



Matematisk statistik
Stockholms universitet

**Förslag till ny estimator
i undersökningen
"IT bland individer"**

Frida Videll

Examensarbete 2005:9

Postal address:

Matematisk statistik
Dept. of Mathematics
Stockholms universitet
SE-106 91 Stockholm
Sweden

Internet:

<http://www.math.su.se/matstat>



Matematisk statistik
Stockholms universitet
Examensarbete 2005:9,
<http://www.math.su.se/matstat>

Förslag till ny estimator i undersökningen ”IT bland individer”

Frida Videll*

November 2005

Abstract

Undersökningen IT bland individer utförs årligen genom att använda ett suburval från arbetskraftsundersökningen, AKU. I nuläget presenterar undersökningen av IT bland individer enbart punktskattningar eftersom beräkningarna av dessa är uppbyggt på ett komplext sätt och inte kan beskrivas med ett vanligt skattningsförfarande. I detta examensarbete undersöker vi hur estimatorn är uppbyggd för att kunna beräkna variansskattningar. Eftersom estimatorn är komplex beräknar vi approximativa variansskattningar. För att få en enklare och effektivare estimator undersöker vi vilka registervariabler som samvarierar med hur individerna svarar på frågorna samt med svarsfrekvensen. Efter denna analys av hjälpvariabler kommer vi fram till ett förslag till en ny estimator. Vid jämförelse mellan de approximativa variansskattningarna och de variansskattningar som fås med den nya estimatorn måste vi vara försiktiga med att dra slutsatser eftersom vi ej vet om de approximativa variansskattningarna under- eller överskattar den sanna variansen. Med den nya estimatorn kan punkt- och variansskattningar beräknas på ett enkelt sätt med en känd programvara vilket leder till slutsatsen att den nya estimatorn bör användas.

*Postal address: Matematisk statistik, Stockholms universitet, SE-106 91, Stockholm, Sweden. E-mail: fridavidell@hotmail.com. Handledare: Anders Björkström

Suggestion for a new estimator in the “Community survey on ICT usage in households and by individuals”

Abstract

The Community survey on ICT usage in households and by individuals is a annual survey that uses a subsample from the Labour force survey. The situation today is that the Community survey on ICT usage in households and by individuals only publish point estimates. The reason for this is that the computation of the point estimates is so complex that it can't be described with a ordinary survey design. In this study we examine how the point estimates are computed with the aim to compute variance estimates. We find that the estimator is so complex that we in this study only compute approximative variance estimates for the point estimates. In order to find a simple estimator we investigate the variables that are known for every individual in the subsample. We examine if these variables effect how the individuals answer on the questions and if they answer or not. This analysis lead to a suggestion for a new estimator that is simple to understand. When we compare the approximative variance estimates with the variance estimates that the new estimator gives we have to be careful when we draw conclusions because we don't know if the approximative variance estimates are under- or overestimates of the true variance. The biggest advantage with the new estimator, compared with the estimator that is used, is that that point- and variance estimates easily can be computed. This lead us to the conclusion that the new estimator should be used.

Förord

Denna rapport är utförd under VT 2005/HT 2005 och utgör mitt examensarbete omfattande 20 poäng för magisterexamen i matematisk statistik från Matematik-ekonomi-linjen på Stockholms Universitet.

Arbetet har utförts på Statistiska Centralbyrån, metodenheten för ekonomisk statistik. Jag vill med detta förord tacka Statistiska Centralbyrån för att de gav mig en arbetsplats och därmed även en inblick i arbetslivet.

Stort tack till min handledare på Statistiska Centralbyrån, Stefan Svanberg, som genom att han alltid fanns tillhands gav mig värdefull rådgivning och vägledning. Jag vill även tacka de på Statistiska Centralbyrån som har gett mig den hjälp som behövts för att genomföra detta examensarbete.

Min handledare på Stockholms Universitet, Anders Björkström, ska också ha ett stort tack för den handledning han har givit mig under arbetets gång.

Innehåll

1	INLEDNING	1
1.1	Undersökningen IT bland individer	1
1.2	Arbetskraftsundersökningen	1
1.3	Syfte och tillvägagångssätt	2
1.4	Disposition	2
2	METODER	3
2.1	Urvalsdesign	3
2.1.1	Obundet slumpurval	3
2.1.2	Stratifierat urval	4
2.1.3	Systematiskt urval	4
2.1.4	Poststratifiering	5
2.1.5	Regressionsestimatorn	5
2.1.6	Kalibrering	6
2.2	Variansestimering	7
2.2.1	CLAN	7
2.2.2	Bootstrap	8
3	ESTIMATORN I DATORVANEUNDERSÖKNINGEN 2004	10
3.1	Beskrivning av uppräkningsvikter och hjälpvariabler i AKU	12
3.1.1	Arbetslösa	13
3.1.2	Sysselsatta och Akstatus 1	13
3.1.3	Akstatus 5	13

3.1.4	Åldersklassen 65-74 år	14
3.2	Beskrivning av uppräkningsvikter och hjälpvariabler i datorvaneundersökningen	15
4	ANALYS	16
4.1	Approximativ variansskattning	16
4.2	Val av hjälpinformation	17
4.2.1	Nationalitet	18
4.2.2	Region	20
4.2.3	Kön	21
4.2.4	Utbildningsnivå	22
4.2.5	Ålder	23
4.2.6	Slutsats	26
4.3	Val av estimator	27
4.4	Jämförelse av estimatorer	32
5	RESULTAT OCH SLUTSATSER	36
A	Pearsons χ^2-test	39
A.1	Exempel på beräkning av χ^2 -statistikan	39
B	Variabler	41
C	Hjälpvektorer	43
D	Rotationsgrupp	46
E	Tabeller	47

Tabeller

3.1	<i>Redovisningsgrupper</i>	10
4.1	<i>Punktskattningar av de som använt en persondator under första kvartalet 2004 samt skattningar av standardavvikelserna</i>	17
4.2	<i>Medborgarskap</i>	18
4.3	χ^2 -värden för variabeln <i>AktNat2</i>	19
4.4	<i>Invandringsår</i>	19
4.5	χ^2 -värden för variabeln <i>SenInvAr2</i>	20
4.6	<i>Regionuppdelning</i>	20
4.7	χ^2 -värden för variabeln <i>Hregg</i>	20
4.8	<i>Grövre indelning av Hregg</i>	21
4.9	χ^2 -värden för den grövre indelningen av <i>Hregg</i>	21
4.10	<i>Kön</i>	22
4.11	χ^2 -värden för variabeln <i>Kon</i>	22
4.12	<i>Utbildningsnivå</i>	22
4.13	χ^2 -värden för variabeln <i>UtbNiv</i>	22
4.14	<i>Utbildningsinriktning</i>	23
4.15	χ^2 -värden för variabeln <i>UtbInr</i>	23
4.16	<i>Åldersindelningen (16-24 år, 25-34 år, 35-44 år, 45-54 år, 55-74 år)</i>	23
4.17	χ^2 -värden för åldersindelningen (16-24 år, 25-34 år, 35-44 år, 45-54 år, 55-74 år)	24
4.18	<i>Svarsfördelningen för individer som är 55-74 år</i>	24
4.19	<i>Åldersindelningen (16-34 år, 35-54 år, 55-64 år, 65-74 år)</i>	24
4.20	χ^2 -värden för åldersindelningen (16-34 år, 35-54 år, 55-64 år, 65-74 år)	25

4.21	<i>Svarsfördelning för olika åldersklasser</i>	25
4.22	<i>Antal individer, som ingår i datorvaneundersökningen, i varje åldersklass vid åldersindelningen (16-34 år, 35-54 år, 55-64 år, 65-74 år)</i>	26
4.23	<i>Antal individer, som ingår i datorvaneundersökningen, i varje åldersklass vid åldersindelningen (16-34 år, 35-54 år, 55-62 år, 63-74 år)</i>	26
4.24	<i>Åldersindelningen (16-34 år, 35-54 år, 55-62 år, 63-74 år)</i>	26
4.25	<i>χ^2-värden för åldersindelningen (16-34 år, 35-54 år, 55-62 år, 63-74 år)</i>	27
4.26	<i>Variationskoefficienter och standardavvikelser för totalerna för fråga B</i>	30
4.27	<i>Sammanfattande tabell för totalerna för fråga B</i>	30
4.28	<i>Sammanfattande tabell för andelarna för fråga B</i>	31
4.29	<i>Sammanfattande tabell för totalerna för fråga C</i>	31
4.30	<i>Sammanfattande tabellen för andelarna för fråga C</i>	32
4.31	<i>Punktskattningar för fråga B</i>	32
4.32	<i>Punktskattningar för fråga C</i>	33
4.33	<i>Variationskoefficienter för fråga B</i>	34
4.34	<i>Variationskoefficienter för fråga C</i>	34
A.1	<i>Indelning av observationerna</i>	39
A.2	<i>Observerade frekvenser för medborgarskap</i>	40
A.3	<i>Förväntade frekvenser för medborgarskap</i>	40
A.4	<i>χ^2-värden i varje cell för medborgarskap</i>	40
B.1	<i>Beskrivning av variabler</i>	41
B.2	<i>Beskrivning av SNI</i>	42
C.1	<i>Hjälpvariabeln Help1</i>	43
C.2	<i>Hjälpvariabeln Help2</i>	44

C.3	<i>Hjälpvariabeln Help3</i>	45
D.1	<i>Rotationsgrupper för AKU-urvalet april, 2004</i>	46
E.1	<i>Gruppering för Akstatus 5</i>	47
E.2	<i>Poststratifieringen i datorvaneundersökningen</i>	47
E.3	<i>Poststratifieringen för Fall 1</i>	48

Figurer

3.1	<i>Andel individer i olika åldrar</i>	11
-----	---------------------------------------	----

1 INLEDNING

1.1 Undersökningen IT bland individer

Undersökningen IT bland individer, kallas fortsättningsvis för datorvaneundersökningen, är en årlig undersökning vars syfte är att beskriva IT-användningen i Sverige¹. Denna undersökning görs även i ett antal andra europeiska länder inom EU och samordnas av Eurostat som är EU:s statistikbyrå. Fyra grupper av variabler framkommer i undersökningen, dessa är tillgång till olika typer av IT, användning av olika typer av IT, upplevda säkerhetsproblem vid användning av Internet samt vilka typer av säkerhetsåtgärder som vidtas vid användningen av Internet. Urvalet till denna undersökning är ett suburval från AKU som får svara på tilläggsfrågor angående IT-vanor. De individer som ingår i datorvaneundersökningen kommer från paneler som blir tillfrågade i april. År 2004 svarade 3 626 personer på frågorna i datorvaneundersökningen.

Informationen som fås av denna undersökning används främst av Eurostat, Näringsdepartementet, Statens institut för kommunikationsanalys, Institutet för tillväxtpolitiska studier, Post- och telestyrelsen, statstkontoret och NUTEK.

1.2 Arbetskraftsundersökningen

Arbetskraftsundersökningen, AKU, genomförs varje månad och avser att ge en bild av sysselsättningen och utvecklingen på arbetsmarknaden i Sverige². Urvalsförfarandet är panelundersökning³ med roterande urval, vilket innebär att urvalsindividerna ingår i undersökningen vid flera tillfällen.

Första gången som Statistiska Centralbyrån, SCB, genomförde AKU var augusti 1961. I början, från och med 1961 till och med 1969, gjordes undersökningarna kvartalsvis men 1970 skedde en övergång till månadsvisa undersökningar.

Tillsammans med annan arbetsmarknadsstatistik används resultaten från AKU som underlag för planering av arbetsmarknadspolitiska program samt för uppföljning av arbetsmarknadsinsatser. AKU används även för allmänna konjunkturbedömningar och internationella jämförelser samt för den ekonomiska och sociala forskningens behov av data om arbetsmarknaden. Även massmedia använder sig av AKU för att ge information inom området till allmänheten. De viktigaste användarna av statistiken från AKU är därför Arbetsmarknadsstyrelsen, AMS, Närings- och Finansdepartementen, Konjunkturinstitutet, arbetsmark-

¹Se [14].

²Se [13].

³För mer information om panelundersökningar se [7], sida 138.

nadens organisationer samt massmedia som press, radio och TV.

1.3 Syfte och tillvägagångssätt

För att skatta totaler och andelar inom datorvaneundersökningen används ett komplext skattningsförfarande som har bidragit till att variansskattningar ej har beräknats.

Syftet med detta examensarbete är att, för datorvaneundersökningen 2004, utreda hur skattningsförfarandet är uppbyggt för att kunna undersöka om detta förfarande kan göras på ett mer effektivt sätt som ger lägre variansskattningar samt bidrar till att variansskattningarna kan beräknas på ett enklare sätt. Vi kommer att jämföra olika skattningsförfaranden genom att se vilka varianser som de genererar och på detta sätt dra slutsatser om vilket skattningsförfarande som är lämpligt att använda samt se vad precisionsvinsten är av att använda ett mer komplext skattningsförfarande, vilket görs idag.

1.4 Disposition

I början av detta examensarbete ges en överblick av AKU och datorvaneundersökningen. I Avsnitt 2 ges en översiktlig bild av olika urvalsdesigner som används i AKU och datorvaneundersökningen samt hur variansskattningar kan beräknas på olika sätt. En beskrivning av AKU och datorvaneundersökningen ges i Avsnitt 3 medan Avsnitt 4 behandlar olika skattningsförfaranden som är aktuella att använda i datorvaneundersökningen. Avslutningsvis presenteras slutsatserna i Avsnitt 5.

2 METODER

2.1 Urvalsdesign

En undersökning görs för att beskriva hur verkligheten ser ut och för att få en fullständig bild av verkligheten skulle en totalundersökning behöva göras, det vill säga alla objekt i målpopulationen skulle behöva tillfrågas. Detta är i de flesta fall omöjligt att göra och för att försöka komma runt detta problem dras ett urval, s , av n objekt som får representera målpopulationen, U , som består av N objekt. Objekten i detta urval kommer därför att få representera fler objekt i målpopulationen än enbart sig själva. För att få fram skattningar kommer därför värdet på undersökningsvariabeln att multipliceras med en uppräkningsvikt, d_k . Denna uppräkningsvikt kan fås fram på olika sätt, den kan vara inversen av urvalssannolikheten och beror därför på urvalsdesignen. Uppräkningsvikten kan också beräknas genom att utnyttja information som finns tillgänglig för objekten för att på så sätt korrigera för variabler som kan samvariera med undersökningsvariabeln. Nedan kommer vi att gå igenom olika skattningsförfaranden som är aktuella i, eller har anknytning till, detta arbete⁴. Skattningar för obundet slumpurval, stratifierat urval och systematiskt urval beräknas genom att använda Horvitz-Thompson estimatorn som ger väntevärdesriktiga eller approximativt väntevärdesriktiga skattningar av populationsstorheterna genom att för varje element i urvalet använda inversen av inklusions sannolikheten som uppräkningsvikt.

2.1.1 Obundet slumpurval

Vid obundet slumpurval, OSU, dras ett urval utan återläggning från målpopulationen. Detta innebär att alla objekt som ingår i urvalsregistret har samma sannolikhet att komma med i urvalet samt att alla objekt som ingår i urvalet har samma uppräkningsvikt, det vill säga varje objekt bidrar med lika mycket till skattningarna.

$$P(\text{objekt } k \text{ tillhör urvalet}) = \pi_k = \frac{n}{N}, \quad k = 1, 2, \dots, N \quad (2.1)$$

Uppräkningsvikten blir nu $d_k = \frac{1}{\pi_k}$ för varje objekt.

⁴För vidare läsning om OSU, stratifiering urval och systematiskt urval se [7], sida 58, 74 samt 95. För mer information om kalibrering se [5]. För fördjupning av poststratifiering, GREG och Bootstrap rekommenderas [11], sida 265, 225 samt 442. En mer ingående beskrivning av variansskattningen i CLAN fås i [1].

Horvitz-Thompson estimatoren för totalen av undersökningsvariabeln Y blir nu $\hat{Y}_{tot} = \sum_s \frac{N}{n} y_k$, där y_k är värdet på undersökningsvariabeln för objekt k .

2.1.2 Stratifierat urval

En stratumindelning av populationen görs innan urvalet dras för att skapa homogena grupper, med avseende på undersökningsvariabeln, för att minska variansen. Populationen delas då upp i disjunkta delmängder, A_1, A_2, \dots, A_H , efter den eller de variabler som tros samvariera med undersökningsvariabeln, dessa delmängder kallas för stratan. Denna uppdelning av objekten görs för att skapa strata där de objekt som ingår i samma strata liknar varandra, med avseende på undersökningsvariabeln, medan skillnaden mellan olika stratum är stor. Om indelningen lyckas skapa homogena strata som skiljer sig från varandra erhålles en lägre varians än vad som fås med OSU. Från vart och ett av dessa stratan dras ett sannolikhetsurval, dessa urval dras oberoende av varandra. Det totala urvalet fås genom att slå ihop de olika stratumurvalen. Vid stratifierat urval har de objekt som tillhör samma stratum samma sannolikhet att komma med i urvalet.

$$P(\text{objekt } k \text{ tillhör urvalet}) = \pi_{hk} = \frac{n_h}{N_h}, \quad k = 1, 2, \dots, N \quad (2.2)$$

där n_h =antalet objekt i urvalet från stratum h och N_h =totala antalet objekt i stratum h .

