

Matematisk statistik  
Stockholms universitet

# Modellering av prevalens som resultat av incidens och mortalitet

Sofia Qvarnström

Examensarbete 2006:4

## **Postadress:**

Matematisk statistik  
Matematiska institutionen  
Stockholms universitet  
106 91 Stockholm  
Sverige

## **Internet:**

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



Matematisk statistik  
Stockholms universitet  
Examensarbete 2006:4,  
<http://www.math.su.se/matstat>

# Modellering av prevalens som resultat av incidens och mortalitet

Sofia Qvarnström\*

april 2006

## Sammanfattning

Föreliggande rapport ger en redogörelse för konstruktion av en modell som beskriver prevalens (antalet sjuka under en tidsperiod) som resultat av mortalitet (antalet som avlider under en tidsperiod) och incidens (antalet individer som blir sjuka under en tidsperiod). Incidens och mortalitet är åldersberoende och modellframställningen inrättas under antagandet att sjukdomen är oåterkallelig. Som underlag till rapporten har datamaterial från "Kungsholmsstudien" använts. Kungsholmsstudien har begränsningar som gör att beräkningar får göras enligt två antaganden. Skattningar av ettåriga övergångssannolikheter samt övergångssannolikheter som sträcker sig över flera år presenteras i rapporten. Resultatet består av ett intervall för prevalensen som drivs av incidens och mortalitet.

---

\*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige. E-post: [qvarnen@hotmail.com](mailto:qvarnen@hotmail.com) Handledare: Anders Björksröm.



## Abstract

Before us lies a report that gives an account of the creation of a model that describes prevalence (the number of cases of a disease during a defined time interval) as a result of mortality (the number of deaths occurring during a defined time interval) and incidence (the number of new cases of disease occurring during a defined time interval). Incidence and mortality are both age dependent and the account of the model is established under the assumption that the disease is irrevocable. To serve as the basis of the report we have used data from “Kungsholmsstudien”. Kungsholmsstudien has limitations at which the calculation is done according to two different assumptions. Estimations of the one-year transition probabilities and also transition probabilities that reach more than one year are presented in the report. The result is an interval for the prevalence that is driven by incidence and mortality.

# FÖRORD

Denna rapport är utförd under HT 2005/ VT 2006 och utgör mitt examensarbete omfattande 20 poäng för magisterexamen i matematisk statistik på Stockholms Universitet.

Arbetet är ett delprojekt i forskningsprogrammet Tredjeliv. Initiativtagare är Birger Rapp, Professor, ekonomiska informationssystem, Institutionen för datavetenskap, Linköpings Universitet (LiU). Birger Rapp är programledare i Tredjeliv.

Docent Mårten Lagergren, Stiftelsen Stockholms Läns Äldrecentrum, har utformat idén till arbetet och har bidragit med datainsamling.

Rapporten har som utgångspunkt att ligga till grund för fortsatt analys av hälsoutvecklingen bland äldre. Malin Düring, student på Stockholms Universitet, ska med anknytning till denna rapport utföra sitt examensarbete i matematisk statistik under VT 2006.

Jag vill passa på att tacka alla som bistått mig under arbetets gång. Först och främst vill jag rikta ett tack till Birger Rapp, som gav mig chansen att genomföra mitt examensarbete som ett led i Tredjeliv och som alltid visade stort engagemang och handlingskraft. Jag vill också tacka Länsförsäkringar för det finansiella stöd jag erhållit. Mårten Lagergren vill jag tacka för uppslaget till arbetet och satsningen du gjorde i projektet. Störst tack ska min handledare Anders Björkström ha för förträfflig handledning genom assistans och stöd. Till sist vill jag ge Malin Düring ett tack för samarbetet och önska henne lycka till med sitt examensarbete.

# INNEHÅLL

1. INLEDNING .....	6
1.1 TREDJELIV OCH ÄLDRECENTRUM .....	6
1.2 BAKGRUND .....	6
1.3 MÅLSÄTTNING .....	7
1.4 DISPOSITION .....	8
2. DATA .....	9
2.1 KUNGSHOLMSSTUDIEN .....	9
2.2 BESKRIVNING AV DATA .....	10
2.3 DATABEARBETNING .....	10
3. METOD OCH RESULTAT .....	12
3.1 TILLSTÅND OCH ÖVERGÅNGAR .....	12
3.2 DET MARKOVSKA VILLKORET .....	13
3.2.1 Modellen i figur 2 .....	14
3.2.2 Modellen i figur 3 .....	15
3.3 INCIDENS OCH MORTALITET .....	16
3.3.1 Skattning av de ettåriga övergångssannolikheterna, $P_{ij}(t)$ .....	16
3.3.2 Minsta kvadrat metoden .....	18
3.3.3 Plottar av de ettåriga övergångssannolikheterna .....	19
3.4 PREVALENS .....	25
3.4.1 Beräkning av övergångssannolikheterna .....	25
3.4.2 Resultat .....	28
4. DISKUSSION .....	31
5. REFERENSER .....	32
6. APPENDIX .....	33
APPENDIX 1 .....	33
APPENDIX 2 .....	34
APPENDIX 3 .....	36

# 1. INLEDNING

## 1.1 Tredjeliv och Äldrecentrum

Föreliggande rapport har utförts på uppdrag av ”Tredjeliv”. Tredjeliv är ett forsknings- utvecklings- och kunskapsspridningsprogram kring alternativa och flexibla bostadslösningar som skall passa äldre utifrån perspektiven ekonomi, organisation och individ<sup>1</sup>. Birger Rapp, som är programledare i Tredjeliv, har frambringat ett samarbete med Mårten Lagergren, docent på Stiftelsen Stockholms Läns Äldrecentrum. Mårten Lagergren har formulerat problemställning och bidragit med datamaterial som möjliggjort nedanstående analys.

## 1.2 Bakgrund

Den demografiska utvecklingen i Sverige går mot en befolkningssammansättning med allt högre medelålder. Enligt svenska befolkningsprognoser kommer nästan var fjärde svensk år 2025 att vara över 65 år<sup>2</sup>. Den höga medelåldern kommer av att den genomsnittliga livslängden ökar. Men hur är det med hälsan? Kan den höga medelåldern kopplas samman med ljusare hälsoutveckling? Eller resulterar den höga medelåldern i en längre livstid med svagare hälsotillstånd?

Bristen på erfarenhetsmässiga analyser runt dessa frågeställningar ger stort utrymme för livliga diskussioner experter emellan. Vissa tror att ökad livslängd leder till att åldersrelaterad skröplighet och sjukdom gör insteg vid allt högre åldrar. Medan andra menar att omsorgsbehovet kommer att etablera sig vid ungefär samma ålder som nu, och att vårdbehovet därför kommer att föreligga under längre tid i takt med att livslängden ökar. Man brukar skilja på tre teorier i diskussionen om hälsoutvecklingen hos de äldre<sup>3</sup>:

- Teorin om komprimerad sjuklighet
- Teorin om uppskjuten sjuklighet
- Teorin om expanderad sjuklighet

---

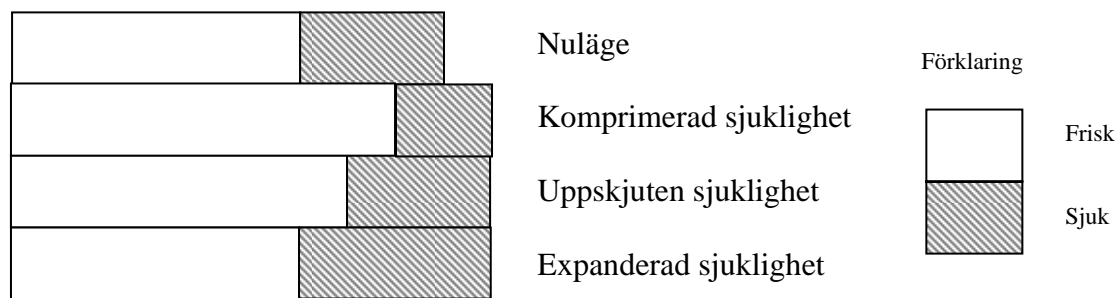
<sup>1</sup> Se [5] s.2

<sup>2</sup> Se [8] s.26

<sup>3</sup> Se [3] s.22



Figur 1 åskådliggör de tre teorierna.



