaden mellan maximum och minimum är relativt stor.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Andel av befolkningen som får sjukpenning	2268	,0032	,1895	,057891	,0296675
Valid N (listwise)	2268				

Tabell 1: Översikt över den beroende variabeln

### 4.2 Variansanalys

Tabell 2 visar resultatet från den fyrsidiga variansanalysen med fixa effekter. Vi ser att många av samspelen har höga p-värden och alltså kan exkluderas från modellen.

### Tests of Between-Subjects Effects

Dependent Variable: Andel av befolkningen med sjukpenning

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1,879 <sup>a</sup>	323	5,818E-03	97,553	,000	,942
Intercept	6,877	1	6,877	115304,7	,000	,983
Länsregion	,108	2	5,421E-02	908,812	,000	,483
Årtal	6,358E-02	5	1,272E-02	213,202	,000	,354
Kön	,552	1	,552	9261,441	,000	,827
Ålder	,891	8	,111	1866,675	,000	,885
Länsregion * Årtal	1,994E-03	10	1,994E-04	3,342	,000	,017
Länsregion * Kön	2,403E-02	2	1,202E-02	201,447	,000	,172
Årtal* Kön	5,221E-03	5	1,044E-03	17,506	,000	,043
Länsregion * Årtal* Kön	2,749E-04	10	2,749E-05	,461	,915	,002
Länsregion * Ålder	1,879E-02	16	1,175E-03	19,694	,000	,139
Årtal * Ålder	1,041E-02	40	2,602E-04	4,362	,000	,082
Länsregion * Årtal * Ålder	1,762E-03	80	2,203E-05	,369	1,000	,015
Kön * Ålder	6,902E-02	8	8,628E-03	144,651	,000	,373
Länsregion * Kön * Ålder	5,567E-03	16	3,480E-04	5,834	,000	,046
Årtal * Kön * Ålder	1,998E-03	40	4,995E-05	,838	,755	,017
Länsregion * Årtal * Kön * Ålder	7,522E-04	80	9,403E-06	,158	1,000	,006
Error	,116	1944	5,964E-05			
Total	9,596	2268				
Corrected Total	1,995	2267				

a. R Squared = ,942 (Adjusted R Squared = ,932)

**Tabell 2:** Variansanalys av den ursprungliga modellen.

För att således få en bättre och enklare modell tas dessa osignifikanta parametrar bort en efter en. Vi börjar med att ta bort fyrfaktorsamspelen och fortsätter därefter tills dess att alla parametrarna är signifikanta. Detta leder till att vår ”nya” modell<sup>1</sup> blir.

$$y_{ijklm} = \mu + \tau_i + \beta_j + \gamma_k + \delta_l + (\tau\beta)_{ij} + (\tau\gamma)_{ik} + (\beta\gamma)_{jk} + (\tau\delta)_{il} + (\beta\delta)_{jl} + (\gamma\delta)_{kl} + (\tau\gamma\delta)_{ikl} + \varepsilon_{ijklm}$$

$$\left. \begin{array}{l} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, c \\ l = 1, 2, \dots, d \\ m = 1, 2, \dots, n_h \end{array} \right\}$$

I modellen har vi sex signifikanta tvåfaktorssamspel och ett signifikant trefaktorssamspel. I tabell 3 nedan redovisas resultatet från variansanalysen i SPSS och i bilaga 2 hittas resultaten från

<sup>1</sup> Vi har valt att döpa denna modell till ”nya” modellen för att senare i uppsatsen inte blanda ihop den med den slutgiltiga modellen.



Tukeys test i SAS. Anledningen till att båda resultaten finns redovisade är att SPSS ger en bra sammanfattande bild av analysen medan SAS redovisar skillnaderna mellan de olika grupperna mera ingående.

### Tests of Between-Subjects Effects

Dependent Variable: Andel av befolkningen med sjukpenning

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1,875 <sup>a</sup>	113	1,659E-02	295,976	,000	,939
Intercept	6,877	1	6,877	122700,7	,000	,983
Ålder	,891	8	,111	1986,410	,000	,881
Årtal	6,358E-02	5	1,272E-02	226,877	,000	,345
Kön	,552	1	,552	9855,501	,000	,821
Länsregion	,108	2	5,421E-02	967,106	,000	,473
Årtal * Ålder	1,231E-02	40	3,078E-04	5,491	,000	,093
Kön * Ålder	6,902E-02	8	8,628E-03	153,929	,000	,364
Årtal * Kön	5,704E-03	5	1,141E-03	20,352	,000	,045
Länsregion * Ålder	1,879E-02	16	1,175E-03	20,957	,000	,135
Länsregion * År	1,994E-03	10	1,994E-04	3,557	,000	,016
Länsregion * Kön	2,403E-02	2	1,202E-02	214,368	,000	,166
Länsregion * Kön * Ålder	5,567E-03	16	3,480E-04	6,208	,000	,044
Error	,121	2154	5,605E-05			
Total	9,596	2268				
Corrected Total	1,995	2267				

a. R Squared = ,939 (Adjusted R Squared = ,936)