Vid stratifierat urval blir uppräkningsvikten $d_k = \frac{1}{\pi_{hk}}$ och Horvitz-Thompson estimatoren för totalen av undersökningsvariabeln Y blir därför $\hat{Y}_{tot} = \sum_{h=1}^H \frac{N_h}{n_h} T_h(\mathbf{y})$. Där $T_h(\mathbf{y}) = \sum_h y_k$.

2.1.3 Systematiskt urval

Vid ett systematiskt urval ordnas de N objekten i urvalsregistret enligt någon variabel. Objektens registernummer bestäms av ordningen i registret, det vill säga det objekt som står först i registret har registernummer 1. Om objekten ordnas efter personnummer blir åldersfördelningen mindre spridd jämfört med ett stratifierat urval med åldersklasser som stratum. För att dra urvalet bestäms en steglängd, t , som $t = \frac{N}{n}$. Därefter väljs en startpunkt, B , på måfå bland de t första objekten i registret, detta innebär att $P(B = k) = \frac{1}{t}$, $k = 1, 2, \dots, t$. Med steglängden t tas steg fram i registret med start i B , vilket innebär att de objekt som ingår i urvalet har registernummer $B, B + t, B + 2t, \dots, B + (n - 1)t$. Vid systematiskt urval är inklusionssannolikheten, och därmed även

uppräkningsvikten, för objekten samma som vid OSU, inklusions sannolikheten för objekt k ges därför av (2.1). Detta bidrar även till att Horvitz-Thompson estimatorn av totalen för undersökningsvariabeln Y blir samma som för OSU.

2.1.4 Poststratifiering

Tanken bakom poststratifiering liknar den bakom stratifiering med den skillnad att stratifiering tas med redan i urvalssteget medan poststratifiering kommer med i skattningssteget. Efter att urvalet har dragits kan det visa sig att en eller flera variabler, vars värden för objekten i urvalet är kända, samvarierar med undersökningsvariabeln. Dessa variabler kan då användas för att dela upp objekten i disjunkta poststrata G_1, G_2, \dots, G_D där de objekt inom samma poststratum liknar varandra när det gäller undersökningsvariabeln. Vid poststratifiering bibehåller objekten sina urvalssannolikheter som bestäms av vilket urvalsförfarande som används. Om urvalet dras med hjälp av OSU kommer de inom samma poststratum att ha samma uppräkningsvikt. Om urvalet i stället dras med hjälp av stratifiering kommer de som tillhör samma kombination av poststrata och strata att få samma uppräkningsvikt. Då samma indelning görs för stratum och poststratum kommer alla objekt som tillhör samma stratum att få samma uppräkningsvikt. Skulle indelningen av poststratum skilja sig från indelningen av stratum kommer det att finnas stratum där alla objekt ej har samma uppräkningsvikt. För vart och ett av de poststrata som bildas görs en skattning av den aktuella populationsstorheten som sedan bidrar till den slutliga skattningen av populationsstorheten.

Skattningen av totalen för undersökningsvariabeln ges av $\hat{Y}_{tot} = \sum_{d=1}^D N_d \bar{Y}_{sd}$ där \bar{Y}_{sd} är y -medelvärdet för de svarande i G_d .

2.1.5 Regressionsestimatorn

Regressionsestimatorn, GREG, använder sig av hjälpinformation och svarssannolikheten, θ_k , för att få fram skattningar. För att bestämma $\hat{\theta}_k$ kan urvalet s delas upp i svarshomogenitetsgrupper, SHG. Denna gruppindelning görs på ett sådant sätt att det finns skäl att tro att svarssannolikheterna inom en given grupp är ungefär lika stora. Vi kallar dessa grupper för $s_1, \dots, s_p, \dots, s_P$ och motsvarande svarsmängder $r_1, \dots, r_p, \dots, r_P$. Storleken på dessa mängder är $n_1, \dots, n_p, \dots, n_P$, respektive $m_1, \dots, m_p, \dots, m_P$. Skattningen av θ_k som används i samband med SHGs är nu $\hat{\theta}_k = \frac{m_p}{n_p}$, för $k \in r_p$. För hjälpinformationen kan två olika situationer förekomma:

- (i) $\mathbf{t}_x = \sum_U \mathbf{x}_k$ är känd och \mathbf{x}_k är känd för alla $k \in$ urvalet, s

(ii) \mathbf{x}_k är känd för alla $k \in s$

Skattningen av populationsstorheten i situation (i) ser därför ut på följande sätt⁵:

$$\hat{Y}_{reg,U} = \sum_r w_{Uk} \frac{y_k}{\hat{\theta}_k} \quad (2.3)$$

$$\text{där } w_{Uk} = d_k \left(1 + c_k \left(\mathbf{t}_x - \sum_r \frac{d_k \mathbf{x}_k}{\hat{\theta}_k}\right)' \left(\sum_r \frac{d_k c_k \mathbf{x}_k \mathbf{x}_k'}{\hat{\theta}_k}\right)^{-1} \mathbf{x}_k\right), \text{ för } k \in s. \quad (2.4)$$

Konstanten c_k anger residualvariansen i regressionsmodellen och gör estimatorn mer flexibel.

I situation (ii) blir skattningen av populationsstorheten:

$$\hat{Y}_{reg,U} = \sum_r w_{sk} \frac{y_k}{\hat{\theta}_k} \quad (2.5)$$

$$\text{där } w_{sk} = d_k \left(1 + c_k \left(\sum_s d_k \mathbf{x}_k - \sum_r \frac{d_k \mathbf{x}_k}{\hat{\theta}_k}\right)' \left(\sum_r \frac{d_k c_k \mathbf{x}_k \mathbf{x}_k'}{\hat{\theta}_k}\right)^{-1} \mathbf{x}_k\right), \text{ för } k \in s. \quad (2.6)$$

2.1.6 Kalibrering

Kalibrering är ett skattningsförfarande som använder sig av hjälpinformation för att reducera urvalsfelet och bortfallsskevheten. Genom att använda kalibreringstekniken kan produkten av ett antal delvikter oftast erhållas i ett enda steg. Dessa delvikter fås, bland annat, av att objekten har olika inklusions sannolikheter, poststratifiering används samt att bortfall förekommer. För att få fram kalibreringsvikterna används hjälpinformation, \mathbf{x}_k . Denna hjälpinformation ska så bra som möjligt förklara variationen i svarssannolikheterna, förklara variationen i de viktigaste undersökningsvariablerna samt identifiera de viktigaste redovisningsgrupperna. I situation (i) i Avsnitt 2.1.5 erhåller vi följande skattning av populationsstorheten:

⁵För en härledning av (2.3) se [3].

$$\hat{Y}_{kal,U} = \sum_r d_k v_{Uk} y_k \quad (2.7)$$

$$\text{där } v_{Uk} = 1 + c_k (\mathbf{t}_x - \sum_r d_k \mathbf{x}_k)' (\sum_r d_k c_k \mathbf{x}_k \mathbf{x}_k')^{-1} \mathbf{x}_k, \text{ för } k \in s. \quad (2.8)$$

Skattningen av populationsstorleken blir i situation (ii):

$$\hat{Y}_{kal,U} = \sum_r d_k v_{sk} y_k \quad (2.9)$$

$$\text{där } v_{sk} = 1 + c_k (\sum_s d_k \mathbf{x}_k - \sum_r d_k \mathbf{x}_k)' (\sum_r d_k c_k \mathbf{x}_k \mathbf{x}_k')^{-1} \mathbf{x}_k, \text{ för } k \in s. \quad (2.10)$$

Den kalibreringsvikt som fås ska vara så lik $d_k = \frac{1}{\pi_k}$ som möjligt och samtidigt uppfylla villkoret $\sum_r d_k v_{Uk} \mathbf{x}_k = \sum_U \mathbf{x}_k$ om vi befinner oss i situation (i). Om vi befinner oss i situation (ii) blir villkoret istället $\sum_r d_k v_{sk} \mathbf{x}_k = \sum_s d_k \mathbf{x}_k$.

Vid jämförelse mellan (2.3) och (2.7) ses att dessa ger samma skattningar om alla i urvalet svarar, det vill säga då vi ej har bortfall.

2.2 Variansestimering

En parameter, Y , ska skattas för en fix ändlig population, U , av storlek N , där Y är en funktion av J totaler $\mathbf{t} = (t_1, t_2, \dots, t_j, \dots, t_J)'$. Detta innebär att $Y = f(t_1, t_2, \dots, t_j, \dots, t_J) = f(\mathbf{t})$ där $t_j = \sum_U y_{jk}$ är totalen för variabeln y_j i populationen och y_{jk} är värdet på y_j för objekt k . Skattningen av Y görs genom att dra ett urval, s , från populationen med en given urvalsdesign för att beräkna skattningar av $t_1, t_2, \dots, t_j, \dots, t_J$. Y skattas nu genom att ersätta \mathbf{t} med $\hat{\mathbf{t}} = (\hat{t}_1, \hat{t}_2, \dots, \hat{t}_j, \dots, \hat{t}_J)$ vilket ger $\hat{Y} = f(\hat{t}_1, \hat{t}_2, \dots, \hat{t}_j, \dots, \hat{t}_J)$. För att beräkna variansskattningen, $\hat{V}(\hat{Y})$, till Y kan olika tillvägagångssätt användas, två av dessa beskrivs i Avsnitten 2.2.1 och 2.2.2.

2.2.1 CLAN

Problem vid beräkning av variansen för Y uppkommer då $f(\mathbf{t})$ är en icke-linjär funktion. I detta arbete används en SAS-baserad programvara, CLAN97, som

löser detta problem genom att använda den första termen i Taylorserieapproximation för funktionen f . Detta tillvägagångssätt lyckas bra om antagandet om att urvalsstorleken är så stor att högre ordningens termer i Taylorserieapproximationen är försumbara håller. Vi får därför följande uttryck:

$$\hat{Y} - Y \approx \sum_{j=1}^J f'_j(\mathbf{t})(\hat{t}_j - t_j) \quad (2.11)$$

där $f'_j(\mathbf{t}) = \frac{\partial f(\mathbf{t})}{\partial t_j}$ är den partiala derivatan av f i \mathbf{t} .

MSE av \hat{Y} approximeras därför med

$$MSE(\hat{Y}) \approx V\left(\sum_{j=1}^J f'_j(\mathbf{t})\hat{t}_j\right) = \sum_{i=1}^J \sum_{j=1}^J f'_i(\mathbf{t})f'_j(\mathbf{t})Cov(\hat{t}_i, \hat{t}_j) \quad (2.12)$$

Eftersom $MSE(\hat{Y}) = V(\hat{Y}) + Bias^2(\hat{Y})$ kan $MSE(\hat{Y})$ användas som en approximation av $V(\hat{Y})$ om $Bias^2(\hat{Y})$ är liten jämfört med $V(\hat{Y})$, vilket är vanligt då urvalsstorleken är stor. Eftersom de partiala derivatorna, $f'_1, f'_2, \dots, f'_j, \dots, f'_J$, vanligtvis beror på okända \mathbf{t} ersätts $f'_j(\mathbf{t})$ och $Cov(\cdot, \cdot)$ av urvalsbaserade $f'_j(\hat{\mathbf{t}})$ och $\widehat{Cov}(\cdot, \cdot)$ för att beräkna variansskattning. Detta leder till estimatorn

$$\hat{V}(\hat{Y})_{CLAN} = \sum_{i=1}^J \sum_{j=1}^J f'_i(\hat{\mathbf{t}})f'_j(\hat{\mathbf{t}})\widehat{Cov}(\hat{t}_i, \hat{t}_j) \quad (2.13)$$

Denna estimator, $\hat{V}(\hat{Y})_{CLAN}$, är vanligtvis inte en väntevärdesriktig skattning av varken den sanna $MSE(\hat{Y})$ eller $V(\hat{Y})$, estimatorn är däremot en konsistent skattning av $MSE(\hat{Y})$ om \hat{t}_j och $\widehat{Cov}(\hat{t}_i, \hat{t}_j)$ är konsistenta skattningar av t_j och $Cov(\hat{t}_i, \hat{t}_j)$.

2.2.2 Bootstrap

Ett annat sätt att beräkna $\hat{V}(\hat{Y})$ är att använda Bootstrap som är en teknik som först användes utanför ramen för urvalsundersökningar för att beräkna approximativa variansskattningar och konfidensintervall. Vid bootstrap skapas populationen U^* som är en återskapning av U . Från U^* dras en serie oberoende urval, "återurval" eller "bootstrapurval", genom att använda samma urvalsdesign som användes då s drogs från U . För att urvalen ska vara oberoende måste varje

bootstrapurval ersätts i U^* innan nästa dras. För vart och ett av dessa bootstrapurval beräknas skattningen \hat{Y}_a^* ($a=1,2,\dots,A$), där A är antalet bootstrapurval som görs, på samma sätt som \hat{Y} beräknas. Med hjälp av den observerade fördelningen för $\hat{Y}_1^*, \hat{Y}_2^*, \dots, \hat{Y}_A^*$, som ses som en skattning av fördelningen för \hat{Y} , skattas $V(\hat{Y})$ av

$$\hat{V}(\hat{Y})_{BS} = \frac{1}{A-1} \sum_{a=1}^A (\hat{Y}_a^* - \hat{Y}^*)^2 \quad (2.14)$$

där $\hat{Y}^* = \frac{1}{A} \sum_{a=1}^A \hat{Y}_a^*$.

3 ESTIMATORN I DATORVANEUNDERSÖKNINGEN 2004

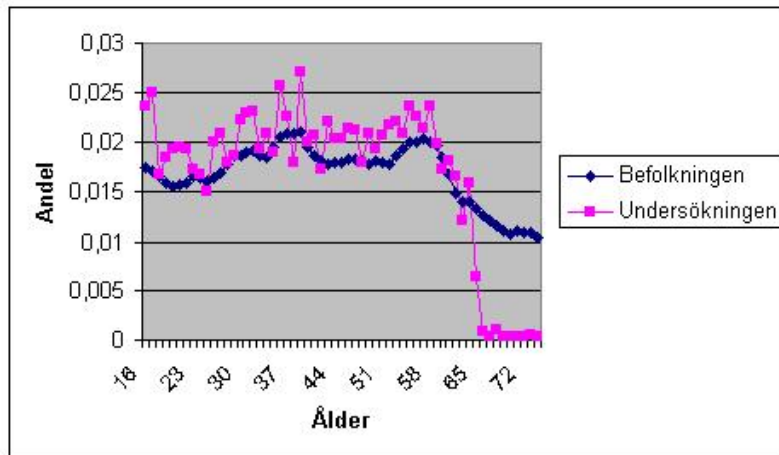
Resultatet från datorvaneundersökningen publiceras varje år, se [8], denna publicering kallas fortsättningsvis för rapporten. I början av rapporten ses en jämförelse mellan Sverige och några andra länder i EU, där denna undersökning också genomförs. Detta görs för att se hur IT-användningen skiljer sig åt mellan dessa länder och Sverige. Därefter ges en mer detaljrik redovisning av de frågor som ställs i Sverige. Dessa frågor presenteras uppdelade på olika redovisningsgrupper enligt Tabell 3.1. För varje redovisningsgrupp beräknas både den skattade andelen och den skattade totalen.

Urvalet av individer som får svara på frågor angående deras datorvanor är ett suburval av AKU-urvalet. Detta suburval kommer från april månad där individerna i vissa slumpceller får svara på de tillägsfrågor som ingår i datorvaneundersökningen, vilka slumpceller som ingår i suburvalet 2004 ses i Appendix D.

När vi tittar närmare på de som har deltagit i datorvaneundersökningen ser vi att åldrarna 65-74 år är underrepresenterade i urvalet, vilket ses i Figur 3.1. Denna situation uppkommer eftersom AKU är intresserade av personer som är aktivt arbetssökande medan datorvaneundersökningen är intresserade av alla personer i åldrarna 16-74 år. I det datamaterial som AKU redovisar till datorvaneundersökningen finns uppräkningsvikter för varje individ redovisade. Hur dessa beräknas beskrivs i Avsnitt 3.1 och 3.2.