**Figur 1:** visualisering av tre hälsoteorier

Teorin om komprimerad sjuklighet innebär att antalet ”friska” år ökar snabbare än det totala antalet levnadsår medan uppskjuten sjuklighet menar att antalet ”friska” år förblir oförändrat då livslängden ökar. Den mer pessimistiska teorin, expanderad sjuklighet, innebär att de levnadsår vi vinner med utökad livslängd består av år med sjukdom.

Diskussionerna kring hälsoutvecklingen bland de äldre skapar en önskan om en modell som skildrar antalet sjuka vid en viss ålder som resultat av två intensiteter; insjuknadsintensitet och dödlighetsintensitet. Modellen kan under dessa förhållanden nyttjas i prognoser om äldres hälsa och svara på vad som sker om den dödligheten ökar/minskar eller om insjuknandet ökar/minskar.

### 1.3 Målsättning

För att beskriva hälsa används tre mått.

Incidens – antalet nya fall under en viss tidsperiod

Mortalitet – antalet personer som avlider under en viss tidsperiod

Prevalens – antalet individer som är drabbade av sjukdomen under en viss tidsperiod.

Syftet med föreliggande rapport är att skapa en modell som beskriver prevalens som resultat av incidens och mortalitet. Modellen är tänkt som ett verktyg för att kunna iscensätta effekterna av olika antaganden om incidens och mortalitet. Frågeställningar om vilken av de tre hälsoutvecklingsteorier som bäst beskriver framtiden om den genomsnittliga åldern ökar/minskar eller om insjuknandet ökar/minskar kommer att kunna granskas med hjälp av modellen.

Incidens kan avse olika saker, såsom nedsatt funktionsförmåga, förekomst av sjukdom, intagning på sjukhem etc. I denna rapport kommer incidens att ämna behovet av hemtjänst och institutionsboende men tanken är att man ska kunna använda samma grundmodell då incidensen åsyftar något annat.

Modellframställningen kommer att inrättas under följande antaganden.

- Sjukdomen är oåterkallelig.
- Incidens och mortalitet beror av ålder.

Att incidens och mortalitet beror av ålder är tillfredsställande men en invändning kan vara könsberoendets frånvaro, av den orsaken att kvinnor och män skiljer sig åt i sjukdoms- och dödlighetsstatistik. Även om målet från början var att skilja könen åt så visade sig att datamaterialet inte gjorde det möjligt. Kvinnorna är kraftigt överrepresenterade i materialet och av den orsaken var en enskild analys av männens situation inte görlig. I denna analys väljer vi således att göra beräkningar oberoende av kön. Det kan diskuteras varför männen inte helt uteslöts ur modellen. Anledningen till att vi valde att behålla männen var att det redan så begränsade materialet skulle begränsas ytterligare och således försvåra möjligheten att åstadkomma ett resultat.

## 1.4 Disposition

I kapitel 2 ges en överblick av datamaterialet som börjar med en presentation av Kungsholmsprojektet och avslutar med avgränsningar och bearbetning av datamaterialet. Första delen i kapitel 3 presenterar en modellframställning. Därefter presenteras de teorier som utgör utgångspunkten för arbetet. Vidare redovisas metoder och resultat om vartannat i kapitlet. Rapporten avslutas med kapitel 4 där tankegångar kring eventuell fortsatt analys kring ämnet äger rum.

## 2. Data

### 2.1 Kungsholmsstudien

Mårten Lagergren på Äldrecentrum har gett bidrag med datamaterialet, som kommer från en studie kallad Kungsholmsstudien<sup>4</sup>. Kungsholmsstudien som startade 1987 är en longitudinell, populationsbaserad studie, där alla personer födda före 1913 och bosatta i Kungsholms församling i Stockholm inbjöds att delta. (Både personer i eget boende och institutionaliserade personer.) Syftet med projektet var att följa en stor grupp med äldre personer och studera sjukdomsförekomst bland dem. Sammanlagt 1810 personer ingick i studien. Efter den första inspektionen (fas 1) 1987-1988 deltog individerna, om de levde, i ytterligare 4 inspektioner (fas 3,4,5 och 6) fram till år 2000. Tabell 1 visar antalet medverkande per observationstillfälle.<sup>5</sup>

*Tabell 1: Antal medverkande vid varje fas*

Fas	Antal
1	1810
3	1105
4	685
5	427
6	274

En farhåga kan tänkas vara att datamaterialet från Kungsholmen inte är representativt för hela Sverige eftersom det kommer från en begränsad del av landet. I rapporten ”Transitions between modes of public care services and mortality among very old in Sweden”<sup>6</sup> tar Daniel Hallberg och Mårten Lagergren upp denna betänklighet. De redogör för en jämförelse av datamaterialet med data från Statistiska centralbyråns inkomstfördelnings- undersökning (HINK). De uppvisar svårigheter med jämförandet av datamaterialen men förevisar ändå ett visst stöd till att Kungsholmsstudien är representativ för Sverige.

---

<sup>4</sup> Se [6]

<sup>5</sup> Se [2]s. 5

<sup>6</sup> Se [2]s. 9

## 2.2 Beskrivning av data

Kungsholmsstudien har bland annat observerat ålder, kön, blodtryck, längd och vikt, funktionsförmåga utifrån dagliga aktiviteter (ADL) och behov av hemhjälp eller institutionsboende bland de äldre. Vid det första mättillfället genomfördes en omfattande klinisk undersökning efter en screeningfas. Vid varje annat tillfälle intervjuades alla deltagare av sjuksköterskor, undersöktes kliniskt av läkare och genomgick en psykologisk bedömning av psykologer<sup>7</sup>.

I förevarande analys kommer vi att inkludera ålder, kön, behov av hemhjälp och institutionsboende. Vi kommer att studera hur personerna i studien förflyttar sig mellan de olika tillstånden samtidigt som vi tar hänsyn till ålder och kön.

De olika tillstånden är

- Status 0: Ingen hemhjälp
- Status 1: Hemhjälp
- Status 2: Institutionsboende med dygnet runt hjälp
- Status 3: Död

Då övergång sker till status 3 (död) finns den exakta tiden då det inträffade noterat. Till de andra tillstånden vet vi inte när övergången skedde bara att det har skett någon gång sen det senaste mättillfället. Vi kommer längre fram i rapporten visa på svårigheter i och med denna ofullkomlighet.