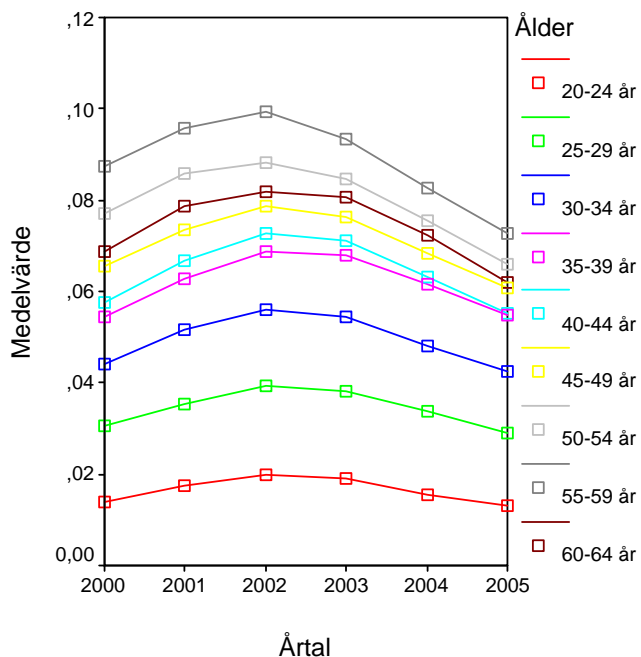
**Tabell 3:** Variansanalys av den ”nya” modellen

## 4.3 Samspelseffekter

Detta delavsnitt syftar till att förklara de olika samspelseffekterna. Genom så kallade samspelsplottar är det möjligt att få en bild av samspelet som gör det lättare att tolka interaktionen. Varje samspel analyseras var för sig i nedanstående avsnitt.

### 4.3.1 Samspelet mellan ålder och årtal

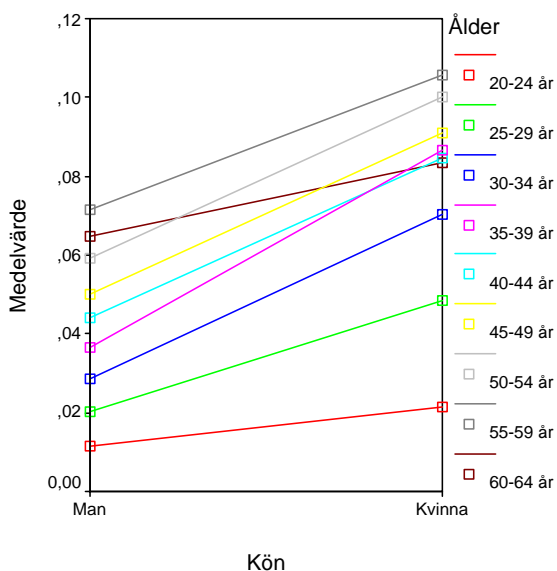
I figur 3 ser vi den första av våra samspelsplottar. Här ser vi tydligt att desto äldre man är, desto större andel av befolkningen får sjukpenning. Ett undantag är emellertid den äldsta åldersgruppen som alltid ”bara” har den tredje största andelen med sjukpenning. Något större samspel ser vi egentligen inte ifrån figuren. Personer i åldersgrupp 1 har alltid minst andel som får sjukpenning och åldersgrupp 8 har alltid högst andel med sjukpenning, oavsett vilket år vi analyserar. Ska man vara petig ser vi däremot indikationer på att skillnaden mellan åldersgrupperna verkar minska med åren och då framförallt för de äldre åldersgrupperna.  $\eta^2$ -värdet säger oss att variationen i ”Andel av befolkningen med sjukpenning” endast till 9,3 procent (se tabell 3) beror på samspelet mellan ålder och årtal. Detta i kombination med att vi inte ser tydligare tendenser till korsade linjer gör att vi väljer att bortse från det signifikanta samspelet.



Figur 3: Samspelsplot mellan ålder och årtal

### 4.3.2 Samspelet mellan ålder och kön

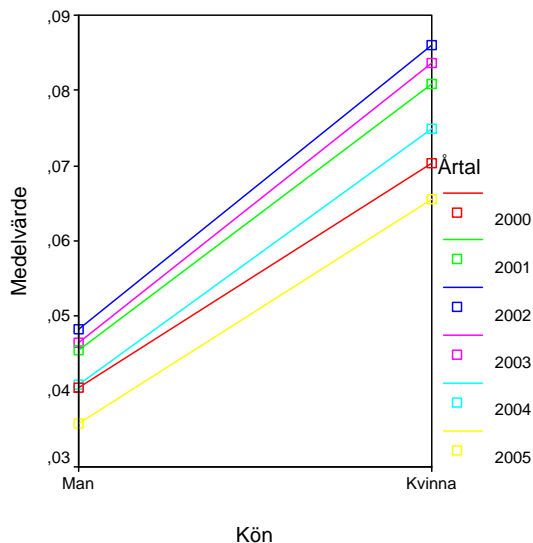
Figur 4 visar på ett tydligt samspel mellan kön och ålder. Först ser vi att kvinnor alltid har högre andel som får sjukpenning oavsett vilken åldersgrupp vi undersöker. Vad som skiljer sig är att åldersgrupp 9 (alltså den äldsta) för män är den åldersgruppen som har den näst högsta andelen med sjukpenning, medan samma åldersgrupp ligger på sjätteplats för kvinnor. Kvinnorna i åldern 35-39 år tilldelas även till större del sjukpenning än vad kvinnor i ålders 40-44 år och 60-64 år gör, vilket inte är fallet för männen. Även  $\eta^2$ -värdet är relativt högt. 36,4 procent av variationen i "Andel av befolkningen med sjukpenning" kan förklaras av detta samspel.