Ålder	Kön		
Utbildningsnivå			
16-24	Totalt	Kvinnor	Män
25-34	Totalt	Kvinnor	Män
35-44	Totalt	Kvinnor	Män
45-54	Totalt	Kvinnor	Män
55-74	Totalt	Kvinnor	Män
Förgymnasial	Totalt	Kvinnor	Män
Gymnasial	Totalt	Kvinnor	Män
Eftergymnasial	Totalt	Kvinnor	Män
Totalt	Totalt	Kvinnor	Män

Tabell 3.1: *Redovisningsgrupper*



Figur 3.1: *Andel individer i olika åldrar*

3.1 Beskrivning av uppräkningsvikter och hjälpvariabler i AKU

Urvalet till AKU dras från en personnummersorterad version av SCB:s register över totalbefolkningen, RTB. Detta register kompletteras med uppgifter från SCB:s sysselsättningsregister, SREG⁶, som ger information om individen är förvärvsarbetande eller ej. Hela urvalet för AKU delas upp i tre olika delar, en för varje månad i kvartalet, där varje del består av åtta rotationsgrupper. För vart och ett av dessa tre “månadsurval” återkommer sju av rotationsgrupperna med tre månaders mellanrum medan en rotationsgrupp byts ut mot en rotationsgrupp med nya individer. Detta medför att varje urvalsindivid deltar i undersökningen totalt åtta gånger, en gång i kvartalet under två år. Vid årsskiftet varje år dras urvalet till AKU. Detta urval är tillräckligt stort för att täcka behovet av urvalsindivider under de kommande 12 månaderna, vilket i detta fall är april-mars. Det behov som behöver täckas av nya urvalsindivider uppkommer av att, som tidigare nämnts, 1/8 av urvalet tas bort från undersökningen. Urvalet dras med hjälp av stratifiering där populationen stratifieras efter variablerna kön, region, medborgarskap (svenskt/ej svenskt) samt förvärvsarbetande (förvärvsarbetande/ej förvärvsarbetande). Variabeln region följer länsindelningen från 1996 och antalet urvalsstrata blir därför $2 \cdot 24 \cdot 2 = 192$. Inom varje stratum sorteras individerna efter personnummer och därefter dras ett systematiskt urval inom varje stratum för att få en jämn spridning på ålder. Det dragna urvalet fördelas därefter på rotationsgrupperna som ligger till grund för månadsundersökningarna. De individer som ingår i urvalet delas in i paneler inom rotationsgrupperna för att lätt kunna genomföra månadsundersökningarna⁷.

Eftersom AKU genomförs varje månad och då alla individer i urvalet ej tillfrågas varje månad måste inklusionssannolikheterna räknas upp för individerna i varje “månadsurval” så att detta “månadsurval” representerar hela målpopulationen. Därefter används GREG för “månadsurvalet”. Detta innebär att för urvalsindivid k observeras (y_k, \mathbf{x}_k) , där y_k är ett mätvärde från undersökningen medan \mathbf{x}_k är en vektor med hjälpinformation.

För att beräkna skattningarna i AKU bildas två uppräkningsvikter för varje individ i urvalet med användning av olika hjälpinformation. För både sysselsatta och arbetslösa används hjälpvariabeln help3 som är en poststratifiering vilken delar in individerna efter en regionindelning. För de som är sysselsatta används även hjälpvariabeln help2 som är en poststratifiering efter kön, åldersgrupp och SNI-grupp, där SNI-grupperna anger vilken näringsgren som individen tillhör. För de arbetslösa individerna används även hjälpvariabeln help1 vilken är en poststratifiering efter kön, åldersgrupp och information om huruvida individen är arbetssökande enligt AMS:s arbetssöknaderegister eller ej. Hur poststratu-

⁶SREG ligger till grund för RAMS, registerbaserad arbetsmarknadsstatistik.

⁷För en närmare beskrivning av hur urvalet går till i AKU samt hur individerna delas in i olika paneler se [6].

mindelningen ser ut för help1, help2 och help3 beskrivs i Appendix C.

3.1.1 Arbetslösa

De som är arbetslösa och de som är sysselsatta får en uppräkningsvikt, $w_{U_k}^{(ams)}$, som erhålles genom att använda GREG med help1 och help3 som hjälpinformation. För att få fram $w_{U_k}^{(ams)}$ innehåller \mathbf{x}_k 41 nollor och två ettor. En etta anger vilken region som individen tillhör och den andra ettan anger vilket poststratum som individen tillhör enligt uppdelningen kön, ålder och arbetssökandestatus. \mathbf{t}_x kommer nu att vara en vektor med poststratumtotaler för hela populationen. Denna uppräkningsvikt beräknas alltså för både arbetslösa och sysselsatta men den används enbart för de som är arbetslösa och i Appendix B ser vi att dessa individer har akstatus 4. Vi sätter $w_{U_k}^{(AKU)} = w_{U_k}^{(ams)}$ då $k \in \{\text{akstatus}=4\}$.

3.1.2 Sysselsatta och Akstatus 1

Genom att använda GREG med help2 och help3 som hjälpinformation fås en uppräkningsvikt, $w_{U_k}^{(sys)}$, för arbetslösa och sysselsatta. Nu kommer \mathbf{x}_k att bestå av 168 nollor och två ettor, där en etta anger vilken region individen tillhör och den andra ettan anger vilket poststratum individen tillhör enligt uppdelningen kön, ålder, och SNI-grupp. \mathbf{t}_x anger poststratumtotalerna för hela populationen. $w_{U_k}^{(sys)}$ används för de som är sysselsatta, det vill säga de som har akstatus 1, 2 eller 3. Detta innebär att $w_{U_k}^{(AKU)} = w_{U_k}^{(sys)}$ då $k \in \{1 \leq \text{akstatus} \leq 3\}$.

3.1.3 Akstatus 5

De individer som har akstatus 5 ingår ej i arbetskraften och deras uppräkningsvikter måste därför beräknas på ett speciellt sätt. En indelning efter kön*ålder görs vilket ger 28 grupper, hur dessa grupper ser ut ses i Tabell E.1.

För att underlätta beskrivningen av hur vikterna skapas för de individer som har akstatus 5 använder vi oss av indikatorvariabeln I_{kg} som bildas på följande sätt:

$$I_{kg} = \begin{cases} 1 & \text{om individ } k \text{ har deltagit i undersökningen och tillhör grupp } g \\ 0 & \text{annars.} \end{cases}$$

Först skattas den totala befolkningen, antalet i hela befolkningen med akstatus 1

samt antalet i hela befolkningen som är sysselsatta, det vill säga de med akstatus 1, 2 eller 3, inom varje grupp genom att använda $w_{Uk}^{(sys)}$.

$$\widehat{\text{Bef}}_g = \sum_{k \in \{\text{akstatus} \geq 1\}} I_{kg} w_{Uk}^{(sys)}, \text{ för } g = 1, 2, \dots, 28.$$

$$\widehat{\text{Ak1}}_g = \sum_{k \in \{\text{akstatus} = 1\}} I_{kg} w_{Uk}^{(sys)}, \text{ för } g = 1, 2, \dots, 28.$$

$$\widehat{\text{Sys}}_g = \sum_{k \in \{1 < \text{akstatus} < 4\}} I_{kg} w_{Uk}^{(sys)}, \text{ för } g = 1, 2, \dots, 28.$$

Genom att använda $w_{Uk}^{(ams)}$ skattas antalet i hela befolkningen med akstatus 4 inom grupperna.

$$\widehat{\text{Ak4}}_g = \sum_{k \in \{\text{akstatus} = 4\}} I_{kg} w_{Uk}^{(ams)}, \text{ för } g = 1, 2, \dots, 28.$$

För att skatta antalet i hela befolkningen med akstatus 5 inom grupperna subtraheras summan av de skattade totalerna för akstatus 1, akstatus 4 och antalet sysselsatta i hela befolkningen från skattningen av den totala befolkningen. Därefter beräknas antalet svarande med akstatus 5 inom grupperna och sedan divideras den skattade totalen med akstatus 5 i gruppen med antalet svarande med akstatus 5. Detta innebär att alla individer, med akstatus 5, inom samma grupp kommer att få samma uppräkningsvikt.

$$\text{Lillan5}_g = \sum_{k \in \{\text{akstatus} = 5\}} I_{kg}, \text{ för } g = 1, 2, \dots, 28.$$

$$\text{Vikt5}_{kg} = \frac{\widehat{\text{Bef}}_g - (\widehat{\text{Ak1}}_g + \widehat{\text{Ak4}}_g + \widehat{\text{Sys}}_g)}{\text{Lillan5}_g}, \text{ för } g = 1, 2, \dots, 28.$$

För de individer som tillhör akstatus 5 sätts $w_{Uk}^{(AKU)} = \text{Vikt5}_{kg}$.

3.1.4 Åldersklassen 65-74 år

De individer som är 65-74 år delas in i poststratan efter variablerna kön, ålder (65-69 år, 70-74 år) samt födelseland (i Sverige/ej i Sverige). Antalet poststrata blir därför $2 \cdot 2 \cdot 2 = 8$ stycken. För att få reda på antalet i respektive poststratum i hela befolkningen används RTB. Uppräkningsvikten fås genom att dividera totala antalet i hela befolkningen inom varje poststratum med antalet svarande inom varje poststratum, denna uppräkningsvikt kallas för $w_{Uk}^{(AKU)}$.

3.2 Beskrivning av uppräkningsvikter och hjälpvariabler i datorvaneundersökningen

De individer som ingår i datorvaneundersökningen tar med sig de uppräkningsvikter som de erhöll i AKU, $w_{U_k}^{(AKU)}$. Vi kan däremot inte använda dessa vikter för att göra punktskattningar i datorvaneundersökningen eftersom vikterna är för små då de är framtagna med ett större urval än det som ingår i datorvaneundersökningen. Individerna i datorvaneundersökningen delas in i poststratan, ps , efter kön, ålder och akstatus, se Tabell E.2.

För att justera vikterna för individerna används $w_{U_k}^{(AKU)}$, dessa uppräkningsvikter summeras för de som ingår i datorvaneundersökningen och som tillhör samma poststratum.

$$\text{Uppant_dv}_{ps} = \sum_r I_{kps} w_{U_k}^{(AKU)}, \text{ för } ps = 1, 2, \dots, 28.$$

$$\text{där } I_{kps} = \begin{cases} 1 & \text{om individ } k \text{ har deltagit i datorvaneundersökningen och tillhör poststratum } ps \\ 0 & \text{annars.} \end{cases}$$

Uppräkningsvikterna summeras även för alla individer i AKU som tillhör samma poststratum för att få totalerna inom poststrata.

$$\text{Uppant_aku}_{ps} = \sum_{k \in \{ps \cap AKU\}} w_{U_k}^{(AKU)}, \text{ för } ps = 1, 2, \dots, 28.$$

För att få uppräkningsvikterna, $w_{U_k}^{(DV)}$, för individerna i datorvaneundersökningen multipliceras uppräkningsvikten som fås i AKU med kvoten mellan totalen, som erhålles från alla individer i AKU och totalen, som erhålles från individerna i datorvaneundersökningen, för det poststratum som individen tillhör.

$$w_{U_k}^{(DV)} = w_{U_k}^{(AKU)} * \frac{\text{Uppant_aku}_{ps}}{\text{Uppant_dv}_{ps}}, \text{ för } k \in \{\text{datorvaneundersökningen} \cap ps\}.$$

4 ANALYS

Från Avsnitt 3 såg vi att estimatorn i datorvaneundersökningen var komplex vilket gör variansskattningar svårberäknade. Vi kommer därför att börja detta avsnitt med att beräkna approximativa variansskattningar, detta görs i Avsnitt 4.1. För att hitta hjälpvariabler som ska användas för att ge bra skattningar i datorvaneundersökningen undersöker vi i Avsnitt 4.2 olika variabler som finns tillgängliga för alla individer som ingår i urvalet till datorvaneundersökningen. Hur vi ska använda oss av dessa hjälpvariabler utreder vi i Avsnitt 4.3. Avsnitt 4 avslutas i Avsnitt 4.4 med en jämförelse mellan de approximativa variansskattningarna och den estimator som vi beslutar oss för att använda i Avsnitt 4.3.

4.1 Approximativ variansskattning

Som vi såg i Avsnitt 3 används ett komplext skattningsförfarande och inom ramen för detta examensarbete har därför enbart approximativa variansskattningar beräknats.

Ett sätt att beräkna variansskattningarna på är att anta att urvalet är draget med OSU, vilket ger uppräkningsvikten d_k , samt att använda GREG med den poststratifiering som beskrivs i Tabell E.2 som hjälpinformation. Detta innebär att \mathbf{x}_k består av 27 nollor och en etta som anger vilket poststratum individ k tillhör och \mathbf{t}_x kommer att bestå av poststratumtotaler för hela populationen. Vi vet däremot inte dessa totaler utan använder därför $w_{Uk}^{(AKU)}$ som fås i Avsnitt 3.1 för att skatta dessa totaler. Genom att anta att urvalet är draget med OSU kommer alla individer att få samma inklusionssannolikhet vilket är felaktigt och $w_{Uk}^{(AKU)}$ ges därför som en vikt för varje individ i datorvaneundersökningen eftersom den korrekta inklusionssannolikheten för individen är en del av $w_{Uk}^{(AKU)}$. Vid användning av GREG kommer vi däremot att få felaktiga skattningar om $w_{Uk}^{(AKU)}$ enbart används som vikt för varje individ eftersom $w_{Uk}^{(AKU)}$ kommer att multipliceras med alla variabler för individen. Vi måste därför sätta $c_k = \frac{1}{w_{Uk}^{(AKU)}}$ i (2.4) för att få korrekta skattningar. Genom att göra på detta sätt erhåller vi samma punktskattningar som i rapporten, vi får även variansskattningar genom att använda CLAN. Punktskattningarna av de som använt en persondator under första kvartalet 2004 och de skattade standardavvikelserna anges i Tabell 4.1.

De skattade standardavvikelser som anges i Tabell 4.1 är approximativa skattningar eftersom vi inte har tagit hänsyn till att det finns en osäkerhet i beräkningen av $w_{Uk}^{(AKU)}$. Det är lätt att få intrycket att de sanna standardavvikelserna är större än de som anges i Tabell 4.1 eftersom vi ej har med osäkerheten i beräkningen av $w_{Uk}^{(AKU)}$ när vi beräknar dessa standardavvikelser. Detta behöver dock

Ålder Utbildningsnivå	Andel i procent			Absoluta tal		
	Totalt	Kvinnor	Män	Totalt	Kvinnor	Män
16-24	99 (0,4504)	98 (0,7553)	99 (0,5029)	972 947 (26 875,38)	475 047 (19 274,70)	497 901 (18 728,90)
25-34	95 (0,8524)	93 (1,3644)	96 (1,0326)	1 073 950 (27 321,49)	515 947 (19 439,21)	558 003 (19 198,46)
35-44	92 (1,2607)	93 (1,4279)	92 (2,0560)	1 189 914 (37 069,77)	587 385 (24 578,97)	602 529 (27 749,64)
45-54	87 (1,4293)	86 (2,1656)	87 (1,8679)	998 401 (35 720,40)	486 308 (23 401,33)	512 093 (26 987,49)
55-74	69 (4,0578)	70 (6,2131)	67 (5,1807)	1 313 724 (77 774,38)	682 318 (60 557,86)	631 407 (48 801,67)
Förgymnasial	74 (3,4632)	77 (5,2824)	71 (4,4922)	1 036 318 (55 046,86)	502 292 (48 801,67)	534 026 (25 449,46)
Gymnasial	84 (2,0638)	81 (3,4792)	87 (2,2841)	2 600 763 (92 812,56)	1 218 662 (70 610,67)	1 382 100 (60 235,41)
Eftergymnasial	96 (1,2498)	97 (0,8143)	95 (2,4921)	1 911 856 (66 685,53)	1 026 050 (51 112,76)	885 807 (42 830,43)
Totalt	86 (1,2638)	85 (1,9628)	86 (1,5988)	5 548 937 (81 871,33)	2 747 004 (63 076,14)	2 801 933 (52 194,98)

Tabell 4.1: Punktskattningar av de som använt en persondator under första kvartalet 2004 samt skattningar av standardavvikelserna

ej vara fallet eftersom vi skapar en osäkerhet i skattningarna genom att ge individerna i urvalet olika vikter. Vi kan därför inte säga om de approximativa variansskattningarna är under- eller överskattningar av de sanna varianserna.

4.2 Val av hjälpinformation

Enligt [10] ska hjälpvektorn i möjligaste mån

- (i) förklara variationen i svarssannolikheterna
- (ii) förklara variationen i de viktigaste undersökningsvariablerna
- (iii) identifiera de viktigaste redovisningsgrupperna.