## 2.3 Databearbetning

Mårten Lagergren på Äldrecentrum har motiverat en hopslagning av Status 1 (Hemhjälp) och Status 2 (Institutionsboende). De möjliga tillstånden är således

- Status 1: Inget behov av hjälp
- Status 2: Behov av hjälp
- Status 3: Död

---

<sup>7</sup> Se [6]

För enkelhetens skull kallas fortsättningsvis tillstånden

- Status 1: Frisk
- Status 2: Sjuk
- Status 3: Död

Visst bortfall p.g.a. defekter i datamaterialet reducerade antalet medverkande. Tabell 2 redovisar antalet medverkande efter reduceringen.

***Tabell 3: Antal medverkande vid varje fas efter reducering***

Fas	Antal
1	1588
3	981
4	599
5	359
6	78

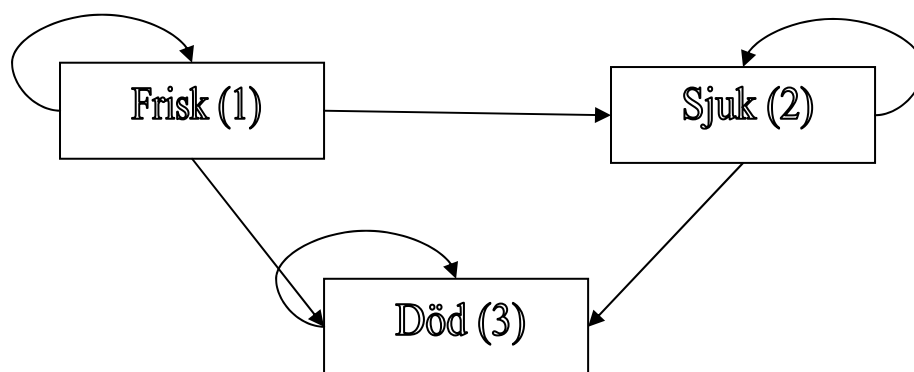
## 3. METOD OCH RESULTAT

### 3.1 Tillstånd och övergångar

För att bringa reda i hälsoutvecklingen ska denna rapport resultera i en modell som beskriver prevalens (antalet sjuka vid en viss tidpunkt) och hur den beror av incidens (antalet inträffade nya sjukdomsfall under en viss period) och mortalitet. Detta resulterar i att blott tre tillstånd finns tillgängliga i vår modell. Modellframställningen kommer att inrättas under följande antaganden.

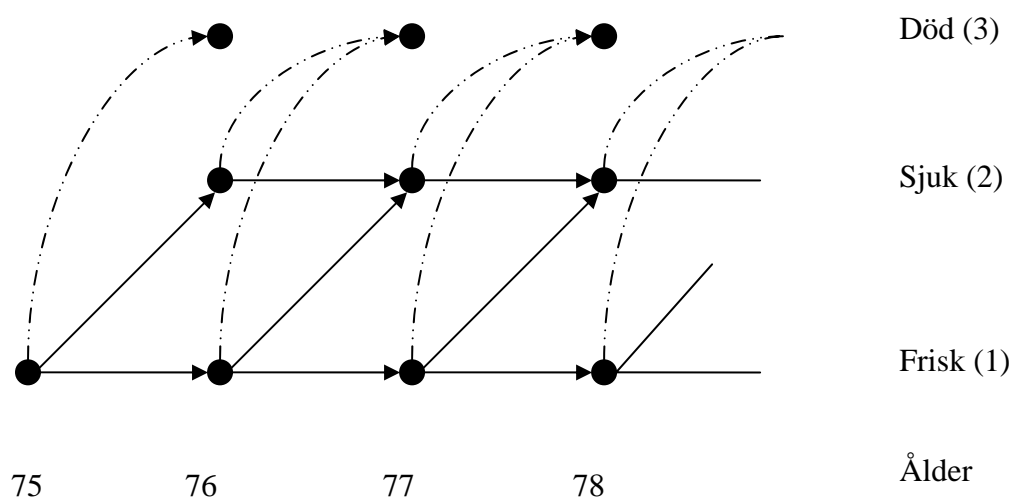
- Sjukdomen är oåterkallelig.
- Incidens och mortalitet beror av ålder.
- Vid 75 år är personerna friska.

Låt oss betrakta en modell med tre tillstånd där 1 = "frisk", 2 = "sjuk" och 3 = "död", där möjliga övergångarna mellan de tre tillstånden är  $1 \rightarrow 1$ ,  $1 \rightarrow 2$ ,  $1 \rightarrow 3$ ,  $2 \rightarrow 2$ ,  $2 \rightarrow 3$ . Figur 2 åskådliggör modellen.



*Figur 2: Modellframställning utan ålderberoende*

Denna modell bygger på att processen är konstant över tiden. Naturligtvis är den konstanta hypotesen orrealistisk, eftersom övergångarna mellan tillstånden beror av åldern på personen. En person som är 75 år besitter inte samma risk som en 95-åring att dö nästa år. Figur 3 visar en åldersberoende modell:



**Figur 3:** Modellframställning med åldersberoende

Begynnelseåldern är 75 och de tre tillstånden Frisk, Sjuk och Död återfinns i varje åldersperiod utom vid ålder 75 eftersom personerna antas friska vid begynnelseåldern.

### 3.2 Det Markovska villkoret

Markovmodeller är av väsentlig betydelse i matematisk modellering. I stort sätt kan man säga att en Markov modell är en modell där ”framtiden är oberoende av det förflutna, givet nutiden”. Denna egenskap förenklar studien av en process och tillåter ofta lösningar på problem som annars inte är lösbara.

Betrakta en tidsdiskret stokastisk process  $\{Y(s), s = 0, 1, 2, \dots\}$  med ett ändligt tillståndsrum  $\xi$ . Vi betecknar uppsättningen övergångar med  $\mathcal{Q}$ . Då  $Y(s) = i$  är processen i tillstånd  $i$  vid tidpunkten  $s$ . Anta att om processen är i tillstånd  $i$  så finns det en sannolikhet,  $P\{X(s) = j | X(t) = i\}$ , att processen är i tillstånd  $j$  vid tidpunkten  $s$ . Den stokastiska processen  $\{Y(s), s = 0, 1, 2, \dots\}$  är en Markov kedja om

$$P\{Y(s) = j | Y(t) = i, Y(t_{n-1}) = i_{n-1}, \dots, Y(1) = i_1, Y(0) = i_0\} = P\{X(s) = j | X(t) = i\}^8$$

<sup>8</sup> Se [4]s. 181

Detta betyder att tillståndet i nästa period endas beror på tillståndet i nuvarande period (det Markovska villkoret).

De betingade sannolikheterna

$$P\{X(s) = j | X(t) = i\}, \quad 0 \leq t \leq s \quad i, j \in \xi$$

kallas övergångssannolikheter och betecknas

$$P_{ij}(t, s) = P\{(X(s) = j | X(t) = i)\}$$

På grund av de Markovska villkoret är dessa sannolikheter konstanta.

Eftersom sannolikheter är icke-negativa och eftersom processen måste göra en övergång till något tillstånd har övergångssannolikheterna följande egenskaper:

$$\begin{aligned} 0 \leq P_{ij}(t, s) \leq 1 & \quad \text{för alla } i, j; \quad 0 \leq t \leq s \\ \sum_{j \in \xi} P_{ij}(t, s) = 1 & \quad \text{för alla } i; \quad 0 \leq t \leq s \end{aligned} \quad ^9$$

Vi låter  $P$  vara en matris med de ettåriga övergångssannolikheterna. Således har  $P$  följande utseende:

$$P = \begin{pmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & \ddots & \vdots \\ P_{m1} & \cdots & P_{mn} \end{pmatrix}$$

### 3.2.1 Modellen i figur 2

Låt oss återvända till vår ovan nämnda modell (Figur 2). Där är tillståndsrummet  $\xi = \{1, 2, 3\}$ .