Figur 4: Samspelsplot mellan kön och ålder

### 4.3.3 Samspelet mellan årtal och kön

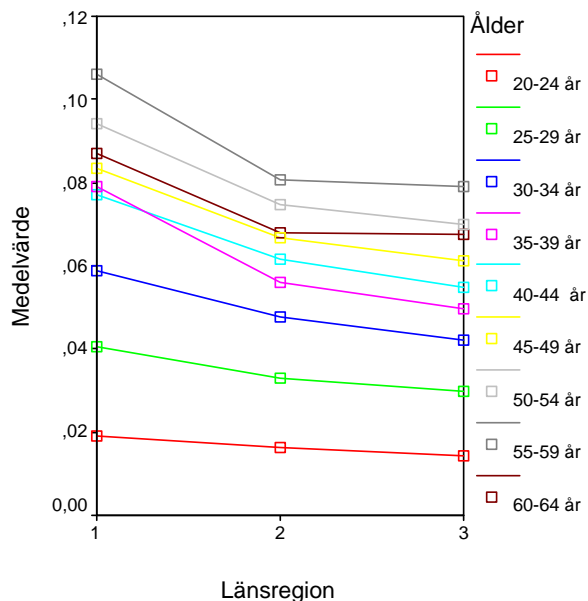
Samspelet mellan kön och årtal verkar, med hänsyn till figur 5, inte speciellt stort. Vi kan tydligt se att kvinnor i alla åren har haft högre andel som fått sjukpenning än männen. Då inga linjer korsar varandra, vilket ju skulle indikera på ett samspel, och det faktumet att  $\eta^2$ -värdet är 4,5 procent gör att vi kan bortse fram detta signifikanta samspel.



Figur 5: Samspeletsplot mellan årtal och kön

### 4.3.4 Samspelet mellan ålder och länsregion

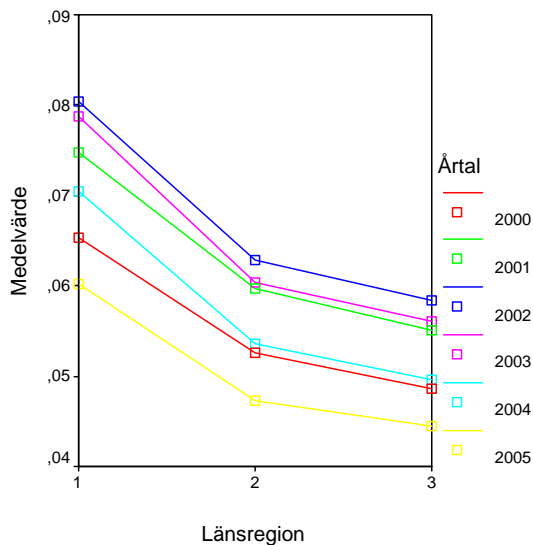
Länsregion 1 har alltid högst ”Andel av befolkningen med sjukpenning”, oavsett vilken ålder som prövas. Länsregion 3 har alltid lägst. Med andra ord tilldelas det en större andel sjukpenning till personer som bor i den norra delen av Sverige, än vad som tilldelas i de sydligare delarna. Utifrån figur 6 kan vi se att två linjer korsas en gång. Det är därför tydligt att det finns ett samspel mellan länsregion 1 och 2 och åldersgrupperna 4 och 5. I länsregion 1 tilldelas alltså individer i åldern 35-39 år procentuellt sett mera sjukpenning än vad som ges individer i åldern 40-44 år. Utöver detta samspel är linjerna i stort sett parallella.  $\eta^2$ -värdet, som är 13,5 procent är relativt högt och vi väljer därför att inte bortse ifrån detta samspel.



Figur 6: Samspelsplot mellan ålder och länsregion

### 4.3.5 Samspelet mellan årtal och länsregion

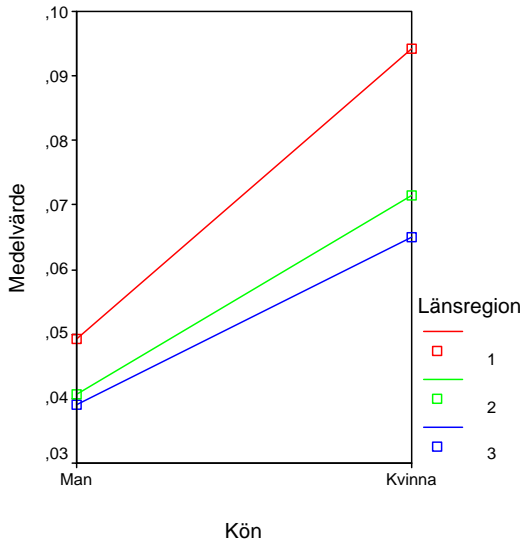
Endast 1,6 procent av variationen i ”Andel av befolkningen med sjukpenning” kan förklaras av samspelet mellan årtal och länsregion. Figur 7 antyder också att linjerna är någorlunda parallella. Detta medför att vi väljer att bortse från detta samspel.