Om (i) uppfylls bidrar detta till att den bias som uppkommer, för alla undersökningsvariabler, från bortfall minskar. Eftersom biasen för alla undersökningsvariabler blir lägre om hjälpvektorn uppfyller (i) är detta villkor särskilt viktigt då undersökningen innehåller många undersökningsvariabler som gör det svårt att utse en eller flera undersökningsvariabler som är viktigast. Då (ii) uppfylls minskar biasen som kommer från bortfallet för de viktigaste undersökningsvariablerna, detta bidrar även till att variansen i skattningarna av de viktigaste undersökningsvariablerna minskar. Problem med att uppfylla detta villkor uppkommer då det ej finns någon eller några självklara undersökningsvariabler som är viktigare än övriga undersökningsvariabler. De viktigaste undersökningsvariablerna identifieras därför genom en subjektiv bedömning. Om hjälpvektorn

	AktNat2=0	AktNat2=1
Fråga B	89,93%	89,02%
Fråga C	85,53%	85,68%
Svarsfrekvens	68,57%	51,41%

Tabell 4.2: *Medborgarskap*

uppfyller (iii) erhåller vi en lägre bias samt en lägre varians för skattningarna av de viktigaste redovisningsgrupperna. I detta avsnitt kommer vi att välja hjälpvariabler efter de tre kriterierna (i), (ii) och (iii) som anges ovan.

För att uppfylla (ii) väljer vi följande två undersökningsvariabler:

- Har du använt en persondator under de tre senaste månaderna
Denna fråga kallas för B, i enlighet med frågeformuläret, och är lika med 1 om individen svarar ja och 0 annars.
- Har du använt Internet under de tre senaste månaderna.
Denna fråga kallas för C, i enlighet med frågeformuläret, och är lika med 1 om individen svarar ja och 0 annars.

Orsaken till att just dessa två frågor väljs är att de är viktiga undersökningsvariabler som ger intressant information.

Vi undersöker närmare de variabler som är kända för alla individer som tillhör urvalet, även de som inte har svarat. Anledningen till att vi enbart undersöker dessa variabler är att vi då kan använda dem på olika sätt i skattningssteget, vi kan till exempel använda dem för att göra en poststratifiering eller använda GREG med hjälpinformation på olika sätt. Dessa variabler är AktNat, Alder, Hregg, Kommun, Kon, Lan, SenInvAr, UtbNiv samt UtbInr. För en närmare beskrivning av dessa variabler se Appendix B. Vi ser redan nu att de variabler som uppfyller (iii) är Alder, Kon och UtbNiva och vi undersöker därför hur bra variablerna uppfyller (i) och (ii).

4.2.1 Nationalitet

AktNat anger vilket medborgarskap individen har eller vilket år individen fick svenskt medborgarskap. Vi gör därför om denna variabel till en indikatorvariabel, AktNat2, som är 1 om individen har eller har haft utländskt medborgarskap och 0 annars. I Tabell 4.2 ser vi hur individerna, uppdelade efter AktNat2, har svarat på B och C samt hur svarsfrekvensen ser ut.

Vi ser ingen större skillnad på hur individerna har svarat på frågorna B och

	χ^2	p-värde
Fråga B	0,3334	0,5637
Fråga C	0,0066	0,9352
Svar	91,0914	<0,0001

Tabell 4.3: χ^2 -värden för variabeln *AktNat2*

	SenInvAr2=0	SenInvAr2=1
Fråga B	90,06%	88,07%
Fråga C	85,77%	83,94%
Svarsfrekvens	68,75%	51,17%

Tabell 4.4: *Invandringsår*

C när de är uppdelade efter medborgarskap, det ser däremot ut som om det finns en skillnad när det gäller svarsfrekvensen. Med hjälp av SAS beräknas därför χ^2 -statistikan för frågorna B och C samt för variabeln Svar som är en indikatorvariabel som är 1 om individen har svarat och 0 annars. I Tabell 4.3 anges χ^2 -värdena samt deras p-värden. Vi ser att medborgarskap inte har någon signifikant betydelse på frågorna B samt C och uppfyller därför inte (ii) men att variabeln uppfyller (i) eftersom den påverkar svarsfrekvensen. Eftersom variabeln *AktNat2* uppfyller villkor (i) kan det vara intressant att ha den som en hjälpvariabel.

Variabeln *SenInvAr* anger vilket år som är det senaste invandringsåret för individen och vi gör därför om även denna variabel till en indikatorvariabel, *SenInvAr2*, som är 1 om individen har invandrat och 0 annars. I Tabell 4.4 ser vi hur individerna har svarat på B och C samt svarsfrekvenserna om de delas upp efter variabeln *SenInvAr2*.

I Tabell 4.4 ser vi att information om huruvida individen har invandrat eller ej inte ger någon större skillnad när det gäller hur individerna har svarat på B och C men precis som för *AktNat2* ser det ut som om *SenInvAr2* påverkar om individerna svarar eller ej. I Tabell 4.5 anges χ^2 -värdena för denna variabel och vi får ett signifikant samband mellan svarsfrekvensen och variabeln *SenInvAr2* och variabeln uppfyller därför (i) men vi ser även att variabeln ej uppfyller (ii). Eftersom variablerna *AktNat2* och *SenInvAr2* är konstruerade på liknande sätt undersöker vi om det räcker med att enbart använda en av dessa variabler som hjälpvariabel. Korrelationen mellan variablerna är 0,77893 med p-värde <0,0001 vilket innebär att vi väljer att enbart ha *SenInvAr2* som hjälpvariabel eftersom denna ger ett högre χ^2 -värde.

	χ^2	p-värde
Fråga B	1,6606	0,1975
Fråga C	1,0312	0,3099
Svar	99,1296	<0,0001

Tabell 4.5: χ^2 -värden för variabeln *SenInvAr2*

Även om denna variabel ej har ett signifikanta samband med hur individerna svarar på frågorna B och C kan vi inte utesluta att det finns ett samband mellan variabeln och hur individerna svarar på övriga frågor. I början av detta avsnitt såg vi att då (i) uppfylls minskar biasen för alla undersökningsvariabler vilket gör denna variabel intressant att fortsätta undersöka.

4.2.2 Region

Precis som variabeln *Hregg* ger variablerna *Kommun* och *Lan* en regionindelning av Sverige. Vi väljer därför att inte undersöka variablerna *Kommun* och *Lan* eftersom vi inte vill ha en alltför detaljerad indelning. Hur individerna har svarat på B och C samt hur svarsfrekvensen ser beroende på vilken region de tillhör ses i Tabell 4.6.

	Hregg=1	Hregg=2	Hregg=3	Hregg=4	Hregg=5	Hregg=6
Fråga B	93,94%	92,96%	90,17%	86,72%	87,98%	82,83%
Fråga C	89,48%	89,65%	85,67%	82,46%	84,50%	77,32%
Svarsfrekvens	62,40%	59,41%	69,11%	66,81%	66,84%	68,97%

Tabell 4.6: *Regionuppdelning*

Vi ser att denna regionuppdelning är en möjlig variabel att ha som hjälpvariabel eftersom hur individerna har svarat på fråga B varierar från 93,94% till 82,83% beroende på vilken region som de tillhör. Vi undersöker därför hur χ^2 -värdena ser ut med avseende på denna variabel, detta ses i Tabell 4.7.

	χ^2	p-värde
Fråga B	38,8305	<0,0001
Fråga C	33,2704	<0,0001
Svar	31,8447	<0,0001

Tabell 4.7: χ^2 -värden för variabeln *Hregg*

I Tabell 4.7 ser vi att det finns signifikanta samband mellan variabeln *Hregg* och frågorna B och C samt svarsfrekvensen vilket innebär att variabeln uppfyller både villkoret (i) samt (ii). Vi vill däremot undersöka om vi kan göra en grövre indelning i olika regioner och vi ser därför vad som händer om vi slår ihop

Hregg=1 med Hregg=2 och Hregg=3 med Hregg=4, Hregg=5 samt Hregg=6. Den regionindelning som vi erhåller är nu Stockholm, Göteborg och Malmö som en region och övriga Sverige som en region. Hur individerna svarar uppdelade på denna grövre indelning ses i Tabell 4.8.

	Stockholm, Göteborg och Malmö	Övriga Sverige
Fråga B	93,49%	88,34%
Fråga C	89,56%	83,93%
Svarsfrekvens	60,98%	68,31%

Tabell 4.8: *Grövre indelning av Hregg*

Det ser ut som om vi fortfarande erhåller en skillnad då individerna delas upp efter denna indelning och vi undersöker detta vidare genom att se på χ^2 -värdena i Tabell 4.9 som visar att denna regionindelning uppfyller (i) och (ii) eftersom det finns signifikanta samband mellan denna grövre regionindelning och frågorna B och C samt svarsfrekvensen. Vi beslutar oss därför för att använda denna regionindelning istället för en mer detaljerad regionindelning.

	χ^2	p-värde
Fråga B	21,5218	<0,0001
Fråga C	19,0751	<0,0001
Svar	28,1883	<0,0001

Tabell 4.9: χ^2 -värden för den grövre indelningen av Hregg

4.2.3 Kön

I Tabell 4.10 ser vi hur individerna svarar på frågorna B och C samt hur svarsfrekvensen ser ut beroende på vilket kön de har. I Tabell 4.10 ser vi inga större skillnader mellan män och kvinnor men vi undersöker ändå hur χ^2 -värdena ser ut för denna variabel, dessa χ^2 -värden anges i Tabell 4.11. Vi ser att individens kön inte ser ut att påverka huruvida individerna svarar eller ej samt hur de har svarat på fråga B men på 2%-nivå fås ett signifikant samband mellan variabeln och fråga C. Tidigare har vi sett att denna variabel uppfyller (iii) och i Tabell 4.11 ser vi att variabeln ej uppfyller (i) men att den uppfyller en del av (ii) eftersom det finns ett samband mellan variabeln och fråga C och vi väljer därför att ha kvar och använda denna som hjälpvariabel. Den främsta orsaken till att vi beslutar oss för att behålla denna variabel är att den uppfyller (iii) samt att den enbart kan anta två värden vilket innebär att om vi vill göra en poststratifiering och lägga på denna variabel kommer den inte bidra till att skapa många ytterligare poststratum.

	Män	Kvinnor
Fråga B	90,54%	89,10%
Fråga C	86,91%	84,18%
Svarsfrekvens	65,37%	66,69%

Tabell 4.10: *Kön*

	χ^2	p-värde
Fråga B	2,0423	0,1530
Fråga C	5,4539	0,0195
Svar	1,0648	0,3021

Tabell 4.11: χ^2 -värden för variabeln *Kon*

4.2.4 Utbildningsnivå

Variabeln *UtbNiv* ger de utbildningsnivåer som finns med i rapporten och hur individerna från de olika utbildningsnivåerna har svarat på frågorna B och C samt respektive svarsfrekvens ses i Tabell 4.12. Det ser ut som om individens utbildningsnivå påverkar hur individerna svarar samt svarsfrekvensen. Vi undersöker därför denna variabel vidare genom att i Tabell 4.13 se hur χ^2 -värdena ser ut för denna variabel. I Tabell 4.13 ser vi att variabeln *UtbNiv* uppfyller villkoren (i) samt (ii) och tidigare har vi konstaterat att variabeln uppfyller villkoret (iii), det är därför ingen tvekan om att denna variabel ska användas som en hjälpvariabel.

	Förgymnasial utbildning	Gymnasial utbildning	Eftergymnasial utbildning
Fråga B	83,51%	87,29%	97,67%
Fråga C	78,76%	81,78%	95,52%
Svarsfrekvens	55,17%	67,71%	73,20%

Tabell 4.12: *Utbildningsnivå*

	χ^2	p-värde
Fråga B	123,3566	<0,0001
Fråga C	166,3669	<0,0001
Svar	112,0010	<0,0001

Tabell 4.13: χ^2 -värden för variabeln *UtbNiv*

Hur individerna svarar på frågorna B och C samt hur svarsfrekvensen ser ut beroende på vilken utbildningsinriktning, *UtbInr*, de har ses i Tabell 4.14. Då det finns en större skillnad på hur individerna har svarat beroende på vilken utbildningsinriktning de har undersöker vi hur χ^2 -värdena ser ut för denna variabel, detta ses i Tabell 4.15. Vi ser i Tabell 4.15 att variabeln *UtbInr* uppfyller

	UtbInr =0	UtbInr =1	UtbInr =2	UtbInr =3	UtbInr =4	UtbInr =5	UtbInr =6	UtbInr =7	UtbInr =8	UtbInr =9
Fråga B	85,73%	95,90%	95,58%	97,39%	98,65%	90,46%	78,13%	87,69%	81,74%	91,23%
Fråga C	81,10%	90,98%	92,92%	95,47%	98,65%	85,83%	75,00%	81,86%	74,78%	88,60%
Svarsfrekvens	58,83%	74,16%	66,86%	71,66%	64,91%	68,53%	75,29%	73,84%	64,61%	56,44%

Tabell 4.14: *Utbildningsinriktning*

(i) och (ii). Det finns skäl att tro att det finns ett samband mellan UtbNiv och UtbInr eftersom vissa utbildningsinriktningar kräver en viss utbildningsnivå⁸. Vi väljer därför att enbart ha med UtbNiv som hjälpvariabel, dels för att denna variabel ger högst χ^2 -värden men framförallt för att den uppfyller villkoret (iii). En annan orsak till att vi väljer utbildningsnivå framför utbildningsinriktning är att vi inte vill inte ha en alltför detaljerad indelning av individerna i undersökningen vilket UtbInr ger.

	χ^2	p-värde
Fråga B	103,72558	<0,0001
Fråga C	116,3523	<0,0001
Svar	93,1047	<0,0001

Tabell 4.15: χ^2 -värden för variabeln *UtbInr*

4.2.5 Ålder

Att fatta beslut angående om variabeln Alder ska vara med som hjälpvariabel eller ej är inte lika rättfram som de föregående variablerna eftersom individerna kan delas in i ett flertal olika åldersklasser. Om GREG används kan Alder användas direkt som en hjälpvariabel. För att uppfylla villkoret (iii) börjar vi med att dela in individerna efter de åldersklasser som används som redovisningsgrupper i rapporten, dvs 16-24 år, 25-34 år, 35-44 år, 45-54 år, 55-74 år, hur dessa olika åldersklasser har svarat på frågorna B och C samt hur svarsfrekvensen ser ut ses i Tabell 4.16.

	16-24 år	25-34 år	35-44 år	45-54 år	55-74 år
Fråga B	98,44%	94,66%	92,88%	86,67%	77,49%
Fråga C	97,03%	92,34%	89,12%	82,00%	68,62%
Svarsfrekvens	73,65%	70,63%	67,31%	68,00%	54,78%

Tabell 4.16: *Åldersindelningen (16-24 år, 25-34 år, 35-44 år, 45-54 år, 55-74 år)*

I Tabell 4.17 undersöker vi vilka χ^2 -värden som fås med denna åldersindelning. Vi ser att vi får signifikanta samband mellan denna åldersindelningen och frågorna B och C samt svarsfrekvensen, vi ser även i Tabell 4.16 att den åldersklass

⁸I detta fall är det inte givande att se på korrelationen mellan dessa två variabler eftersom UtbInr är en nominalvariabel.

	χ^2	p-värde
Fråga B	208,7245	<0,0001
Fråga C	278,9384	<0,0001
Svar	110,4138	<0,0001

Tabell 4.17: χ^2 -värden för åldersindelningen (16-24 år, 25-34 år, 35-44 år, 45-54 år, 55-74 år)

som skiljer sig mest från de övriga är den sista, det vill säga de som är 55-74 år. I Avsnitt 3 nämndes att de som är 65-74 år är underrepresenterade i datorvaneundersökningen och vi väljer därför att se hur dessa svarar på frågorna B och C jämfört med de som är 55-64 år, detta ses i Tabell 4.18.

	55-64 år	65-74 år
Fråga B	78,99%	50,00%
Fråga C	69,93%	44,74%

Tabell 4.18: Svartsfördelningen för individer som är 55-74 år

De som är 65-74 år drar ned siffrorna för åldersklassen 55-74 år men då dessa är underrepresenterade i urvalet påverkas procenttalen för åldersklassen 55-74 år inte mycket. Vi får därför procenttal för åldersklass 55-74 år som ligger nära de som fås för de som är 55-64 år.

Åldersindelningen (16-24 år, 25-34 år, 35-44 år, 45-54 år, 55-74 år) tar ej hänsyn till att de som är 65-74 år underrepresenterade samt tenderar att svara på ett annorlunda sätt än övriga åldrar. För att försöka komma tillrätta med detta problem undersöker vi hur individerna har svarat på frågorna B och C samt hur svarsfrekvensen ser ut om de delas upp på samma sätt som görs i den poststratifiering som används i rapporten för att beräkna skattningarna, detta ses i Tabell 4.19.

	16-34 år	35-54 år	55-64 år	65-74 år
Fråga B	96,43%	89,82%	78,99%	50,00%
Fråga C	94,53%	85,61%	69,93%	44,74%
Svarsfrekvens	72,01%	67,64%	64,23%	14,84%

Tabell 4.19: Åldersindelningen (16-34 år, 35-54 år, 55-64 år, 65-74 år)

Det ser ut som om individerna i de olika åldersklasserna skiljer sig från varandra när det gäller svarsfrekvensen samt hur de har svarat på frågorna B och C. Att svarsfrekvensen blir låg för åldersklassen 65-74 år beror på det vi observerade i Avsnitt 3 om att dessa åldrar är underrepresenterade, detta innebär att det ej finns många individer i dessa åldrar i urvalet och om några av dessa ej svarar påverkas svarsfrekvensen mycket. I Tabell 4.20 anges hur χ^2 -värdena ser ut.