Vidare är mängden möjliga övergångar  $\mathcal{G} = \{(1,1), (1,2), (1,3), (2,2), (2,3)\}$ . Vi presenterar de fem olika sannolikhetsövergångarna.

Givet att man är frisk så är

- $P_{11}$  = sannolikheten att man fortsätter vara frisk
- $P_{12}$  = sannolikheten att man är sjuk
- $P_{13}$  = sannolikheten att man är död

---

<sup>9</sup> Se [4] s. 181



Givet att man är sjuk så är

- $P_{22}$  = sannolikheten att man fortsätter vara sjuk
- $P_{23}$  = sannolikheten att man är död

Vår modell i figur 2 har den ettåriga övergångssannolikhetsmatrisen:

$$P = \begin{pmatrix} P_{11} & P_{12} & P_{13} \\ 0 & P_{22} & P_{23} \\ 0 & 0 & 1 \end{pmatrix}$$

Dessa sannolikheter är desamma oavsett vilken ålder personen är i. Nedan ska vi presentera modellen i figur 3 som är åldersberoende.

### 3.2.2 Modellen i figur 3

Låt oss nu betrakta modellen i figur 3. På samma sätt som i figur 2 har modellen

tillståndsrummet  $\xi = \{1, 2, 3\}$  och mängden möjliga övergångar är  $\mathcal{G} = \{(1,1), (1,2), (1,3), (2,2), (2,3)\}$ .

Eftersom den här modellen är åldersberoende har sannolikhetsövergångarna följande utseende och betydelse.

Givet att man är frisk vid tidpunkten  $t$ , ( $t = 75, 76, 77, \dots$ ) så är

- $P_{11}(t+1)$  = sannolikheten att man fortsätter vara frisk i nästa punkt
- $P_{12}(t+1)$  = sannolikheten att man är sjuk i nästa punkt
- $P_{13}(t+1)$  = sannolikheten att man är död i nästa punkt

Givet att man är sjuk vid tidpunkten  $t$ , ( $t = 75, 76, 77, \dots$ ) så är

- $P_{22}(t+1)$  = sannolikheten att man fortsätter vara sjuk i nästa punkt
- $P_{23}(t+1)$  = sannolikheten att man är död i nästa punkt

Matriserna med ettårsövergångssannolikheter har följande utseende:

$$P[t] = \begin{pmatrix} P_{11}[t] & P_{12}[t] & P_{13}[t] \\ 0 & P_{22}[t] & P_{23}[t] \\ 0 & 0 & 1 \end{pmatrix} \quad t = 1, 2, 3, \dots, N$$

Modellen i figur 3 har konstanta övergångssintensiteter och således är modellen en Markovmodell.

Nu har vi redovisat olika tillstånd och övergångar i vår modell och har på det sättet överblicken som behövs för att kunna fördjupa oss och efteråt kunna framhålla användningen av vår modell.

### 3.3 Incidens och mortalitet

Titeln på förevarande rapport lyder: *Modellering av prevalens som resultat av incidens och mortalitet*. I detta avsnitt betraktar vi incidens och mortalitet och i nästkommande avsnitt prevalens.

För att beräkna incidensen vid  $t$  fordras den ettåriga övergångssannolikheten  $P_{12}(t)$ . Likaledes krävs övergångssannolikheterna  $P_{13}(t)$  och  $P_{23}(t)$  vid beräkning av mortaliteten givet att man är frisk respektive mortaliteten givet att man är sjuk.

#### 3.3.1 Skattning av de ettåriga övergångssannolikheterna, $P_{ij}(t)$

I föregående avsnitt betonades nödvändighet att veta de ettåriga övergångssannolikheterna. Nu ska vi bringa klarhet i hur skattningen av dessa ska ske. De ettåriga övergångssannolikheterna skattas med antalet inträffade fall genom antalet möjliga fall. I symboler skriver man

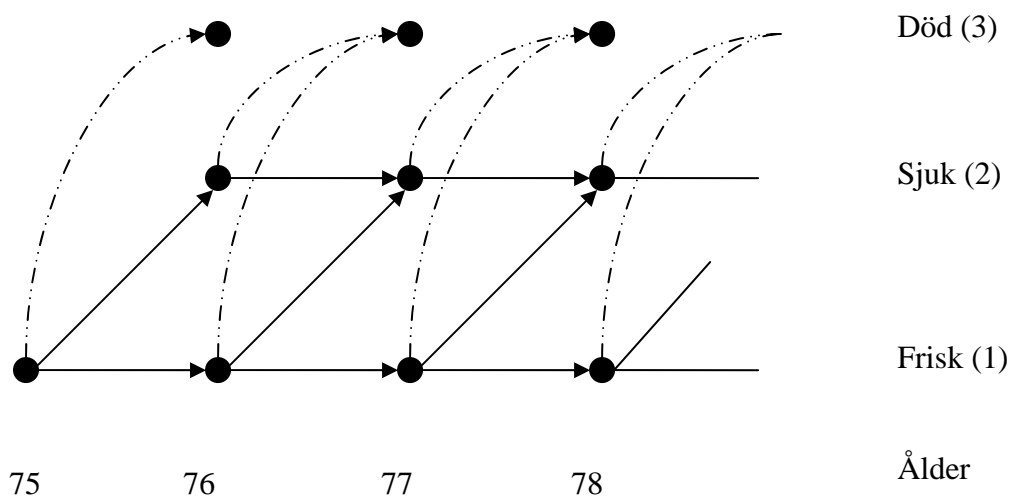
$$\hat{\Pr}(A) = \frac{(\text{antalet gynnsamma utfall})}{(\text{antalet möjliga utfall})}$$

För att beräkna alla ettåriga övergångssannolikheter

$$P_{ij}(t) \quad i = 1, 2 \quad j = 1, 2, 3$$

behöver vi således vid varje nytt år veta vilka övergångar som skett. Men vi vet från tidigare att mättillfällena sker med 3 till 5 års mellanrum vilket resulterar i svårighet att veta vilka

övergångar som skett under varje år. Vi har emellertid viss information som aktualiseras varje år. Närmare bestämt, då övergång sker till tillstånd 3 (död) finns den exakta tiden då det inträffade noterat. Vi vet även att om en person observerats i tillstånd 2 (sjuk) vid två mättillfällen så har hon varit i det tillståndet däremellan också eftersom modellen utesluter tillfrisknande. Motsvarande gäller för en person som observerats i tillstånd 1 (frisk) vid två tillfällen. Men när vi observerat en person i tillstånd 1 (frisk) för att några år senare observera denna i tillstånd 2 (sjuk) vet vi inte vilket år personen blev sjuk, bara att det skett någon gång sen den senaste inspektionen informerade oss om att individen var frisk.



Vi belyser problemet med övergången mellan tillstånd 1 och tillstånd 2 genom att betrakta en individ ur datauppsättningen. Vi har information om att individen vid ålder 75 är frisk och att individen 3 år senare är sjuk. Hur har individen i fråga vandrat för att vid åldern 78 vara sjuk? Tre möjliga vägar finns till handa:

1. Individen blir sjuk när han är 76 år och är sjuk i två år till.
2. Individen blir sjuk när han är 77 år och är sjuk i ett år till.
3. Individen blir sjuk när han är 78 år.

Vi väljer att genomföra två olika beräkningar baserade på följande två antaganden.