Figur 7: Samspelsplot mellan årtal och länsregion

### 4.3.6 Samspelet mellan kön och länsregion

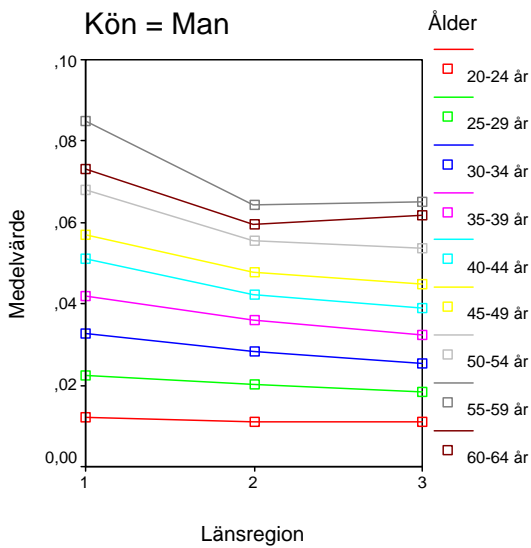
Inte heller här ser vi att linjerna korsar varandra. Däremot visar figur 8 på att skillnaderna mellan regionerna är mycket större för kvinnorna än för männen, vilket även det indikerar på ett samspel.  $\eta^2$ -värdet är 16,6 procent och F-värdet är högt (214,368). Ett högt F-värde tyder på att samspelseffekterna är stora. Vi kan alltså inte bortse ifrån detta samspel.



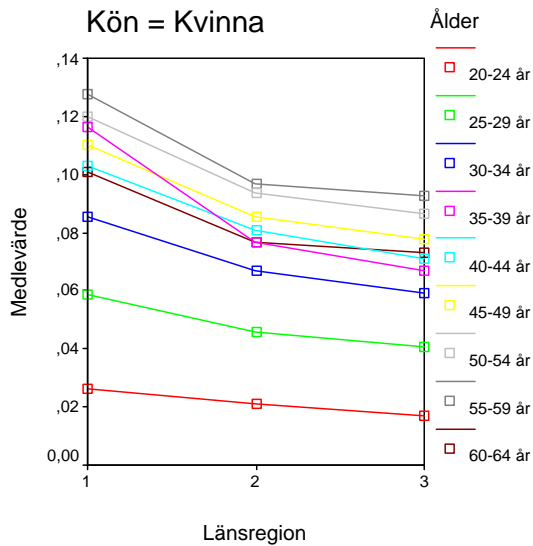
**Figur 8:** Samspeletsplot mellan länsregion och kön

### 4.3.7 Samspelet mellan ålder, kön och länsregion

I Figur 9 ser vi inga tecken på samspel, då alla linjerna verkar parallella. Ett samspel går däremot att utläsa från figur 10, men då endast 4,4 procent av variationen i ”Andel av befolkningen med sjukpenning” kan förklaras av samspelet och det faktumet att F-värdet är relativt lågt gör att vi väljer att bortse ifrån detta. Ett bättre sätt att avbilda ett trefaktorsamspel vore med en tredimensionell graf, men på grund av bristande kunskaper kunde en sådan graf inte redovisas.



**Figur 9:** Samspeletsplot mellan ålder, länsregion och män



Figur 10: Samspelsplot mellan ålder, länsregion och kvinnor

#### 4.4 Sammanfattande resultat

Eftersom många av samspelen inte var speciellt betydande är det nu möjligt att förenkla vår modell ytterligare. Från ovanstående delavsnitt framgick det att endast tre samspel var väsentliga. Samspelet mellan kön och ålder, mellan kön och region och slutligen mellan region och ålder. Inkluderas endast dessa tillsammans med de enskilda faktorerna blir vår slutgiltiga modell;

$$y_{ijklm} = \mu + \tau_i + \beta_j + \gamma_k + \delta_l + (\tau\gamma)_{ik} + (\tau\delta)_{il} + (\gamma\delta)_{kl} + \varepsilon_{ijklm} \begin{cases} i = 1, 2, \dots, a \\ j = 1, 2, \dots, b \\ k = 1, 2, \dots, c \\ l = 1, 2, \dots, d \\ m = 1, 2, \dots, n_h \end{cases}$$

I tabell 4 ser vi resultatet av en variansanalys på vår slutgiltiga modell.