	χ^2	p-värde
Fråga B	220,4997	<0,0001
Fråga C	257,6761	<0,0001
Svar	322,7463	<0,0001

Tabell 4.20: χ^2 -värden för åldersindelningen (16-34 år, 35-54 år, 55-64 år, 65-74 år)

Vi ser i Tabell 4.20 ser vi att denna åldersindelning uppfyller (i) samt (ii) och tidigare konstaterade vi att även (iii) uppfylls.

Med åldersindelningen (16-24 år, 25-34 år, 35-44 år, 45-54 år, 55-74 år) fick vi åldersklasser som liknade varandra när det gäller hur individerna svarade, med åldersindelningen (16-34 år, 35-54 år, 55-64 år, 65-74 år) får vi istället åldersklasser som skiljer sig från varandra. Om individerna delas in efter åldersindelningen (16-34 år, 35-54 år, 55-64 år, 65-74 år) tas hänsyn till att de som är 65-74 år tenderar att svara annorlunda än övriga individer. Ett av problemen med detta tillvägagångssätt är att de som är 65-74 år och som har svarat på frågorna i datorvaneundersökningen kommer att representera många individer i populationen vilket innebär att vad de svarar kommer att påverka skattningarna mycket samtidigt som dessa individer kanske ej är representativa för åldersgruppen. Ett annat problem är att om vi gör en poststratifiering med denna åldersindelning som en variabel kommer vi att i vissa poststrata erhålla få individer eftersom det inte finns många individer som är 65-74 år i urvalet. För att försöka komma tillrätta med det sista problemet ska vi se om det går att bredda åldersklassen 65-74 år, det vill säga om vi kan göra den undre åldersgränsen lägre för att på så sätt öka antalet individer. Vi vill dock att de i åldersklassen ska svara på liknande sätt som tidigare trots att vi ändrar den lägre åldersgränsen. I Tabell 4.19 ser vi hur de i denna åldersklass har svarat på frågorna B och C. Tabell 4.21 visar hur de i åldrarna 64-74 år, 63-74 år respektive 62-74 år har svarat på frågorna B och C.

	64-74 år	63-74 år	62-74 år
Fråga B	54,17%	57,86%	62,50%
Fråga C	47,92%	48,57%	52,00%

Tabell 4.21: Svartsfördelning för olika åldersklasser

Vi ser att trenden är att procenttalet ökar för varje ålder vi lägger till och frågan är nu hur stor den sista åldersklassen ska vara. Det finns inget konkret svar på denna fråga men vi bestämmer oss genast för att den nedre åldersgränsen ej ska vara 62 år eller lägre eftersom denna åldersindelning skiljer sig mycket från hur individerna svarade när åldersindelningen är 65-74 år. Vi observerar att vi inte får någon större skillnad mellan åldersklasserna 64-74 år och 63-74 år samt att dessa inte skiljer sig alltför mycket från åldersklassen 65-74 år. Eftersom vi vill ha en så låg gräns som möjligt väljer vi därför åldersklassen 63-74 år. Då

individerna i datorvaneundersökningen delas upp efter åldersindelningen (16-34 år, 35-54 år, 55-64 år, 65-74 år) fås de antal i varje åldersklass som anges i Tabell 4.22. Antalen i varje åldersklass som har deltagit i datorvaneundersökningen ses i Tabell 4.23 uppdelad efter åldersindelningen (16-34 år, 35-54 år, 55-62 år, 63-74 år).

	16-34 år	35-54 år	55-64 år	65-74 år
Antal som deltagit i datorvaneundersökningen	1 371	1 522	695	38

Tabell 4.22: *Antal individer, som ingår i datorvaneundersökningen, i varje åldersklass vid åldersindelningen (16-34 år, 35-54 år, 55-64 år, 65-74 år)*

	16-34 år	35-54 år	55-62 år	63-74 år
Antal som deltagit i datorvaneundersökningen	1 371	1 522	593	140

Tabell 4.23: *Antal individer, som ingår i datorvaneundersökningen, i varje åldersklass vid åldersindelningen (16-34 år, 35-54 år, 55-62 år, 63-74 år)*

Om individerna delas upp efter åldersindelningen (16-34 år, 35-54 år, 55-62 år, 63-74 år) ser vi i Tabell 4.24 hur de olika åldersklasserna svarar på frågorna B och C samt hur svarsfrekvensen ser ut. Precis som för åldersindelningen (16-34 år, 35-54 år, 55-64 år, 65-74 år) ser vi att denna modifierade åldersindelning (16-34 år, 35-54 år, 55-62 år, 63-74 år) påverkar hur individerna svarar och vi beräknar därför χ^2 -värden som ses i Tabell 4.25. Vid jämförelse av Tabell 4.20 och Tabell 4.25 ser vi att åldersindelningen (16-34 år, 35-54 år, 55-62 år, 63-74 år) ger högre χ^2 -värde på frågorna B och C samt ett lägre χ^2 -värde för svarsfrekvensen än åldersindelningen (16-34 år, 35-54 år, 55-64 år, 65-74 år), vi väljer därför att spara både åldersindelningen (16-34 år, 35-54 år, 55-64 år, 65-74 år) och (16-34 år, 35-54 år, 55-62 år, 63-74 år) för att undersöka hur de fungerar som hjälpinformation.

	16-34 år	35-54 år	55-62 år	63-74 år
Fråga B	96,43%	89,82%	82,12%	57,86%
Fråga C	94,53%	85,61%	73,36%	48,57%
Svarsfrekvens	72,01%	67,64%	64,95%	32,94%

Tabell 4.24: *Åldersindelningen (16-34 år, 35-54 år, 55-62 år, 63-74 år)*

4.2.6 Slutsats

Vi har beslutat oss för att använda variablerna SenInvAr2, Hregg(1+2, 3+4+5+6), Kon, UtbNiv samt Alder som hjälpvariabler när vi skapar en ny estimator.

	χ^2	p-värde
Fråga B	260,3372	<0,0001
Fråga C	315,5982	<0,0001
Svar	240,8338	<0,0001

Tabell 4.25: χ^2 -värden för åldersindelningen (16-34 år, 35-54 år, 55-62 år, 63-74 år)

4.3 Val av estimator

I föregående avsnitt beslutade vi oss för att använda variablerna SenInvAr2, Hregg(1+2, 3+4+5+6), Kon, UtbNiv samt Alder som hjälpinformation. Vi såg att variabeln SenInvAr2 enbart uppfyllde (i) och vi väljer därför att använda denna variabel för att bilda svarshomogenitetsgrupper medan de övriga variablerna kommer att användas i hjälpvektorn.

Variansskattningar beräknar vi genom att använda GREG-kommandot i CLAN. Inklusionssannolikheterna beräknas genom att använda den stratifiering som görs för att dra urvalet till AKU. Alla individer i datorvaneundersökningen är indelade i olika stratum utom de som är äldre än 64 år, dessa antar vi dras med OSU. För att få fram varje individs inklusionssannolikhet, π_k , beräknar vi antalet i urvalet i varje stratum och dividerar detta antal med det totala antalet i varje stratum, detta antal är känt. Vi skapar även svarshomogenitetsgrupper genom att dela in individerna efter variabeln SenInvAr2, vi får då två svarshomogenitetsgrupper som baseras på variabeln SenInvAr2 och stratumindelningen. Vi erhåller individens svarssannolikhet, θ_k , genom att använda dessa svarshomogenitetsgrupper. Skattningen av Y -totalen ser ut på följande sätt:

$$\hat{Y} = \sum_r w_{Uk} y_k \quad (4.1)$$

$$\text{där } w_{Uk} = d_k \left(1 + \left(\hat{\mathbf{t}}_x - \sum_r \frac{d_k \mathbf{x}_k}{\hat{\theta}_k} \right)' \left(\sum_r \frac{d_k \mathbf{x}_k \mathbf{x}_k'}{\hat{\theta}_k} \right)^{-1} \mathbf{x}_k \right). \quad (4.2)$$

Hjälpvariablerna som vi beslutade oss för att använda i Avsnitt 4.2 kan vi nu tänka oss att använda på följande olika sätt:

1. Individerna delas in i olika poststratum efter variablerna Kon, Hregg(1+2, 3+4+5+6), UtbNiv och åldersklass (16-34 år, 35-54 år, 55-62 år, 63-74 år) vilket innebär att vi får $2*2*3*4=48$ poststrata. Vi kallar detta alternativ för Fall 1.
2. Vi bildar en hjälpvektor med variablerna Kon, Hregg(1+2, 3+4+5+6), UtbNiv och åldersklass (16-34 år, 35-54 år, 55-62 år, 63-74 år) som gör

att varje individ har en hjälpvektor som består av $2+2+3+4=11$ element. Detta alternativ kallas för Fall 2.

3. Vi bildar en hjälpvektor med variablerna Kon, Hregg(1+2, 3+4+5+6), UtbNiv och ålder där ett element i hjälpvektorn kommer att ange individens ålder vilket ger $2+2+3+1=8$ element i varje individs hjälpvektor. Vi kallar detta alternativ för Fall 3.
4. Precis som i Fall 2 har varje individ en hjälpvektor som innehåller 11 element. Det som skiljer dessa fall åt är att vi i detta fall har åldersindelningen (16-34 år, 35-54 år, 55-64 år, 65-74 år). Detta alternativ kallas för Fall 4.
5. Detta fall liknar Fall 1 med skillnaden att vi gör poststratifieringen efter åldersindelningen (16-34 år, 35-54 år, 55-64 år, 65-74 år) vilket, precis som i Fall 1, ger $2*2*3*4=48$ poststrata. Vi kallar detta alternativ för Fall 5.

Det som skiljer de olika fallen från varandra är utseendet på hjälpvektorn, \mathbf{x}_k , detta innebär även att $\hat{\mathbf{t}}_x$ kommer att se ut på olika sätt beroende på vilket fall som används. För att skatta $\hat{\mathbf{t}}_x$ kan vi göra på olika sätt, vi kan till exempel använda individernas ursprungliga inklusionssannolikheter, vi kan även använda oss av $w_{U_k}^{(AKU)}$ eller $w_{U_k}^{(DV)}$ för att beräkna skattningarna av x-totalerna. I detta avsnitt använder vi oss av $w_{U_k}^{(DV)}$ för att beräkna dessa skattningar.

Med Fall 1 kommer varje individ att ha en hjälpvektor, $\mathbf{x}_k^{(1)}$, som innehåller 48 element där ett element är en etta och 47 element är nollor, ettan kommer i detta fall att representera vilken poststratumtillhörighet som individen har. Detta bidrar till att $\hat{\mathbf{t}}_x^{(1)}$ består av skattade poststratumtotaler. I detta fall får vi en skattning av Y-totalen genom att i Ekvation (4.2) använda $\mathbf{x}_k^{(1)}$ och $\hat{\mathbf{t}}_x^{(1)}$ i stället för \mathbf{x}_k respektive $\hat{\mathbf{t}}_x$.

Vid Fall 2 består hjälpvektorn, $\mathbf{x}_k^{(2)}$, av elva element där fyra element är ettor och sju element är nollor. En etta står för vilken region individen tillhör, den andra ettan står för vilket kön individen har, den tredje ettan står för vilken utbildningsnivå som individen har och den sista ettan står för vilken åldersklass, enligt den indelning som anges i Fall 2, som individen tillhör. Nu består $\hat{\mathbf{t}}_x^{(2)}$ av skattade grupptotaler för varje hjälpvariabel. Vi använder oss av $\mathbf{x}_k^{(2)}$ och $\hat{\mathbf{t}}_x^{(2)}$ i stället för \mathbf{x}_k och $\hat{\mathbf{t}}_x$ i Ekvation (4.2) för att erhålla en skattning av Y-totalen i Fall 2.

När vi använder Fall 3 har individen en hjälpvektor, $\mathbf{x}_k^{(3)}$, som innehåller åtta element där tre element är ettor, som anger individens kön, regiontillhörighet respektive utbildningsnivå, ett element anger individens ålder och fyra element är nollor. I detta fall består $\hat{\mathbf{t}}_x^{(3)}$ av skattningar av grupptotaler samt en skattning

av summan av individernas ålder. Genom att byta ut \mathbf{x}_k samt $\hat{\mathbf{t}}_x$ mot $\mathbf{x}_k^{(3)}$ och $\hat{\mathbf{t}}_x^{(3)}$ i Ekvation (4.2) erhåller vi en skattning av Y -totalen för detta fall.

Vid Fall 4 får varje individ en hjälpvektor, $\mathbf{x}_k^{(4)}$, som består av elva element där fyra av dessa är ettor och sju är nollor. Den första ettan står för vilken region individen tillhör, den andra ettan står för vilket kön individen har, den tredje ettan står för vilken utbildningsnivå individen har och den fjärde ettan står för den åldersklass som individen tillhör, enligt den åldersindelning som anges i Fall 4. Nu består $\hat{\mathbf{t}}_x^{(4)}$ av skattningar av grupptotaler. I detta fall erhåller vi en skattning av Y -totalen genom att i stället för \mathbf{x}_k och $\hat{\mathbf{t}}_x$ använda $\mathbf{x}_k^{(4)}$ respektive $\hat{\mathbf{t}}_x^{(4)}$ i Ekvation (4.2).

Med Fall 5 kommer varje individ att ha en hjälpvektor, $\mathbf{x}_k^{(5)}$, som innehåller en etta och 47 nollor, det vill säga totalt 48 element, där ettan anger vilket poststratum som individen tillhör enligt indelningen som anges i Fall 5. Vi erhåller $\hat{\mathbf{t}}_x^{(5)}$ genom att skatta poststratumtotalerna. Precis som i de övriga fallen får vi en skattning av Y -totalen genom att i Ekvation (4.2) byta ut \mathbf{x}_k och $\hat{\mathbf{t}}_x$ mot $\mathbf{x}_k^{(5)}$ samt $\hat{\mathbf{t}}_x^{(5)}$.

I Avsnitt 3 såg vi att de i åldrarna 65-74 år är underrepresenterade i undersökningen och i Avsnitt 4.2.5 försökte vi att slå ihop dessa individer med individer i närliggande åldrar och vi fick därför den åldersindelning som anges i Fall 1 och Fall 2. När vi nu ska skapa poststratifieringar kan vi få problem eftersom vi kan erhålla få eller inga individer som ingår i urvalet i vissa poststratum. Eftersom vi ej vill ha poststratum som inte har några individer börjar vi med att kontrollera detta för Fall 1 och Fall 5. I Fall 1 får vi individer i alla poststratum men i Fall 5 får vi poststratum som ej innehåller några individer och vi väljer därför att inte undersöka Fall 5 vidare.

Eftersom datorvaneundersökningen består av många frågor kommer vi att, precis som i föregående avsnitt, att koncentrera oss på två frågor, B och C. För att kunna jämföra de olika skattningsförfarandena med varandra kommer den empiriska variationskoefficienten, kallas fortsättningsvis för variationskoefficienten, att beräknas för varje skattning. Variationskoefficienten ser ut på följande sätt:

$$\frac{\sqrt{\hat{V}(\hat{Y})}}{\hat{Y}} \quad (4.3)$$

Uttryckt med ord är variationskoefficienten skattningen av standardavvikelsen för populationsstorheten Y dividerad med skattningen av Y . Det är önskvärt att denna koefficient är liten.

Vi börjar med att se på variationskoefficienterna och standardavvikelserna som fås genom att använda de skattningar som fås av CLAN för totalerna för fråga B, dessa ses i Tabell 4.26 där fetmarkering anger lägsta variationskoefficient respektive lägsta standardavvikelse för varje redovisningsgrupp.