#### Antagande 1:

Alla övergångar mellan frisk och sjuk sker enligt väg 3, d.v.s. sjukdom debuterar det år som vi först observerat sjukdom.

## Antagande 2:

Alla övergångar mellan frisk och sjuk sker enligt väg 1, d.v.s. sjukdom debuterar 1 år efter det observerade friska året.

Om vi till att börja med beräknar alla ettåriga övergångssannolikheterna  $P_{11}(t)$ ,  $P_{12}(t)$ ,  $P_{13}(t)$ ,  $P_{22}(t)$  och  $P_{23}(t)$  under antagande 1 för att sen beräkna dem under antagande 2 så kommer vi skapa ett intervall för dessa sannolikheter där den sanna sannolikheten troligen ligger någonstans i intervallet. Vi utför således två olika beräkningar och får på så sätt två uppsättningar av varje övergångssannolikhet. Innan vi tittar närmare på resultatet av beräkningarna så redogör vi först för utjämningen av de ettåriga övergångssannolikheterna.

### 3.3.2 Minsta kvadrat metoden

Skattningarna av ettåriga övergångssannolikheterna ligger inte på en perfekt linje utan är spridda som ett avlångt moln. Minsta kvadrat metoden är en metod för att räkna sig fram till var linjen, som på ”bästa möjliga sätt” ansluter till datamaterialet, ska ligga. Grundprincipen är att punkterna ska ligga så nära linjen som möjligt. Avståndet mellan punkt och linje ska alltså vara så litet som möjligt. Man räknar då samman alla avstånden i exempelvis y led mellan punkter och tänkt linje. För att undvika att positiva och negativa avstånd tar ut varandra tal kvadrerar man alla värdena. Man strävar efter att göra summan av kvadraterna på skillnaderna mellan punkterna och kurvan så liten som möjligt. Då kommer linjen att ligga så nära punkterna som möjligt.<sup>10</sup>

Innan vi genomför utjämningen av övergångssannolikheterna är det angeläget att titta närmare på antalet övergångar som sker varje år. Få observationer kan ge osäkra skattningar som i sin tur kan ge stor påverkan på linjen. Därför utesluter vi en borttagning av de äldsta åldrarna 98, 99, ... osv. Övertygelse av detta ges i Appendix 1 som påvisar det låga antalet övergångar i de högsta åldrarna. I nästa avsnitt presenteras resultaten av de skattade ettåriga övergångssannolikheterna och utjämningen av dem.

---

<sup>10</sup> Se [7]

### 3.3.3 Plottar av de ettåriga övergångssannolikheterna

Nästkommande sidor visar övergångssannolikheterna  $P_{11}(t)$ ,  $P_{12}(t)$ ,  $P_{13}(t)$ ,  $P_{22}(t)$  och  $P_{23}(t)$  under respektive antaganden (antagande 1 och antagande 2) och en utjämning av sannolikheterna enligt minsta kvadrat metoden.

Förnuftsensligt borde plottarna vi snart ska se åskådliggöra följande effekt:

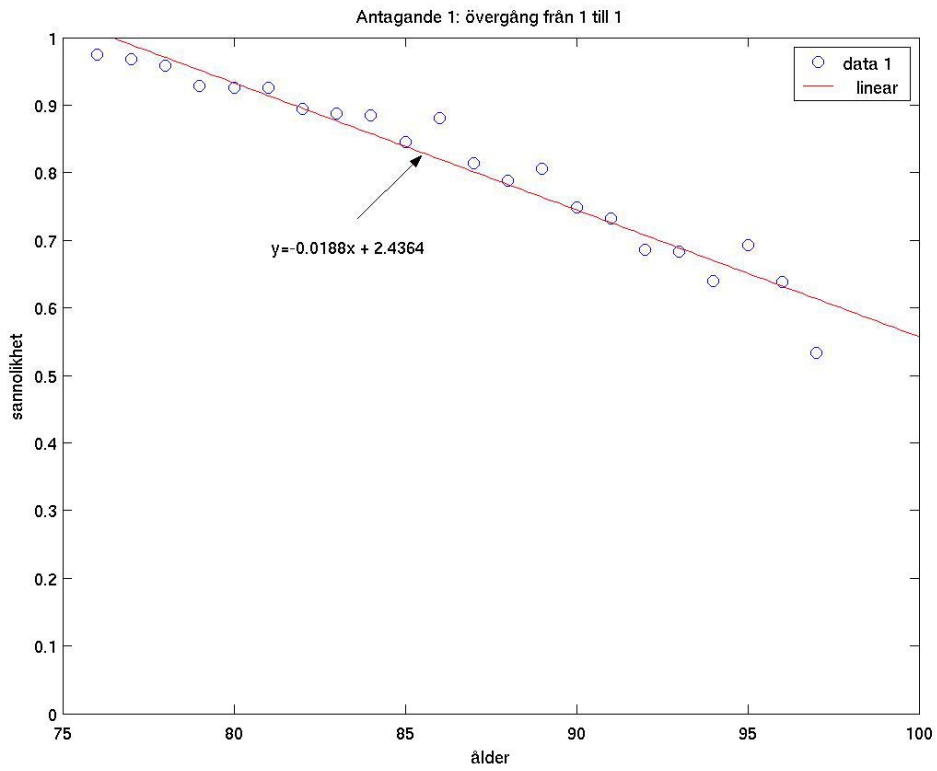
#### Övergång:

- 1 till 1:** Sannolikheten att vara frisk givet att man var frisk året innan borde **minska** då åldern ökar.
- 1 till 2:** Sannolikheten att vara sjuk givet att man var frisk året innan borde **öka** då åldern ökar.
- 1 till 3:** Sannolikheten att vara död givet att man var frisk året innan borde **öka** då åldern ökar.
- 2 till 2:** Sannolikheten att vara sjuk givet att man var sjuk året innan borde **minska** då åldern ökar.
- 2 till 3:** Sannolikheten att vara död givet att man var frisk året innan borde **öka** då åldern ökar.

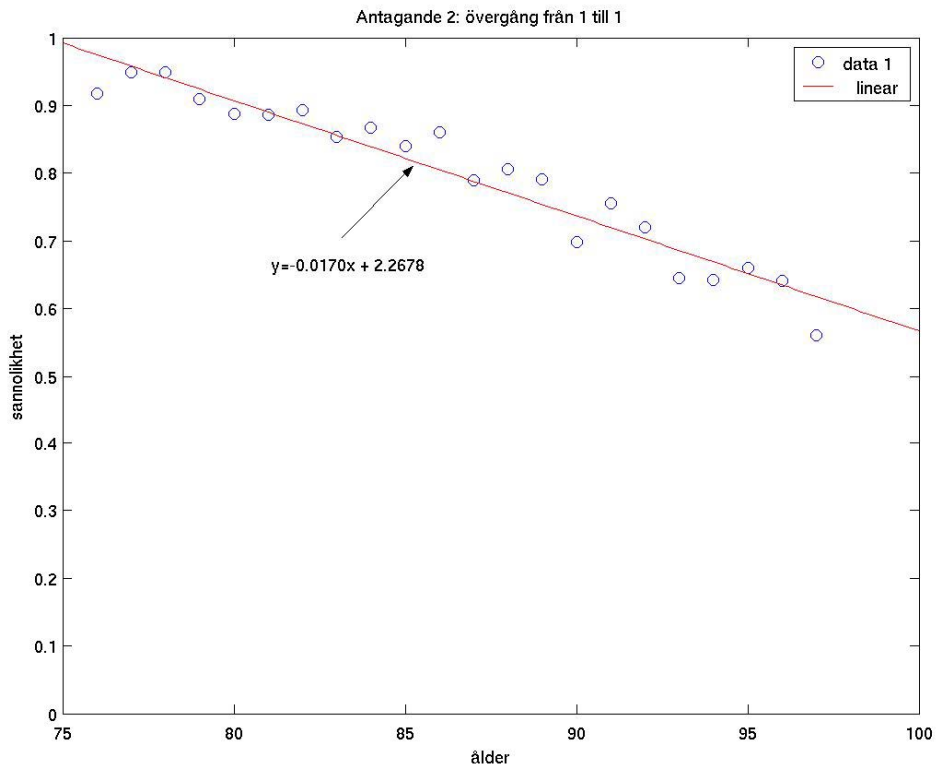
I det följande ser vi att plottarna överensstämmer med vad vi förnuftsensligt förmodar för effekt. Först ges ett förtydligande.