### Tests of Between-Subjects Effects

Dependent Variable: Andel av befolkningen med sjukpenning

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	1,849 <sup>a</sup>	42	4,402E-02	669,510	,000	,927
Intercept	6,877	1	6,877	104588,7	,000	,979
Ålder	,891	8	,111	1693,194	,000	,859
Årtal	6,967E-02	5	1,393E-02	211,912	,000	,323
Kön	,552	1	,552	8400,718	,000	,791
Länsregion	,108	2	5,421E-02	824,350	,000	,426
Kön * Ålder	7,116E-02	8	8,895E-03	135,278	,000	,327
Länsregion * Ålder	1,879E-02	16	1,175E-03	17,864	,000	,114
Länsregion * Kön	2,403E-02	2	1,202E-02	182,725	,000	,141
Error	,146	2225	6,576E-05			
Total	9,596	2268				
Corrected Total	1,995	2267				

a. R Squared = ,927 (Adjusted R Squared = ,925)

**Tabell 4:** Variansanalys med den slutgiltiga modellen

I bilaga 3 redovisas resultaten från Tukeys test. Jämför vi denna modell med modellen som vi kallade ”nya” modellen ser vi att den justerade förklaringsgraden ”bara” har sjunkit från 0,936 till 0,925. Eftersom faktorn Årtal inte finns med i något av de tre samspelet är det alltså möjligt att analysera denna separat. Eftersom  $\eta^2$ -värdet är 32,3 procent förklarar alltså variabeln Årtal 32,3 procent av den totala variationen i ”Andel av befolkningen med sjukpenning”. I bilaga 3 kan vi med hjälp av Tukeys test se att det finns signifikanta skillnader mellan alla åren förutom mellan år 2001 och år 2003. Från och med år 2002 till år 2005 finns signifikanta skillnader mellan alla åren. Vi ser även att ”Andel av befolkningen med sjukpenning” blir allt lägre.

Ifrån samspeletsplottarna i ovanstående delavschnitt 4.3.2, 4.3.4 och 4.3.6 ser vi att det finns tydliga samspel mellan kön och länsregion, mellan kön och ålder samt mellan ålder och länsregion. Kombination kvinnor och länsregion 1 ger till exempel en hög andel som får sjukpenning. Ser man på samspelet mellan kön och ålder, visar det sig att skillnaderna mellan män och kvinnor inte är lika stora för individer i åldersgruppen 20-24 år och 60-64 år, som hos de övriga åldersgrupperna. Dessa åldersgrupper ser ut att ha någorlunda lika skillnader. Störst skillnad finns dock hos åldersgruppen 35-39 år, där andelen med sjukskrivning ungefär är dubbelt så stor för kvinnorna. I figur 6 åskådliggjordes interaktionen mellan ålder och länsregion. Det visade sig där att individer i åldern 35-39 år som är bosatta i länsregion 1 till större del tilldelas sjukpenning än individer i åldern 40-44. Inga tendenser till samspel kunde däremot ses mellan de olika åldersgrupperna och länsregion 1 och 2. Detta eftersom linjerna i samspeletsplotten var nästintill parallella.

Ifrån tabell 4 kan vi se att 85,9 procent av variationen i ”Andel av befolkningen med sjukpenning” kan förklaras av de olika åldersgrupperna. Lite av denna variation kommer däremot också från kön samt lite från länsregion. 77 procent av variationen i ”Andel av befolkningen med sjukpenning” kan förklaras av vilket kön en individ tillhör. På samma sätt beror alltså lite av

denna variation på vilken länsregion en individ bor i samt vilken ålder han/hon har. Länsregion förklarar 42,6 procent av variationen i den beroende variabeln, men denna andel måste precis som ovan tolkas som om att en del av variationen beror på kön och ålder.



## 5 Slutdiskussion

Att det finns skillnader mellan hur stor andel av männen och kvinnorna som får sjukpenning hade vi, med hänsyn till det inledande avsnittet, redan en klar bild av. Däremot var det intressant att se hur stora dessa skillnader egentligen är. Förutom skillnader mellan könen lyckades vi i denna uppsats även visa på att det finns skillnader mellan olika regioner, olika åldrar och olika årtal. Att äldre människor till större grad tilldelas sjukpenning kommer ju inte som någon överraskning, men att de personer som arbetar sina sista år innan pensioneringen faktiskt har en längre andel som tilldelas sjukpenning än de som är något yngre var förbluffande. Tukeys test visade på signifikanta skillnader mellan alla åldersgrupperna (se bilaga 3), men vi måste som sagt ta denna signifikans med en nypa salt eftersom lite av variationen beror på vilket kön en individ har samt i vilken del av Sverige denne bor. Samspelet mellan ålder och kön gav oss en bild av att skillnaderna mellan könen ”*Andel av befolkningen med sjukpenning*” var mindre för den äldsta och den yngsta åldersgruppen, samtidigt som skillnaderna var som allra störst för de mellersta åldersgrupperna.