	Variationskoefficienter, totaler				Standardavvikelser, totaler			
	Fall 1	Fall 2	Fall 3	Fall 4	Fall 1	Fall 2	Fall 3	Fall 4
Totalt, Aldk=1	0,02423	0,02898	0,02804	0,03083	24 420,87	29 875,58	29 245,51	29 823,00
Kvinnor, Aldk=1	0,03435	0,04798	0,04724	0,04951	16 785,49	24 773,50	24 677,29	24 476,33
Män, Aldk=1	0,03415	0,04938	0,04870	0,05102	17 737,71	25 412,03	25 366,65	24 132,98
Totalt, Aldk=2	0,02372	0,02946	0,03447	0,02770	24 740,29	30 008,82	36 720,15	29 925,98
Kvinnor, Aldk=2	0,03426	0,05263	0,05506	0,05063	17 261,20	27 249,03	29 778,44	27 483,94
Män, Aldk=2	0,03288	0,05075	0,05303	0,04828	17 723,79	25 417,26	27 811,70	25 957,35
Totalt, Aldk=3	0,02781	0,02828	0,03528	0,02770	32 135,60	32 796,98	39 087,35	32 211,46
Kvinnor, Aldk=3	0,03902	0,04879	0,05468	0,04851	22 210,39	28 792,28	30 862,09	28 467,04
Män, Aldk=3	0,03961	0,04884	0,05264	0,04816	23 224,90	27 822,52	28 605,10	27 738,44
Totalt, Aldk=4	0,03127	0,03155	0,03560	0,03065	32 338,87	32 488,35	36 109,80	31 758,88
Kvinnor, Aldk=4	0,04358	0,05110	0,05458	0,05127	22 133,08	26 680,03	28 094,72	26 895,70
Män, Aldk=4	0,04480	0,05353	0,05536	0,05223	23 578,16	27 181,52	27 658,53	26 713,61
Totalt, Aldk=5	0,05013	0,05328	0,05790	0,04820	63 432,58	68 278,01	74 682,11	62 508,36
Kvinnor, Aldk=5	0,08304	0,08662	0,08944	0,07877	51 154,38	51 253,88	53 220,40	47 247,32
Män, Aldk=5	0,05743	0,08503	0,08790	0,07874	37 289,18	58 650,48	61 065,46	54 880,17
Totalt, UtbNiv=1	0,03174	0,04313	0,04246	0,04232	33 230,79	45 387,82	44 671,91	44 604,26
Kvinnor, UtbNiv=1	0,05518	0,05952	0,05935	0,05975	27 410,88	29 550,25	29 467,23	30 133,85
Män, UtbNiv=1	0,03386	0,06634	0,06571	0,06413	18 636,05	36 877,03	36 508,72	35 247,47
Totalt, UtbNiv=2	0,02043	0,02244	0,02264	0,01962	52 245,79	57 744,05	58 260,14	50 989,08
Kvinnor, UtbNiv=2	0,03477	0,04132	0,04166	0,03745	41 054,44	49 735,12	50 075,47	45 840,03
Män, UtbNiv=2	0,02316	0,03465	0,03476	0,03196	31 889,40	47 462,37	47 670,99	43 933,82
Totalt, UtbNiv=3	0,01334	0,01568	0,01579	0,01742	25 371,46	29 706,08	29 915,44	32 940,46
Kvinnor, UtbNiv=3	0,01837	0,03513	0,03531	0,03724	18 520,92	36 461,70	36 682,48	37 991,59
Män, UtbNiv=3	0,01863	0,04328	0,04333	0,04119	16 634,75	37 087,77	37 079,81	35 871,32
Totalt	0,01220	0,01290	0,01282	0,01180	67179,44	71 198,94	70 793,39	65 423,96
Totalt, Kvinnor	0,01985	0,01995	0,02020	0,01833	53 318,18	54 614,88	55 300,76	50 374,02
Totalt, Män	0,01442	0,02072	0,02047	0,01952	40 667,07	57 663,31	56 965,50	54 554,50

Tabell 4.26: *Variationskoefficienter och standardavvikelser för totalerna för fråga B*

Tabell 4.27 är en sammanfattning av Tabell 4.26 och ger en överblick av hur de olika skattningsförfarandena presterar. I Tabell 4.27 ser vi att Fall 1 genererar flest variationskoefficienter som är lägst och om vi enbart ser på kolumnen som anger antalet variationskoefficienter som är lägst i Tabell 4.27 skulle Fall 1 väljas.

Fall	Antal variationskoefficienter som är lägst	Summan av alla variationskoefficienter
1	20	0,89622
2	0	1,16123
3	0	1,20442
4	7	1,12292

Tabell 4.27: *Sammanfattande tabell för totalerna för fråga B*

Vi vill däremot veta hur Fall 1-Fall 4 fungerar överlag, det kan till exempel vara så att ett skattningsförfarande inte ger många variationskoefficienter som är lägst men som i de flesta fall ger låga variationskoefficienter. Det kan även vara så att ett skattningsförfarande som ger många variationskoefficienter som

är lägst även kan ge många variationskoefficienter som är högst. Vi summerar därför variationskoefficienterna över redovisningsgrupperna för Fall 1, Fall 2, Fall 3, respektive Fall 4, vilket ses i kolumnen längst till höger i Tabell 4.27. Eftersom en lägre variationskoefficient innebär en mindre osäker skattning vill vi att summan ska vara så låg som möjligt, vilket vid jämförelse mellan Fall 1-Fall 4 fås då Fall 1 används. Genom att se på totalerna för fråga B fås resultatet att Fall 1 är bäst av Fall 1-Fall 4. Vi kan däremot inte dra några större slutsatser av detta eftersom vi enbart har tittat på en fråga, det kan vara så att Fall 1 enbart presterar bra på totalerna för fråga B. Vi väljer därför att även se på andelarna för fråga B, den sammanfattade tabellen för denna situation ses i Tabell 4.28.

Fall	Antal variationskoefficienter som är lägst	Summan av alla variationskoefficienter
1	12	0,63556
2	3	0,72192
3	0	0,72177
4	12	0,67903

Tabell 4.28: Sammanfattande tabell för andelarna för fråga B

Vi ser i Tabell 4.28 att Fall 1 är bäst när det gäller summan av alla variationskoefficienter samt lika bra som Fall 4 när vi ser på antalet variationskoefficienter som är lägst, vi undersöker därför hur situationen är om vi byter fråga. Vi börjar med att se på den sammanfattande tabellen, Tabell 4.29 för totalerna för fråga C.

Fall	Antal variationskoefficienter som är lägst	Summan av alla variationskoefficienter
1	21	0,97353
2	0	1,21772
3	0	1,25932
4	6	1,17859

Tabell 4.29: Sammanfattande tabell för totalerna för fråga C

I Tabell 4.29 ser vi att Fall 1 är bäst även i denna situation men vi tittar även på den sammanfattande tabellen med andelarna för fråga C, Tabell 4.30. Fall 1 fortsätter att vara bäst och vi bestämmer oss därför för att använda hjälpvariablerna på det sätt som beskrivs i Fall 1 med en poststratifiering efter kön, region, utbildningsnivå och ålder, vilket ses i Tabell E.3.

Fall	Antal variationskoefficienter som är lägst	Summan av alla variationskoefficienter
1	15	0,73752
2	0	0,81241
3	0	0,81252
4	12	0,76555

Tabell 4.30: Sammanfattande tabellen för andelarna för fråga C

4.4 Jämförelse av estimatorer

I detta avsnitt ligger fokus på att jämföra osäkerheten i punktskattningarna som fås i rapporten med de som fås om Fall 1 används. Vi börjar dock med att se hur punktskattningarna skiljer sig åt mellan då två olika skattningsförfarandena vilket visas för fråga B i Tabell 4.31. Vi ser att de två olika skattningsförfarandena ger liknande punktskattningar, största värdet på kvoten för totalerna är 1,10757 medan det minsta värdet är 0,95752, för andelarna är motsvarande värden 1,10757 och 0,97031. Vi ser även att båda summorna av kvoterna ligger nära 27 vilket även det pekar på liknande punktskattningar. Vid beräkning av 95%-iga konfidensintervall, för de punktskattningar som erhålles med Fall 1, för varje redovisningsgrupp erhåller vi för Fall 1 enbart konfidensintervall som innehåller den punktskattning som finns i rapporten både när det gäller totalerna samt andelarna.

	Totaler			Andelar		
	Rapporten	Fall 1	Rapporten/Fall 1	Rapporten	Fall 1	Rapporten/Fall 1
Totalt, Aldk=1	972 947	1 008 076	0,96515	0,98635	0,98806	0,99827
Kvinnor, Aldk=1	475 047	488 656	0,97215	0,98218	0,98447	0,99767
Män, Aldk=1	497 901	519 419	0,95857	0,99036	0,99146	0,99889
Totalt, Aldk=2	1 073 950	1 042 939	1,02973	0,94726	0,94822	0,99899
Kvinnor, Aldk=2	515 947	503 870	1,02397	0,92881	0,92829	1,00056
Män, Aldk=2	558 003	539 069	1,03512	0,96499	0,96763	0,99727
Totalt, Aldk=3	1 189 914	1 155 577	1,02971	0,92376	0,92332	1,00048
Kvinnor, Aldk=3	587 385	569 229	1,03190	0,92941	0,93243	0,99676
Män, Aldk=3	602 529	586 348	1,02760	0,91831	0,91465	1,00400
Totalt, Aldk=4	998 401	1 034 145	0,96544	0,86566	0,86909	0,99605
Kvinnor, Aldk=4	486 308	507 884	0,95752	0,85660	0,86193	0,99382
Män, Aldk=4	512 093	526 262	0,97308	0,87444	0,87610	0,99811
Totalt, Aldk=5	1 313 724	1 265 326	1,03825	0,68542	0,66017	1,03825
Kvinnor, Aldk=5	682 318	616 049	1,10757	0,70004	0,63205	1,10757
Män, Aldk=5	631 407	649 276	0,97248	0,67029	0,68926	0,97248
Totalt, UtbNiv=1	1 036 318	1 047 094	0,98971	0,73933	0,74702	0,98971
Kvinnor, UtbNiv=1	502 292	496 723	1,01121	0,76746	0,75895	1,01121
Män, UtbNiv=1	534 026	550 371	0,97030	0,71470	0,73657	0,97031
Totalt, UtbNiv=2	2 600 763	2 557 621	1,01687	0,84189	0,82792	1,01687
Kvinnor, UtbNiv=2	1 218 662	1 180 644	1,03220	0,81241	0,78706	1,03221
Män, UtbNiv=2	1 382 100	1 376 977	1,00372	0,86971	0,86649	1,00372
Totalt, UtbNiv=3	1 911 856	1 901 347	1,00553	0,96200	0,95671	1,00553
Kvinnor, UtbNiv=3	1 026 050	1 008 321	1,01758	0,96888	0,95214	1,01758
Män, UtbNiv=3	885 807	893 026	0,99192	0,95414	0,96192	0,99191
Totalt	5 548 937	5 506 062	1,00779	0,85654	0,84993	1,00778
Totalt, Kvinnor	2 747 004	2 685 688	1,02283	0,85482	0,83574	1,02283
Totalt, Män	2 801 933	2 820 375	0,99346	0,85824	0,86389	0,99346
Summa			27,15135			27,16228

Tabell 4.31: Punktskattningar för fråga B

I Tabell 4.32 ses punktskattningarna för fråga C. För denna fråga ser vi att punktskattningarna liknar varandra, vilket ses på kvoterna, som ligger nära ett, och

	Totaler			Andelar		
	Rapporten	Fall 1	Rapporten/Fall 1	Rapporten	Fall 1	Rapporten/Fall 1
Totalt, Aldk=1	959 060	991 994	0,96680	0,97227	0,97230	0,99997
Kvinnor, Aldk=1	468 630	481 541	0,97319	0,96892	0,97014	0,99874
Män, Aldk=1	490 430	510 453	0,96077	0,97550	0,97434	1,00119
Totalt, Aldk=2	1 050 155	1 020 145	1,02942	0,92628	0,92750	0,99868
Kvinnor, Aldk=2	507 806	496 236	1,02332	0,91416	0,91423	0,99992
Män, Aldk=2	542 349	523 909	1,03520	0,93792	0,94042	0,99734
Totalt, Aldk=3	1 151 334	1 113 568	1,03391	0,89381	0,88976	1,00455
Kvinnor, Aldk=3	562 259	541 033	1,03923	0,88966	0,88624	1,00386
Män, Aldk=3	589 074	572 535	1,02889	0,89781	0,89310	1,00527
Totalt, Aldk=4	945 321	971 562	0,97299	0,81963	0,81649	1,00385
Kvinnor, Aldk=4	456 043	471 582	0,96705	0,80329	0,80033	1,00370
Män, Aldk=4	489 278	499 980	0,97859	0,83548	0,83235	1,00376
Totalt, Aldk=5	1 176 056	1 107 890	1,06153	0,61360	0,57803	1,06154
Kvinnor, Aldk=5	590 825	519 989	1,13623	0,60618	0,53350	1,13623
Män, Aldk=5	585 231	587 901	0,99546	0,62127	0,62411	0,99545
Totalt, UtbNiv=1	978 881	988 076	0,99069	0,69836	0,70492	0,99069
Kvinnor, UtbNiv=1	479 273	469 805	1,02015	0,73229	0,71782	1,02016
Män, UtbNiv=1	499 608	518 271	0,96399	0,66864	0,69361	0,96400
Totalt, UtbNiv=2	2 428 315	2 350 211	1,03323	0,78606	0,76078	1,03323
Kvinnor, UtbNiv=2	1 108 032	1 057 277	1,04800	0,73866	0,70482	1,04801
Män, UtbNiv=2	1 320 283	1 292 933	1,02115	0,83081	0,81360	1,02115
Totalt, UtbNiv=3	1 874 730	1 866 872	1,00421	0,94332	0,93936	1,00422
Kvinnor, UtbNiv=3	998 259	983 299	1,01521	0,94264	0,92851	1,01522
Män, UtbNiv=3	876 471	883 574	0,99196	0,94409	0,95174	0,99196
Totalt	5 281 926	5 205 159	1,01475	0,81533	0,80348	1,01475
Totalt, Kvinnor	2 585 564	2 510 381	1,02995	0,80458	0,78119	1,02994
Totalt, Män	2 696 362	2 694 778	1,00059	0,82591	0,82542	1,00059
Summa			27,33647			27,34799

Tabell 4.32: Punktskattningar för fråga C

summan av kvoterna, som ligger nära 27. Även för fråga C innehåller alla 95%-iga konfidensintervall, som vi erhåller med Fall 1, rapportens punktskattning. Vi har i Tabellerna 4.31 och 4.32 sett att punktskattningarna inte skiljer sig mycket från varandra och vi undersöker därför vidare hur osäkra dessa punktskattningar är genom att se på de variationskoefficienter som de två olika skattningsförfarandena genererar. I Tabellerna 4.33 och 4.34 ses de variationskoefficienter som fås med hjälp av de approximativa variansskattningarna som beskrevs i Avsnitt 4.1 samt de variationskoefficienter som fås med Fall 1. I Tabellerna 4.33 och 4.34 ser vi att Fall 1 ger flest variationskoefficienter som är lägst samt att alla summor av kvoterna är större än 27 vilket indikerar att Fall 1 ger effektivare skattningar än vad de approximativa variansskattningarna gör. Vid en närmare undersökning av de olika redovisningsgrupperna ser vi att kvoterna för andelarna mellan rapporten och Fall 1 är 0,45759 respektive 0,56295 för frågorna B respektive C för redovisningsgruppen kvinnor med UtbNiv=3. Dessa kvoter blir låga eftersom de skattade standardavvikelserna i denna redovisningsgrupp blir lägre med de approximativa variansskattningarna än vid Fall 1. Varför vi får lägre standardavvikelser med de approximativa variansskattningarna än med Fall 1 är svårt att finna en förklaring till men anledningen kan vara att vi ger varje individ i datorvaneundersökningens urval en vikt då vi beräknar de approximativa variansskattningarna. Detta bidrar troligtvis till att vi för de approximativa variansskattningarna skapar negativa kovarianser mellan de skattade totalerna i denna redovisningsgrupp och de skattade antalen som har svarat ja på frågorna medan vi för Fall 1 troligtvis får positiva kovarianser.