*Förtydligande: Ringarna är de skattade ettåriga övergångssannolikheterna i de olika åldrarna 76 till 97. Den räta linjen är en utjämning enligt minsta kvadratmetoden.*

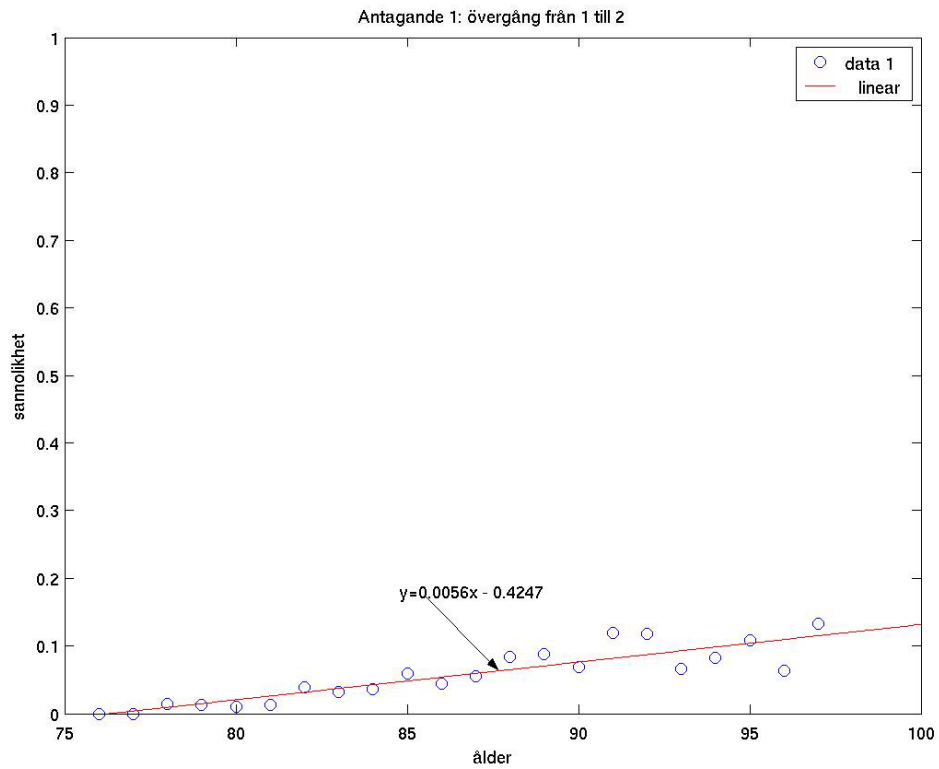
**Bild 1:**  $P_{11}(t); t = 76, 77, \dots, 97$  under **antagande 1** och utjämning av dem



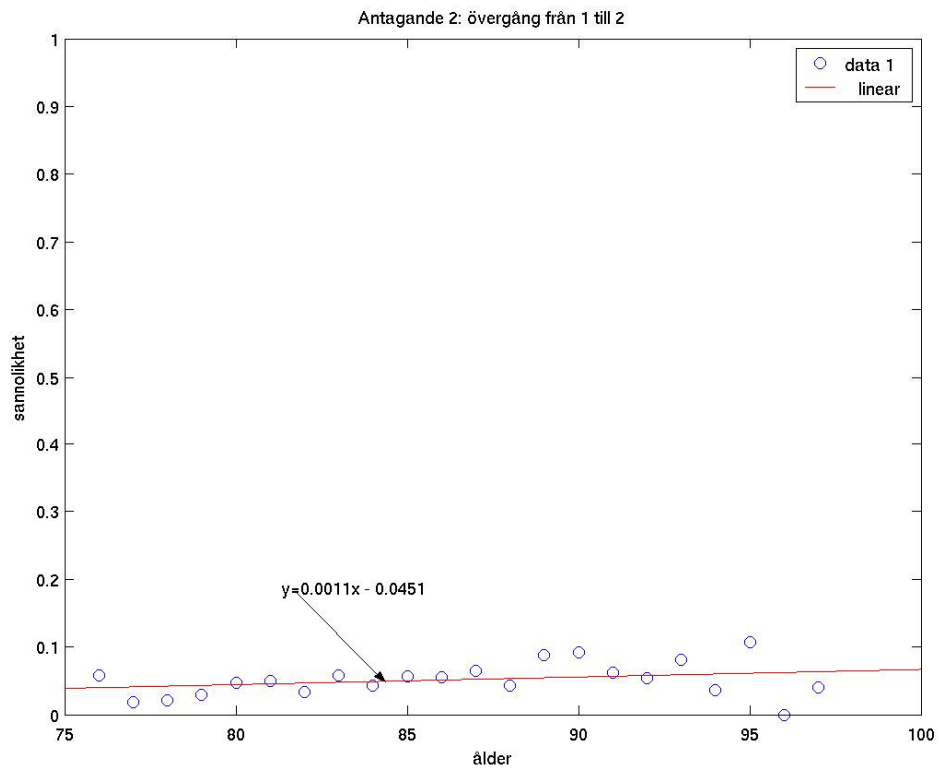
**Bild 2:**  $P_{11}(t); t = 76, 77, \dots, 97$  under **antagande 2** och utjämning av dem



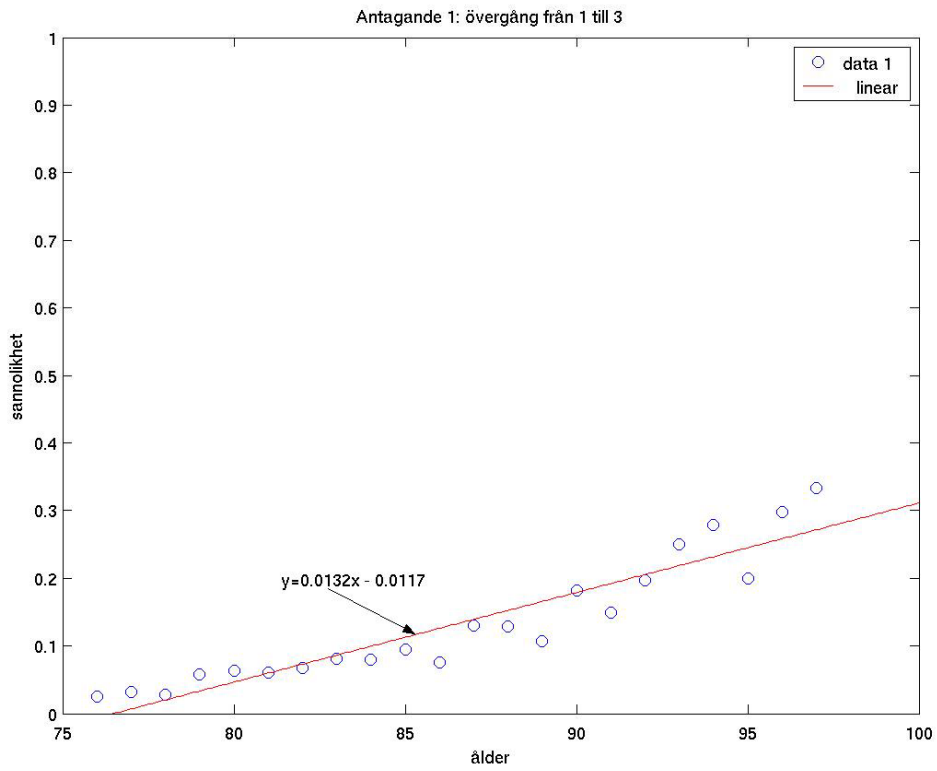
**Bild 3:**  $P_{12}(t); t = 76, 77, \dots, 97$  under **antagande 1** och utjämning av dem