De regionala skillnaderna var överraskande. Vad som gör att svenskar som är bosatta längre norrut till större del tilldelas sjukpenning är svårt att avgöra. Kanske kan det bero på att man har en annan typ av sysselsättning som är mera belastande och stressande och därav har en högre andel sjukskrivningar. Detta är däremot bara spekulationer. Ifrån bilaga 3 kan vi utläsa att Tukeys test visar på signifikanta skillnader mellan alla tre länsregionerna. Minst skillnad finns mellan länsregion 2 och 3. Störst skillnad ses mellan den nordligaste och den sydligaste länsregionen. Precis som för skillnaderna mellan åldersgrupperna ska vi alltså inte lägga för stor vikt vid denna signifikans eftersom en del av variationen beror på de könsmässiga och åldersmässiga skillnaderna. Samspelet mellan länsregion och kön gjorde det möjligt att se att kvinnor i norra delarna av landet har en högre andel som tilldelas sjukpenning än i övriga Sverige.

Ytterligare en sak som var förvånande är att samspelet mellan ålder och länsregion pekade på att det finns skillnader mellan hur stor andel som tilldelas sjukpenning i de olika åldersgrupperna i de olika länsregionerna. Hur det kan komma sig att en yngre åldersgrupp, 35-39 år, i länsregion 1 får procentuellt sett mera sjukpenning än en något äldre åldersgrupp, 40-44 år, är svårtolkat. I de två andra länsregionerna förekommer nämligen inte detta särdrag.

Även fast det inte heller är möjligt att analysera faktorn kön för sig själv, på grund av samspelseffekterna mellan kön och åldersgrupper samt kön och länsregion, pekar mycket av undersökningen på att det finns stora skillnader mellan könen. Fastän  $\eta^2 = 0,791$  för variabeln kön och en del av denna variation beror på variationen i de tre länsregionerna och i de olika åldersgrupperna, tycks ändå mycket av variationen bero på endast de könsmässiga skillnaderna. Tittar man till exempel en extra gång på de samspelsplottar som analyserar bland annat kön, ser vi att linjerna alltid är tilltagande. Detta betyder alltså att kvinnor alltid har högre andel utbetald sjukpenning än vad männen har.

Att det finns signifikanta skillnader mellan åren 2000-2005 har påvisats och vi ser en tydlig neråtående trend efter år 2002. Undersökningen visar därmed att Sveriges välfärdsmål, att halvera ohälsotalet, går åt rätt håll.

## Sammanfattning

Enligt TCO (2003) är det värsta hotet mot Sveriges välfärd att så många svenskar till följd av sjukdom och arbetslöshet inte befinner sig i arbetskraften. En halvering av det så kallade ohälsotalen skulle innebära att bruttonationalprodukten skulle stiga med mer än fem procent. I reda pengar innebär detta 110 miljarder kronor om året. SCB anser att skillnader mellan bland annat män och kvinnors sjukskrivningar bör studeras mera ingående. Detta för att man ska skaffa sig en bättre bild av varför sjukskrivningarna uppkommer. Ett bra sätt att studera Sveriges sjukskrivningar på är genom att titta på vem som tilldelas sjukpenning.

Syftet med denna uppsats är att se efter vart skillnaderna i tilldelad sjukpenning finns, samt att studera och försöka tolka eventuella samspel. För att få en bättre förståelse för hur mycket varje analyserad faktor har för betydelse för skillnaden i tilldelad sjukpenning beräknas styrkan av effekten. Detta mått ( $\eta^2$ ), tillsammans med signifikanta skillnader, ger oss en mera informativ bild av hur stor vikt man ska lägga på de signifikanta skillnaderna.

Det faktorer som i förhållande till, vad vi valt att kalla, ”*Andel av befolkningen med sjukpenning*” analyseras i uppsatsen är ålder, kön, länsregion och tid. Åldersvariabeln är indelad i 9 olika åldersgrupper, länsregionsvariabeln i tre grupper och tiden som analyseras är åren 2000 till och med 2005. Anledningen till att en tidsvariabel också har tagit med är för att analysera om sjukskrivningarna tenderar att sjunka med tiden.

Analysen görs med hjälp av en fyrsidig variansanalysmodell med fixa effekter. Undersökningen börjar med att eliminera osignifikanta samspelseffekter. Därefter analyseras vikten av de resterande signifikanta samspelen med hjälp av samspeleffekter och  $\eta^2$ -värdet. I en den slutliga modellen ser vi att det finns signifikanta samspel mellan kön och ålder, kön och länsregion samt mellan ålder och länsregion. Det finns till exempel en klar tendens till att kvinnor i norra Sverige i högre grad tilldelas sjukpenning. Ser man på samspelet mellan kön och ålder, visar det sig att skillnaderna mellan män och kvinnor inte är lika stora för individer i åldersgruppen 20-24 år och 60-64 år. Däremellan är skillnaderna någorlunda lika. Störst är skillnaden däremot mellan könen när det kommer till åldersgruppen 35-39 år, här är andelen med sjukskrivning ungefär dubbelt så stor. Ytterligare en slutsats vi kunnat dra var att en yngre åldersgrupp, 35-39 år, i länsregion 1 procentuellt sett får mera sjukpenning än en något äldre åldersgrupp, 40-44 år. I de två andra länsregionerna förekommer däremot inte detta särdrag.

Klara tendenser till att ”*Andel av befolkningen med sjukpenning*” sjunker med tiden har hittats. Detta tyder alltså på att Sveriges mål att halvera ohälsotalen har lyckats bra under de senaste fyra åren.