	Totaler			Andelar		
	Rapporten	Fall 1	Rapporten/Fall 1	Rapporten	Fall 1	Rapporten/Fall 1
Totalt, Aldk=1	0,02762	0,02423	1,14027	0,00457	0,00393	1,16120
Kvinnor, Aldk=1	0,04057	0,03435	1,18119	0,00769	0,00663	1,16005
Män, Aldk=1	0,03762	0,03415	1,10153	0,00508	0,00440	1,15435
Totalt, Aldk=2	0,02544	0,02372	1,07242	0,00900	0,00908	0,99097
Kvinnor, Aldk=2	0,03768	0,03426	1,09983	0,01469	0,01566	0,93836
Män, Aldk=2	0,03441	0,03288	1,04644	0,01070	0,00969	1,10399
Totalt, Aldk=3	0,03115	0,02781	1,12025	0,01365	0,01204	1,13355
Kvinnor, Aldk=3	0,04185	0,03902	1,07245	0,01536	0,01324	1,16025
Män, Aldk=3	0,04606	0,03961	1,16274	0,02239	0,01989	1,12581
Totalt, Aldk=4	0,03578	0,03127	1,14413	0,01651	0,01422	1,16087
Kvinnor, Aldk=4	0,04812	0,04358	1,10420	0,02528	0,02113	1,19622
Män, Aldk=4	0,05270	0,04480	1,17626	0,02136	0,01909	1,11879
Totalt, Aldk=5	0,05920	0,05013	1,18093	0,05920	0,05013	1,18093
Kvinnor, Aldk=5	0,08875	0,08304	1,06885	0,08875	0,08304	1,06885
Män, Aldk=5	0,07729	0,05743	1,34577	0,07729	0,05743	1,34577
Totalt, UtbNiv=1	0,05312	0,03174	1,67375	0,04684	0,03174	1,47599
Kvinnor, UtbNiv=1	0,09718	0,05518	1,76098	0,06883	0,05518	1,24729
Män, UtbNiv=1	0,04766	0,03386	1,40740	0,06286	0,03386	1,85627
Totalt, UtbNiv=2	0,03569	0,02043	1,74705	0,02452	0,02043	1,20013
Kvinnor, UtbNiv=2	0,05794	0,03477	1,66626	0,04283	0,03477	1,23159
Män, UtbNiv=2	0,04358	0,02316	1,88190	0,02626	0,02316	1,13403
Totalt, UtbNiv=3	0,03488	0,01334	2,61391	0,01299	0,01334	0,97362
Kvinnor, UtbNiv=3	0,04982	0,01837	2,71205	0,00841	0,01837	0,45759
Män, UtbNiv=3	0,04835	0,01863	2,59580	0,02612	0,01863	1,40221
Totalt	0,01475	0,01220	1,20925	0,01475	0,01220	1,20925
Totalt, Kvinnor	0,02296	0,01985	1,15660	0,02296	0,01985	1,15660
Totalt, Män	0,01863	0,01442	1,29191	0,01863	0,01442	1,29191
Summa			38,73412			31,63642

Tabell 4.33: Variationskoefficienter för fråga B

	Totaler			Andelar		
	Rapporten	Fall 1	Rapporten/Fall 1	Rapporten	Fall 1	Rapporten/Fall 1
Totalt, Aldk=1	0,02808	0,02468	1,13763	0,00672	0,00607	1,10781
Kvinnor, Aldk=1	0,04121	0,03492	1,18006	0,01035	0,00991	1,04419
Män, Aldk=1	0,03828	0,03487	1,09792	0,00864	0,00718	1,20418
Totalt, Aldk=2	0,02600	0,02412	1,07776	0,01074	0,01049	1,02432
Kvinnor, Aldk=2	0,03791	0,03425	1,10683	0,01627	0,01636	0,99462
Män, Aldk=2	0,03571	0,03397	1,05116	0,01412	0,01328	1,06333
Totalt, Aldk=3	0,03219	0,02873	1,12063	0,01529	0,01393	1,09732
Kvinnor, Aldk=3	0,04376	0,04094	1,06901	0,01886	0,01794	1,05152
Män, Aldk=3	0,04704	0,04031	1,16693	0,02386	0,02117	1,12728
Totalt, Aldk=4	0,03754	0,03253	1,15394	0,01942	0,01802	1,07799
Kvinnor, Aldk=4	0,05090	0,04640	1,09705	0,02933	0,02568	1,14205
Män, Aldk=4	0,05486	0,04562	1,20252	0,02556	0,02532	1,00952
Totalt, Aldk=5	0,06850	0,05921	1,15694	0,06850	0,05921	1,15694
Kvinnor, Aldk=5	0,10795	0,09942	1,08582	0,10795	0,09942	1,08582
Män, Aldk=5	0,08408	0,06822	1,23250	0,08408	0,06822	1,23250
Totalt, UtbNiv=1	0,05555	0,03415	1,62660	0,04854	0,03415	1,42133
Kvinnor, UtbNiv=1	0,10132	0,05902	1,71668	0,07152	0,05902	1,21177
Män, UtbNiv=1	0,04897	0,03675	1,33259	0,06446	0,03675	1,75411
Totalt, UtbNiv=2	0,03730	0,02360	1,58084	0,02785	0,02360	1,18033
Kvinnor, UtbNiv=2	0,06121	0,03983	1,53667	0,05045	0,03983	1,26654
Män, UtbNiv=2	0,04548	0,02753	1,65226	0,02809	0,02753	1,02049
Totalt, UtbNiv=3	0,03546	0,01403	2,52762	0,01371	0,01403	0,97726
Kvinnor, UtbNiv=3	0,05100	0,01975	2,58189	0,01112	0,01975	0,56295
Män, UtbNiv=3	0,04879	0,01910	2,55418	0,02646	0,01910	1,38520
Totalt	0,01619	0,01351	1,19873	0,01619	0,01351	1,19873
Totalt, Kvinnor	0,02584	0,02180	1,18554	0,02584	0,02180	1,18554
Totalt, Män	0,01980	0,01629	1,21562	0,01980	0,01629	1,21562
Summa			37,64590			30,79925

Tabell 4.34: Variationskoefficienter för fråga C

Detta är ett bra tillfälle att påminna om det som nämdes i Avsnitt 4.1 om att vi ej vet om de approximativa variansskattningarna ger under- eller överskattningar av variansen. Om det är en underskattning är de sanna variationskoefficienterna större än vad som anges i Tabellerna 4.33 och 4.34. Detta innebär att Fall 1 i själva verket presterar ännu bättre än vad estimatorn som används i rapporten gör än vad Tabellerna 4.33 och 4.34 antyder. Om det i stället är en överskattning av variansen är variationskoefficienterna för rapporten längre än vad som anges i Tabellerna 4.33 och 4.34. Vi vet inte hur mycket lägre dessa variationskoefficienter är då vi har en överskattning men vi ser att det i de flesta redovisningsgrupper skiljer mycket mellan variationskoefficienterna då Fall 1 ger lägre variationskoefficienter. Detta innebär att även om de approximativa variansskattningarna ger en överskattning av variansen måste den sanna variansen i de flesta redovisningsgrupperna vara mycket lägre än de som fås med den approximativa variansskattningen för att prestera bättre precisionsmässigt än Fall 1.

Följande argument talar för Fall 1 jämfört med den nuvarande estimatorn:

- Fall 1 är uppbyggd på ett enklare sätt.
- I Fall 1 erhåller vi ett explicit uttryck för variansskattningarna.
- Beräkning av variansskattningar med Bootstrap, då den nuvarande estimatorn används, kräver mycket tid till simulering eftersom datorvaneundersökningen består av många frågor och redovisningsgrupper.

5 RESULTAT OCH SLUTSATSER

Syftet med detta examensarbete var att utreda den nuvarande estimatorn i datorvaneundersökningen. Estimatorn är komplex eftersom uppräkningsvikterna för varje individ i datorvaneundersökningen beräknas fram i två steg. Först beräknas uppräkningsvikter för varje individ som har svarat på frågorna i AKU och som tillhör det aktuella "månadsurvalet" som i detta fall är april 2004, på olika sätt beroende på vilka egenskaper de har. För att de individer som ingår i datorvaneundersökningen ska representera hela målpopulationen måste de uppräkningsvikter som individerna får i AKU räknas upp. Detta görs genom att göra en poststratifiering och använda alla individer som har svarat på frågorna i AKU för det aktuella "månadsurvalet". Denna estimator liknar ej någon av de vanligaste estimatorerna vilket ställer till problem vid beräkning av variansskattningar. Vi beräknade därför approximativa variansskattningar genom att använda GREG och för varje individ i datorvaneundersökningen lägga på de uppräkningsvikter som de erhåller i AKU.

Vi ville även undersöka om vi kan erhålla punktskattningar för datorvaneundersökningen på ett mindre komplicerat sätt och i Avsnitt 4.2 valde vi hjälpvariabler efter följande tre kriterier:

- (i) hjälpvektorn ska i möjligaste mån förklara variationen i svarssannolikheterna
- (ii) hjälpvektorn ska i möjligaste mån förklara variationen i de viktigaste undersökningsvariablerna
- (iii) hjälpvektorn ska i möjligaste mån identifiera de viktigaste redovisningsgrupperna.

Med dessa tre kriterier i åtanke undersöktes de variabler som är kända för alla individer som ingår i urvalet till datorvaneundersökningen. För var och en av dessa variabler tittade vi på hur individerna hade svarat på frågorna B och C, som beskrivs i Avsnitt 4.2, samt hur svarsfrekvensen såg ut om de delades in efter den aktuella variabeln, till exempel då de delas upp efter män och kvinnor. Vi beslutade oss för att använda variablerna kön, regiontillhörighet, utbildningsnivå samt åldersklasser för att bilda en hjälpvektor och variabeln medborgarskap för att skapa svarshomogenitetsgrupper. Hjälpvektorn kan skapas på olika sätt med hjälp av dessa variabler och efter att ha undersökt hur dessa olika fall presterar bestämde vi oss för att göra en poststratifiering, där ålder delas upp efter indelningen (16-34 år, 35-54 år, 55-62 år, 63-74 år), vilket gav $2*2*3*4=48$ poststrata. Vi avslutade med att jämföra denna nya estimator med de approximativa variansskattningarna och vi fann att den nya estimatorn

gav skattningar med lägre variationskoefficienter än vad de approximativa variansskattningarna gav. Med enbart denna information kunde vi inte dra slutsatsen att den nya estimatorn är bättre än den estimator som används eftersom vi ej vet om de approximativa variansskattningarna är under- eller överskattningar av de sanna varianserna. Resultaten indikerar dock att den nya estimatorn är bättre eftersom den för de flesta redovisningsgrupper ger mindre varians. Ett sätt att utvärdera detta närmare på är att beräkna variansskattningar med Bootstrap vilket skulle kunna ge ett mer definitivt svar. Vi kan däremot konstatera att den nya estimatorn är mer lätthanterlig och enklare att förstå än den gamla estimatorn. En annan fördel med den nya estimatorn är att vi kan beräkna variansskattningar på ett enkelt sätt med en känd programvara vilket antyder att det går att förbättra den estimator som används. Slutsatsen blir därför att den nuvarande estimatorn bör ersättas med den föreslagna estimatorn.

Referenser

- [1] Claes Andersson och Lennart Nordberg (1994): A method for variance estimation of non-linear functions of totals in surveys - theory and software implementation, *Journal of official statistics*, pp 395-405.
- [2] Claes Andersson och Lennart Nordberg (1998): *A user's guide to CLAN97 - a SAS-program for computation of point- and standard error estimates in sample surveys*, Statistiska Centralbyrån.
- [3] Victor M. Estevao och Carl-Erik Särndal (2000): A functional form approach to calibration. *Journal of official statistics*, pp 379-399.
- [4] Bernard W. Lindgren (1998): *Statistical theory*, fourth edition, Chapman & Hall.
- [5] Sixten Lunström m.fl. (2000): *Kalibrering av vikter - beskrivning av tekniken och de SCB fall den prövats i*, Statistiska Centralbyrån, 2000:1.
- [6] Hassan Mirza och Jan Hörngren (2001): *Urvals- och estimationsförfarandet i de svenska arbetskraftsundersökningarna (AKU)*, Statistiska Centralbyrån 2001:5.
- [7] Bengt Rosén (2005): *Teori och praktik för urvalsundersökningar*, Kompendium, januari 2005.
- [8] SCB (2004): *Privatpersoners användning av datorer och Internet 2004*, Statistiska Centralbyrån.
- [9] Peter Skatt (2004): *Community survey on ICT usage in households and by individuals 2004*, Statistiska Centralbyrån, Interim report.
- [10] Carl-Erik Särndal och Sixten Lunström (2005): *Estimation in surveys with nonresponse*, Wiley.
- [11] Carl-Erik Särndal, Bengt Swensson och Jan Wretman (2003): *Model assisted survey sampling*, Springer-Verlag.
- [12] Olle Wessberg och Åsa Greijer (2003): *Joint standard quality report for labour force and survey and regional labour market statistics*, Statistiska Centralbyrån.
- [13] www.scb.se: Arbetskraftsundersökningar, beskrivning av statistiken.
- [14] www.scb.se: IT bland individer, beskrivning av statistiken.

Appendix

A Pearsons χ^2 -test

Vid Pearsons χ^2 -test testas en observerad fördelning mot en ansatt fördelning. För att göra detta test används en statistika, X^2 , som mäter avvikelser mellan den observerade och den ansatta fördelningen. Antag att observationerna kan delas in i olika kategorier på det sätt som anges i Tabell A.1.

Kategori 1	Kategori 2					Totalt
	1	2	3	...	k_2	
1	x_{11}	x_{12}	x_{13}	...	x_{1k_2}	n_1
2	x_{21}	x_{22}	x_{23}	...	x_{2k_2}	n_2
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
k_1	$x_{k_1 1}$	$x_{k_1 2}$	$x_{k_1 3}$...	$x_{k_1 k_2}$	n_{k_1}

Tabell A.1: Indelning av observationerna

χ^2 statistikan definieras nu på följande sätt:

$$X^2 = \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} \frac{(x_{ij} - n_i \hat{p}_{0j})^2}{n_i \hat{p}_{0j}} \quad (\text{A.1})$$

där $\hat{p}_{0j} = \frac{1}{N} \sum_{i=1}^{k_1} x_{ij}$, $j = 1, 2, \dots, k_2$, $N = \sum_{i=1}^{k_1} n_i$.

Vi vill testa $H_0 : p_{ij} = \hat{p}_{0j}$ för alla i mot $H_A : p_{ij} \neq \hat{p}_{0j}$ för något i . Under H_0 är X^2 approx $\chi^2((k_1 - 1)(k_2 - 1))$ vilket innebär att stora värden på X^2 leder till en förkastning av H_0 .

A.1 Exempel på beräkning av χ^2 -statistikan

Vi ska i detta avsnitt beräkna χ^2 -statistikan för Fråga B i Tabell 4.2. H_0 : *Individens medborgarskap påverkar inte hur individerna svarar på fråga B*. De observerade frekvenserna anges i Tabell A.2 och de förväntade frekvenserna anges i Tabell A.3, där cellvärdet $326,36 = \frac{3207}{3626} * 369$.

	AktNat2=0	AktNat2=1	Totalt
B=0	323	46	369
B=1	2 884	373	3 257
Totalt	3 207	419	3 626

Tabell A.2: Observerade frekvenser för medborgarskap

	AktNat2=0	AktNat2=1	Totalt
B=0	326,36	42,64	369
B=1	2880,64	376,36	3 257
Totalt	3 207	419	3 626

Tabell A.3: Förväntade frekvenser för medborgarskap

I Tabell A.4 anges de χ^2 -värden som vi erhåller i varje cell, där $0,03460 = \frac{(323-326,36)^2}{326,36}$. χ^2 -statistikan, X^2 , är summan av cellvärdena som anges i Tabell A.4. Detta innebär att $X^2 = 0,03460 + 0,26484 + 0,00392 + 0,03000 = 0,33336$ och antalet frihetsgrader blir $(2-1)(2-1) = 1$ vilket, enligt en χ^2 -fördelningstabell, leder till p-värdet 0,5637 som anges i Tabell 4.3. Vi kan därför inte förkasta H_0 .