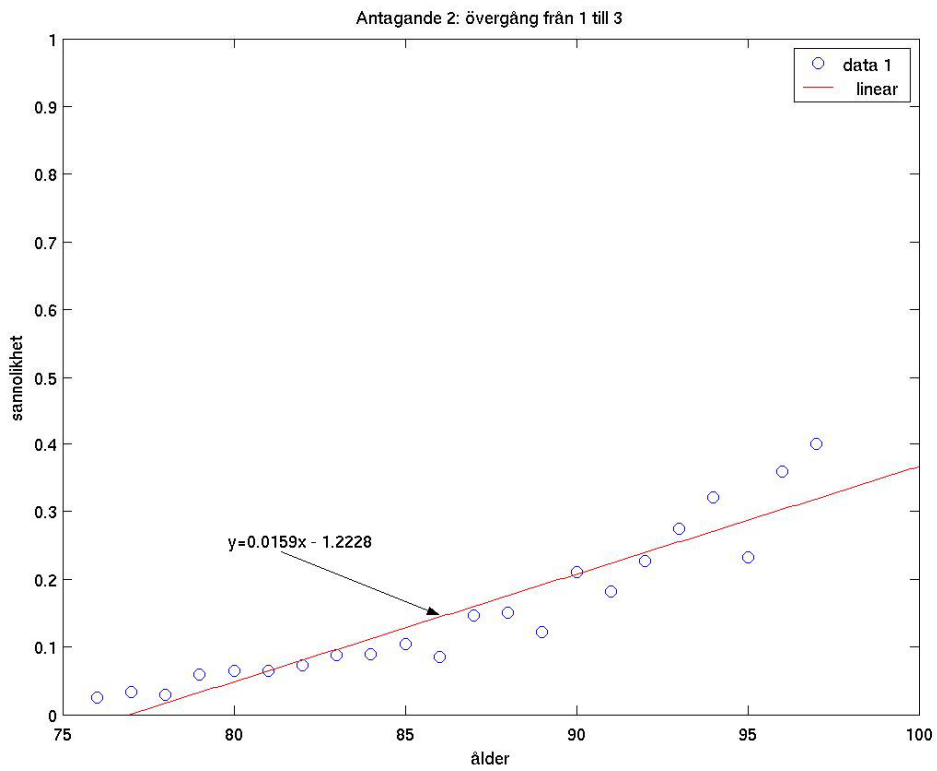
**Bild 4:**  $P_{12}(t); t = 76, 77, \dots, 97$  under **antagande 2** och utjämning av dem