## Litteraturförteckning

Hays, W.L., *Statistics for the social sciences*, Holt, Rinehart and Winston, London. 1974.

Marcoulides, G.A. & Hershberger, S.L., *Multivariate Statistical Methods*, Lawrence Erlbaum Associates, Inc., Mahwah, 1997.

Montgomery, D.C., *Design and Analysis of Experiments*, John Wiley & Sons, Inc., Danvers, 2005.

Winer, B.J., *Statistical principles in experimental design*, McGraw-Hill Kogakusha, LTD., Tokyo, 1971.

SCB, *Sjukfrånvaro och ohälsa i Sverige – en belysning utifrån SCB:s statistik*, SCB-tryck, Örebro, 2004:3

TCO, *Ohälsan kräver akutgrupp*, [http://www.sif.se/EPiBrowser/Dokumentarkiv/Press/Mari-Anns%20artiklar/Debart\\_SvD\\_20030408.pdf](http://www.sif.se/EPiBrowser/Dokumentarkiv/Press/Mari-Anns%20artiklar/Debart_SvD_20030408.pdf), 2003-04-08

LO, *ohälsans trappa*, EO print AB, Stockholm, 2004

Försäkringskassan, *Socialförsäkringens omfattning och finansiering 2003-2006*, AB Danagårds Grafiska, 2005

## Bilaga 1

### Sveriges län

Stockholms län	länsregion 2
Uppsala län	länsregion 2
Södermanlands län	länsregion 2
Östergötlands län	länsregion 2
Jönköpings län	länsregion 3
Kronobergs län	länsregion 3
Kalmar län	länsregion 3
Gotlands län	länsregion 3
Blekinge län	länsregion 3
Skåne län	länsregion 3
Hallands län	länsregion 2
Västra Götalands län	länsregion 2
Värmlands län	länsregion 2
Örebro län	länsregion 2
Västmanlands län	länsregion 2
Dalarnas län	länsregion 2
Gävleborgs län	länsregion 2
Västernorrlands län	länsregion 1
Jämtlands län	länsregion 1
Västerbottens län	länsregion 1
Norrbottens län	länsregion 1



### Åldersgrupper

- 20-24 år
- 25-29 år
- 30-34 år
- 35-39 år
- 40-44 år
- 45-49 år
- 50-54 år
- 55-59 år
- 60-64 år

## Bilaga 2 "Nya" modellen

### SAS-kommandon

```
OPTIONS NODATE NOCENTER;
DATA SJUKPENNINGEN;
INPUT Aldersgrupp Ar Kon ProcentMedSjukpenning Lansregion;
CARDS;
1          2000          1          0.016811765          2
1          2000          1          0.012809042          2
1          2000          1          0.016107561          2
.
.
.
9          2005          0          0.058194316          1
;
PROC GLM;
CLASS Aldersgrupp Ar Kon Lansregion;
MODEL ProcentMedSjukpenning = Aldersgrupp Ar Kon Lansregion Aldersgrupp*Ar
Aldersgrupp*Kon Ar*Kon Aldersgrupp*Lansregion Ar*Lansregion Kon*Lansregion
Aldersgrupp*Kon*Lansregion;
MEANS Kon Ar Lansregion Aldersgrupp / TUKEY;
MANOVA H=Aldersgrupp Ar Kon Lansregion Aldersgrupp*Ar Aldersgrupp*Kon Ar*Kon
Aldersgrupp*Lansregion Ar*Lansregion Kon*Lansregion Aldersgrupp*Kon*Lansregion
/ PRINTH PRINTE;
RUN;
```

### Resultat

The GLM Procedure

Tukey's Studentized Range (HSD) Test for ProcentMedSjukpenning

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	2154
Error Mean Square	0.000056
Critical Value of Studentized Range	2.77337
Minimum Significant Difference	0.0006

Means with the same letter are not significantly different.

T  
u  
k  
e  
y

G  
r  
o  
u  
p  
i  
n  
g

	Mean	N	Kon
A	0.0740012	1134	1
B	0.0417806	1134	0

he GLM Procedure

Tukey's Studentized Range (HSD) Test for ProcentMedSjukpenning

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	2154
Error Mean Square	0.000056
Critical Value of Studentized Range	4.03377
Minimum Significant Difference	0.0016

Means with the same letter are not significantly different.

T  
u  
k  
e  
y

G  
r  
o  
u  
p  
i  
n  
g

	Mean	N	Ar
A	0.0649073	378	2002
B	0.0626573	378	2003
B	0.0612883	378	2001
C	0.0556637	378	2004
D	0.0538615	378	2000

E 0.0489673 378 2005

The GLM Procedure

Tukey's Studentized Range (HSD) Test for ProcentMedSjukpenning

NOTE: This test controls the Type I experimentwise error rate.

Alpha 0.05  
Error Degrees of Freedom 2154  
Error Mean Square 0.000056  
Critical Value of Studentized Range 3.31679

Comparisons significant at the 0.05 level are indicated by \*\*\*.