	AktNat2=0	AktNat2=1
B=0	0,03460	0,26484
B=1	0,00392	0,03000

Tabell A.4: χ^2 -värden i varje cell för medborgarskap

B Variabler

Variabel	Beskrivning
Akstatus=1	Ej i arbetskraften(utom Grad EX)
Akstatus=2	I arbete
Akstatus=3	Frånvarande
Akstatus=4	Arbetslös
Akstatus=5	Ej i arbetskraften(Grad EX, intagen, utomlands)
AktNat	Medborgarskap eller år för svenskt medborgarskap
Aldk=1	15-24 år
Aldk=2	25-34 år
Aldk=3	35-44 år
Ams=1	Arbetssökande enligt AMS:s arbetssökanderegister
Ams=2	Ej arbetssökande enligt AMS: arbetssökanderegister
Grad EX	Arbete utomlands mindre än 1 år framåt; Arbete utomlands minst 1 år framåt, men mindre än 2 år totalt; Utlandsvistelse mindre än 1 år framåt; Utlandsvistelse minst 1 år framåt, men mindre än 2 år totalt; Långvarigt sjuk i 1 år framåt eller mer inklusive förtidspensionär av hälsoskäl; Intagen för vård 1 år framåt eller mer
Hregg=1	Stockholm
Hregg=2	Göteborg/Malmö
Hregg=3	Större städer
Hregg=4	Södra mellanbygden
Hregg=5	Norra tätbygden
Hregg=6	Norra glesbygden
Kon=1	Man
Kon=2	Kvinna
SenInvAr	Senaste invandringsår
SNIGrupp=1	01, 02, 05, 10-14, 40-41, 45
SNIGrupp=2	15-26, 27, 36
SNIGrupp=3	28, 29-35
SNIGrupp=4	51, 52, 60-64
SNIGrupp=5	66-67, 70-72, 74
SNIGrupp=6	73, 75, 80, 85, 91
SNIGrupp=7	37, 50, 55, 90, 92-93, 95, 99
SNIGrupp=8	Övriga och de som är arbetslösa
UtbNiv=1	Förgymnasial utbildning
UtbNiv=2	Gymnasial utbildning
UtbNiv=3	Eftergymnasial utbildning

Tabell B.1: *Beskrivning av variabler*

Variabel	Beskrivning
SNI=01	Jordbruk, jakt och service i anslutning härtill
SNI=02	Skogsbruk och service till skogsbruk
SNI=05	Fiske, vattenbruk samt service i anslutning härtill
SNI=10	Kol- och torvutvinning
SNI=11	Utvinning av råpetroleum och naturgas samt service i anslutning härtill
SNI=12	Utvinning av uran- och toriummalm
SNI=13	Utvinning av metallmalmer
SNI=14	Annan mineralutvinning
SNI=15	Livsmedels- och dryckesvaruframställning
SNI=16	Tobaksvarutillverkning
SNI=17	Textilvarutillverkning
SNI=18	Tillverkning av kläder, pälsberedning
SNI=19	Garvning och annan läderberedning, tillverkning av reseffekter, handväskor, skodon o.d.
SNI=20	Tillverkning av trä och varor av trä, kork och rottning o.d. utom möbler
SNI=21	Massa-, pappers- och pappersvarutillverkning
SNI=22	Förlagsverksamhet; grafisk produktion och reproduktion av inspelningar
SNI=23	Tillverkning av stenkolsprodukter, raffinerade petroleumprodukter och kärnbränsle
SNI=24	Tillverkning av kemikalier och kemiska produkter
SNI=25	Tillverkning av kemikalier och kemiska produkter
SNI=26	Tillverkning av icke-metallsika mineraliska produkter
SNI=27	Stål- och metallframställning
SNI=28	Tillverkning av metallvaror utom maskiner och apparater
SNI=29	Tillverkning av maskiner som ej ingår i annan underavdelning
SNI=30	Tillverkning av kontorsmaskiner och datorer
SNI=31	Tillverkning av andra elektriska maskiner och artiklar
SNI=32	Tillverkning av teleprodukter
SNI=33	Tillverkning av precisionsinstrument, medicinska och optiska instrument samt ur
SNI=34	Tillverkning av motorfordon, släpfordon och påhängsvagnar
SNI=35	Tillverkning av andra transportmedel
SNI=36	Tillverkning av möbler, annan tillverkning
SNI=37	Återvinning
SNI=40	El-, gas-, ång- och hetvattenförsörjning
SNI=41	Vattenförsörjning
SNI=45	Byggverksamhet
SNI=50	Handel med och service av motorfordon; detaljhandel med drivmedel
SNI=51	Parti- och agenturhandel utom med motorfordon
SNI=52	Detaljhandel utom med motorfordon; reparation av hushållsartiklar och personliga artiklar
SNI=55	Hotell- och restaurangverksamhet
SNI=60	Landtransport; transport i rörsystem
SNI=61	Sjötransport
SNI=62	Lufttransport
SNI=63	Stödtjänster till transport; resebyråverksamhet
SNI=64	Post- och telekommunikationer
SNI=65	Finansförmedling utom försäkring och pensionsfondsverksamhet
SNI=66	Försäkring och pensionsfondsverksamhet utom obligatorisk socialförsäkring
SNI=67	Stödtjänster till finansiell verksamhet
SNI=70	Fastighetsverksamhet
SNI=71	Uthyrning av fordon och maskiner utan bemanning samt av hushållsartiklar och varor för personligt bruk
SNI=72	Databehandlingsverksamhet m.m.
SNI=73	Forskning och utveckling
SNI=74	Andra företagstjänster
SNI=75	Offentlig förvaltning och försvar, obligatorisk socialförsäkring
SNI=80	Utbildning
SNI=85	Hälso- och sjukvård, sociala tjänster; veterinärverksamhet
SNI=90	Avloppsrening, avfallshantering, renhållning o.d.
SNI=91	Intrassebevakning; religiös verksamhet
SNI=92	Rekreations-, kultur- och sportverksamhet
SNI=93	Annan serviceverksamhet
SNI=95	Förvärvsarbete i hushåll
SNI=99	Verksamhet vid internationella organisationer, utländska ambassader o.d.

Tabell B.2: Beskrivning av SNI

C Hjälpvektorer

Ams=1/ej Ams=2	Kön	Ålder	Poststratum
1	Män	16-24	1
1	Män	25-34	2
1	Män	35-44	3
1	Män	45-64	4
1	Kvinnor	16-24	5
1	Kvinnor	25-34	6
1	Kvinnor	35-44	7
1	Kvinnor	45-64	8
2	Män	16-24	9
2	Män	25-34	10
2	Män	35-44	11
2	Män	45-64	12
2	Kvinnor	16-24	13
2	Kvinnor	25-34	14
2	Kvinnor	35-44	15
2	Kvinnor	45-64	16

Tabell C.1: *Hjälpvariabeln Help1*

Kön	SNI grupp	Ålder	Post-stratum	Kön	SNI grupp	Ålder	Post-stratum	Kön	SNI grupp	Ålder	Post-stratum
Män	1	16-24	1	Män	6	35-39	49	Kvinnor	3	55-59	97
Män	1	25-29	2	Män	6	40-44	50	Kvinnor	3	60-64	98
Män	1	30-34	3	Män	6	45-49	51	Kvinnor	4	16-24	99
Män	1	35-39	4	Män	6	50-54	52	Kvinnor	4	25-29	100
Män	1	40-44	5	Män	6	55-59	53	Kvinnor	4	30-34	101
Män	1	45-49	6	Män	6	60-64	54	Kvinnor	4	35-39	102
Män	1	50-54	7	Män	7	16-24	55	Kvinnor	4	40-44	103
Män	1	55-59	8	Män	7	25-29	56	Kvinnor	4	45-49	104
Män	1	60-64	9	Män	7	30-34	57	Kvinnor	4	50-54	105
Män	2	16-24	10	Män	7	35-39	58	Kvinnor	4	55-59	106
Män	2	25-29	11	Män	7	40-44	59	Kvinnor	4	60-64	107
Män	2	30-34	12	Män	7	45-54	60	Kvinnor	5	16-24	108
Män	2	35-39	13	Män	7	55-64	61	Kvinnor	5	25-29	109
Män	2	40-44	14	Män	8	16-19	62	Kvinnor	5	30-34	110
Män	2	45-49	15	Män	8	20-24	63	Kvinnor	5	35-39	111
Män	2	50-54	16	Män	8	25-29	64	Kvinnor	5	40-44	112
Män	2	55-59	17	Män	8	30-34	65	Kvinnor	5	45-49	113
Män	2	60-64	18	Män	8	35-39	66	Kvinnor	5	50-54	114
Män	3	16-24	19	Män	8	40-44	67	Kvinnor	5	55-59	115
Män	3	25-29	20	Män	8	45-49	68	Kvinnor	5	60-64	116
Män	3	30-34	21	Män	8	50-54	69	Kvinnor	6	16-24	117
Män	3	35-39	22	Män	8	55-59	70	Kvinnor	6	25-29	118
Män	3	40-44	23	Män	8	60-64	71	Kvinnor	6	30-34	119
Män	3	45-49	24	Kvinnor	1	16-24	72	Kvinnor	6	35-39	120
Män	3	50-54	25	Kvinnor	1	25-29	73	Kvinnor	6	40-44	121
Män	3	55-59	26	Kvinnor	1	30-34	74	Kvinnor	6	45-49	122
Män	3	60-64	27	Kvinnor	1	35-39	75	Kvinnor	6	50-54	123
Män	4	16-24	28	Kvinnor	1	40-44	76	Kvinnor	6	55-59	124
Män	4	25-29	29	Kvinnor	1	45-49	77	Kvinnor	6	60-64	125
Män	4	30-34	30	Kvinnor	1	50-54	78	Kvinnor	7	16-24	126
Män	4	35-39	31	Kvinnor	1	55-59	79	Kvinnor	7	25-29	127
Män	4	40-44	32	Kvinnor	1	60-64	80	Kvinnor	7	30-34	128
Män	4	45-49	33	Kvinnor	2	16-24	81	Kvinnor	7	35-39	129
Män	4	50-54	34	Kvinnor	2	25-29	82	Kvinnor	7	40-44	130
Män	4	55-59	35	Kvinnor	2	30-34	83	Kvinnor	7	45-49	131
Män	4	60-64	36	Kvinnor	2	35-39	84	Kvinnor	7	50-54	132
Män	5	16-24	37	Kvinnor	2	40-44	85	Kvinnor	7	55-64	133
Män	5	25-29	38	Kvinnor	2	45-49	86	Kvinnor	8	16-19	134
Män	5	30-34	39	Kvinnor	2	50-54	87	Kvinnor	8	20-24	135
Män	5	35-39	40	Kvinnor	2	55-59	88	Kvinnor	8	25-29	136
Män	5	40-44	41	Kvinnor	2	60-64	89	Kvinnor	8	30-34	137
Män	5	45-49	42	Kvinnor	3	16-24	90	Kvinnor	8	35-39	138
Män	5	50-54	43	Kvinnor	3	25-29	91	Kvinnor	8	40-44	139
Män	5	55-59	44	Kvinnor	3	30-34	92	Kvinnor	8	45-49	140
Män	5	60-64	45	Kvinnor	3	35-39	93	Kvinnor	8	50-54	141
Män	6	16-24	46	Kvinnor	3	40-44	94	Kvinnor	8	55-59	142
Män	6	25-29	47	Kvinnor	3	45-49	95	Kvinnor	8	60-64	143
Män	6	30-34	48	Kvinnor	3	50-54	96				

Tabell C.2: *Hjälparvariabeln Help2*

Benämning	Poststratum
Stockholms län exkl Stockholms kommun	1
Stockholms kommun	2
Uppsala län	3
Södermanlands län	4
Östergötlands län	5
Jönköpings län	6
Kronobergs län	7
Kalmar län	8
Gotlands län	9
Blekinge län	10
Kristianstads län	11
Malmöhus län exkl Malmö kommun	12
Hallands län	13
Göteborg- och Bohus län exkl Göteborgs kommun	14
Älvsborgs län	15
Skaraborgs län	16
Värmlands län	17
Örebro län	18
Västmanlands län	19
Dalarnas län	20
Gävleborgs län	21
Västernorrlands län	22
Jämtlands län	23
Västerbottens län	24
Norrbottnens län	25
Malmö kommun	26
Göteborgs kommun	27

Tabell C.3: *Hjälpvariabeln Help3*

D Rotationsgrupp

		Rotationsgrupp																															
		1				2				3				4				5				6				7				8			
		Panelnummer																															
Mätvecka	Slumpgrupp	12	16	11	15	34	38	33	37	32	36	31	35	24	28	23	27																
1	10					DV	DV	DV	DV																								
1	11					DV	DV	DV	DV																								
1	12				DV	DV	DV	DV	DV																								
1	13				DV	DV	DV	DV	DV																								
1	14				DV	DV	DV	DV	DV																								
1	15					DV	DV	DV	DV																								
1	16					DV	DV	DV	DV																								
1	17					DV	DV	DV	DV																								
1	18					DV	DV	DV	DV																								
1	19					DV	DV	DV	DV																								
2	20					DV	DV	DV	DV																								
2	21					DV	DV	DV	DV																								
2	22				DV	DV	DV	DV	DV																								
2	23				DV	DV	DV	DV	DV																								
2	24				DV	DV	DV	DV	DV																								
2	25					DV	DV	DV	DV																								
2	26					DV	DV	DV	DV																								
2	27					DV	DV	DV	DV																								
2	28					DV	DV	DV	DV																								
2	29					DV	DV	DV	DV																								
3	30					DV	DV	DV	DV																								
3	31					DV	DV	DV	DV																								
3	32				DV	DV	DV	DV	DV																								
3	33				DV	DV	DV	DV	DV																								
3	34				DV	DV	DV	DV	DV																								
3	35					DV	DV	DV	DV																								
3	36					DV	DV	DV	DV																								
3	37					DV	DV	DV	DV																								
3	38					DV	DV	DV	DV																								
3	39					DV	DV	DV	DV																								
4	40					DV	DV	DV	DV																								
4	41					DV	DV	DV	DV																								
4	42				DV	DV	DV	DV	DV																								
4	43				DV	DV	DV	DV	DV																								
4	44				DV	DV	DV	DV	DV																								
4	45					DV	DV	DV	DV																								
4	46					DV	DV	DV	DV																								
4	47					DV	DV	DV	DV																								
4	48					DV	DV	DV	DV																								
4	49					DV	DV	DV	DV																								

Tabell D.1: Rotationsgrupper för AKU-urvalet april, 2004

E Tabeller

Kön	Ålder	Grupp, g	Kön	Ålder	Grupp, g
Män	16-19	1	Kvinnor	16-19	15
Män	20-24	2	Kvinnor	20-24	16
Män	25-29	3	Kvinnor	25-29	17
Män	30-34	4	Kvinnor	30-34	18
Män	35-39	5	Kvinnor	35-39	19
Män	40-44	6	Kvinnor	40-44	20
Män	45-49	7	Kvinnor	45-49	21
Män	50-54	8	Kvinnor	50-54	22
Män	55-59	9	Kvinnor	55-59	23
Män	60	10	Kvinnor	60	24
Män	61	11	Kvinnor	61	25
Män	62	12	Kvinnor	62	26
Män	63	13	Kvinnor	63	27
Män	64	14	Kvinnor	64	28

Tabell E.1: Gruppering för Akstatus 5

Kön	Akstatus	Ålder	ps	Kön	Akstauts	Ålder	ps
Män	1 eller 5	16-34	1	Kvinnor	3	35-54	15
Män	2	16-34	2	Kvinnor	4	35-54	16
Män	3	16-34	3	Män	1 eller 5	55-64	17
Män	4	16-34	4	Män	2	55-64	18
Kvinnor	1 eller 5	16-34	5	Män	3	55-64	19
Kvinnor	2	16-34	6	Män	4	55-64	20
Kvinnor	3	16-34	7	Kvinnor	1 eller 5	55-64	21
Kvinnor	4	16-34	8	Kvinnor	2	55-64	22
Män	1 eller 5	35-54	9	Kvinnor	3	55-64	23
Män	2	35-54	10	Kvinnor	4	55-64	24
Män	3	35-54	11	Män	1 eller 5	65-74	25
Män	4	35-54	12	Män	2, 3 eller 4	65-74	26
Kvinnor	1 eller 5	35-54	13	Kvinnor	1 eller 5	65-74	27
Kvinnor	2	35-54	14	Kvinnor	2, 3 eller 4	65-74	28

Tabell E.2: Poststratifieringen i datorvaneundersökningen

Hregg	Kön	UtbNiv	Ålder	ps	Hregg	Kön	UtbNiv	Ålder	ps
1 eller 2	Män	1	16-34	1	3,4,5 eller 6	Män	1	16-34	25
1 eller 2	Män	1	35-54	2	3,4,5 eller 6	Män	1	35-54	26
1 eller 2	Män	1	55-62	3	3,4,5 eller 6	Män	1	55-62	27
1 eller 2	Män	1	63-74	4	3,4,5 eller 6	Män	1	63-74	28
1 eller 2	Män	2	16-34	5	3,4,5 eller 6	Män	2	16-34	29
1 eller 2	Män	2	63-74	6	3,4,5 eller 6	Män	2	35-54	30
1 eller 2	Män	2	55-62	7	3,4,5 eller 6	Män	2	55-62	31
1 eller 2	Män	2	63-74	8	3,4,5 eller 6	Män	2	63-74	32
1 eller 2	Män	3	16-34	9	3,4,5 eller 6	Män	3	16-34	33
1 eller 2	Män	3	63-74	10	3,4,5 eller 6	Män	3	35-54	34
1 eller 2	Män	3	55-62	11	3,4,5 eller 6	Män	3	55-62	35
1 eller 2	Män	3	63-74	12	3,4,5 eller 6	Män	3	63-74	36
1 eller 2	Kvinnor	1	16-34	13	3,4,5 eller 6	Kvinnor	1	16-34	37
1 eller 2	Kvinnor	1	63-74	14	3,4,5 eller 6	Kvinnor	1	35-54	38
1 eller 2	Kvinnor	1	55-62	15	3,4,5 eller 6	Kvinnor	1	55-62	39
1 eller 2	Kvinnor	1	63-74	16	3,4,5 eller 6	Kvinnor	1	63-74	40
1 eller 2	Kvinnor	2	16-34	17	3,4,5 eller 6	Kvinnor	2	16-34	41
1 eller 2	Kvinnor	2	63-74	18	3,4,5 eller 6	Kvinnor	2	35-54	42
1 eller 2	Kvinnor	2	55-62	19	3,4,5 eller 6	Kvinnor	2	55-62	43
1 eller 2	Kvinnor	2	63-74	20	3,4,5 eller 6	Kvinnor	2	63-74	44
1 eller 2	Kvinnor	3	16-34	21	3,4,5 eller 6	Kvinnor	3	16-34	45
1 eller 2	Kvinnor	3	63-74	22	3,4,5 eller 6	Kvinnor	3	35-54	46
1 eller 2	Kvinnor	3	55-62	23	3,4,5 eller 6	Kvinnor	3	55-62	47
1 eller 2	Kvinnor	3	63-74	24	3,4,5 eller 6	Kvinnor	3	63-74	48

Tabell E.3: Poststratifieringen för Fall 1