Lansregion Comparison	Difference Between Means	Simultaneous 95% Confidence Limits
1 - 2	0.0156074	0.0146209 0.0165939 ***
1 - 3	0.0196611	0.0185705 0.0207518 ***
2 - 1	-0.0156074	-0.0165939 -0.0146209 ***
2 - 3	0.0040538	0.0031963 0.0049112 ***
3 - 1	-0.0196611	-0.0207518 -0.0185705 ***
3 - 2	-0.0040538	-0.0049112 -0.0031963 ***

The GLM Procedure

Tukey's Studentized Range (HSD) Test for ProcentMedSjukpenning

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha 0.05  
Error Degrees of Freedom 2154  
Error Mean Square 0.000056  
Critical Value of Studentized Range 4.39101  
Minimum Significant Difference 0.0021  
Means with the same letter are not significantly different.

T  
u  
k  
e  
y  
  
G  
r  
o

u p i n g	Mean	N	Aldersgrupp
A	0.0849831	252	8
B	0.0770517	252	7
C	0.0714267	252	9
D	0.0683403	252	6
E	0.0625875	252	5
F	0.0587086	252	4
G	0.0482053	252	3
H	0.0335177	252	2
I	0.0161975	252	1



## Bilaga 3 Slutgiltiga modellen

### SAS-kommandon

```
OPTIONS NODATE NOCENTER;
DATA SJUKPENNINGEN;
INPUT Aldersgrupp Ar Kon ProcentMedSjukpenning Lansregion;
CARDS;
1          2000          1          0.016811765          2
1          2000          1          0.012809042          2
1          2000          1          0.016107561          2
.
.
.
9          2005          0          0.058194316          1
;
PROC GLM;
CLASS Aldersgrupp Ar Kon Lansregion;
MODEL ProcentMedSjukpenning = Aldersgrupp Ar Kon Lansregion Aldersgrupp*Kon
Aldersgrupp*Lansregion Kon*Lansregion;
MEANS Kon Ar Lansregion Aldersgrupp / TUKEY;
MANOVA H=Aldersgrupp Ar Kon Lansregion Aldersgrupp*Kon Aldersgrupp*Lansregion
Kon*Lansregion / PRINTH PRINTE;
RUN;
```

### Resultat

The GLM Procedure

Tukey's Studentized Range (HSD) Test for ProcentMedSjukpenning

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	2225
Error Mean Square	0.000066
Critical Value of Studentized Range	2.77332
Minimum Significant Difference	0.0007

Means with the same letter are not significantly different.

T  
u  
k  
e  
y  
  
G  
r  
o  
u

p  
 i  
 n  
 g

	Mean	N	Kon
--	------	---	-----

A	0.0740012	1134	1
---	-----------	------	---

B	0.0417806	1134	0
---	-----------	------	---

The GLM Procedure

Tukey's Studentized Range (HSD) Test for ProcentMedSjukpenning

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha	0.05
Error Degrees of Freedom	2225
Error Mean Square	0.000066
Critical Value of Studentized Range	4.03365
Minimum Significant Difference	0.0017

Means with the same letter are not significantly different.

T  
 u  
 k  
 e  
 y

G  
 r  
 o  
 u  
 p  
 i  
 n  
 g

	Mean	N	Ar
--	------	---	----

A	0.0649073	378	2002
---	-----------	-----	------

B	0.0626573	378	2003
---	-----------	-----	------

B	0.0612883	378	2001
---	-----------	-----	------

C	0.0556637	378	2004
---	-----------	-----	------

D	0.0538615	378	2000
---	-----------	-----	------

E	0.0489673	378	2005
---	-----------	-----	------

The GLM Procedure

Tukey's Studentized Range (HSD) Test for ProcentMedSjukpenning

NOTE: This test controls the Type I experimentwise error rate.

Alpha 0.05  
 Error Degrees of Freedom 2225  
 Error Mean Square 0.000066  
 Critical Value of Studentized Range 3.31672

Comparisons significant at the 0.05 level are indicated by \*\*\*.

Lansregion Comparison	Difference Between Means	Simultaneous 95% Confidence Limits	
1 - 2	0.0156074	0.0145389 0.0166759	***
1 - 3	0.0196611	0.0184799 0.0208424	***
2 - 1	-0.0156074	-0.0166759 -0.0145389	***
2 - 3	0.0040538	0.0031250 0.0049825	***
3 - 1	-0.0196611	-0.0208424 -0.0184799	***
3 - 2	-0.0040538	-0.0049825 -0.0031250	***

The GLM Procedure

Tukey's Studentized Range (HSD) Test for ProcentMedSjukpenning

NOTE: This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ.

Alpha 0.05  
 Error Degrees of Freedom 2225  
 Error Mean Square 0.000066  
 Critical Value of Studentized Range 4.39087  
 Minimum Significant Difference 0.0022

Means with the same letter are not significantly different.

T  
u  
k  
e  
y

G  
r  
o  
u  
p  
i  
n  
g

	Mean	N	Aldersgrupp
A	0.0849831	252	8

B	0.0770517	252	7
C	0.0714267	252	9
D	0.0683403	252	6
E	0.0625875	252	5
F	0.0587086	252	4
G	0.0482053	252	3
H	0.0335177	252	2
I	0.0161975	252	1