**Bild 5:**  $P_{13}(t); t = 76, 77, \dots, 97$  under **antagande 1** och utjämning av dem

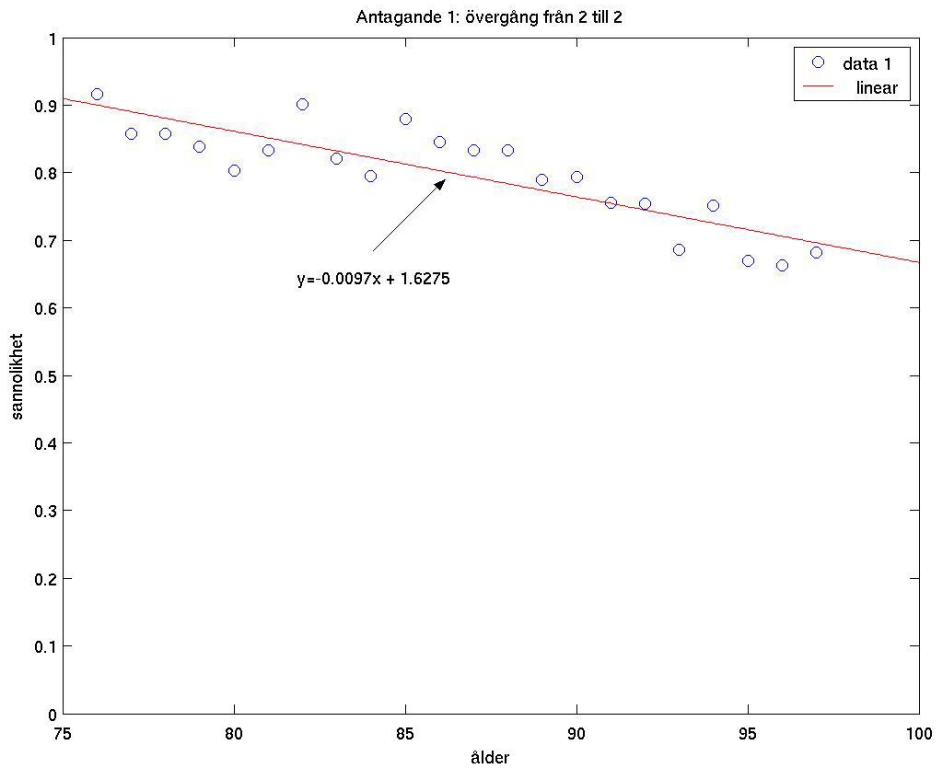


**Bild 6:**  $P_{13}(t); t = 76, 77, \dots, 97$  under **antagande 2** och utjämning av dem

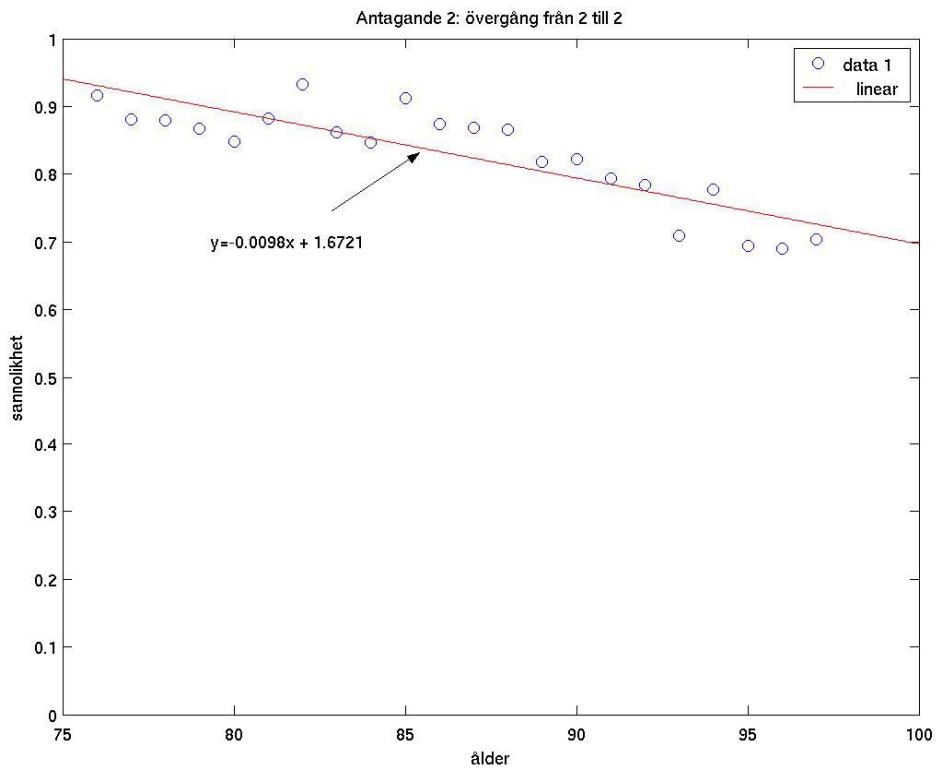




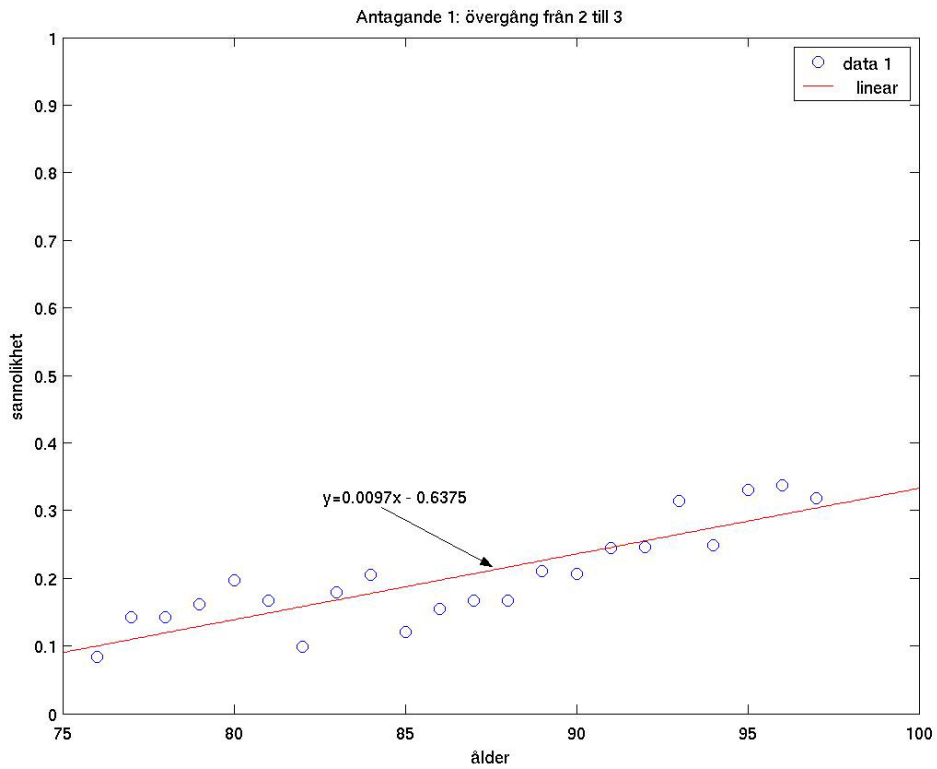
**Bild 7:**  $P_{22}(t); t = 76, 77, \dots, 97$  under **antagande 1** och utjämning av dem



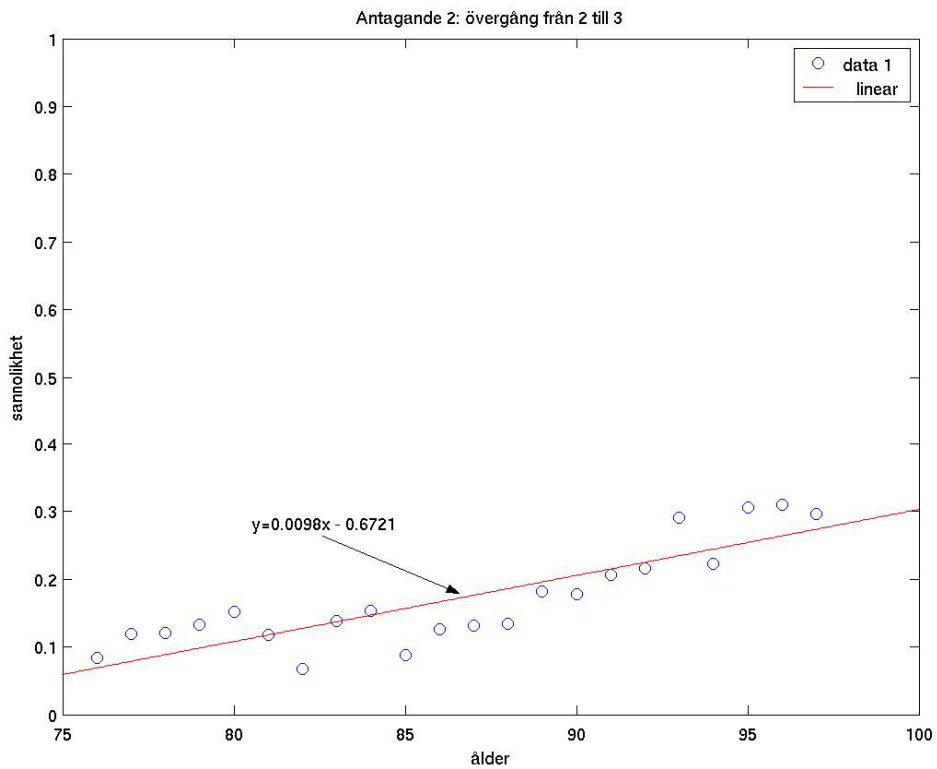
**Bild 8:**  $P_{22}(t); t = 76, 77, \dots, 97$  under **antagande 2** och utjämning av dem



**Bild 9:**  $P_{23}(t); t = 76, 77, \dots, 97$  under **antagande 1** och utjämning av dem



**Bild 10:**  $P_{23}(t); t = 76, 77, \dots, 97$  under **antagande 2** och utjämning av dem



Det kan diskuteras huruvida utjämning ska ske enligt en rätlinjig kurva, i synnerhet då övergång sker till tillstånd 3 (Död), då vi från teorin om livslängdsmodeller välbekant använder en exponentiell kurva. Betraktar vi plottarna av  $P_{13}(t)$  och  $P_{23}(t)$  så kan vi ana ett krökt samband. I varje fall tycks  $P_{13}(t)$  växa snabbare än en rät linje. Dock ger begränsningar i materialet en osäkerhet. Trots att vi har reducerat materialet genom att bara betrakta åldrarna upp till 97, kvarstår problemet med för få observationer i de höga åldrarna. Det kan således vara förrädiskt att vid utjämning av övergångssannolikheterna lägga stor vikt i att sannolikheter i de höga åldrarna skapar ett exponentiellt utseende. Vi väljer att utföra utjämningen efter en rätlinjig kurva men framför ändå en reservation att med förbättrat material bör diskussionen återupptas.

När vi nu är i besittning av övergångssannolikheterna kan vi beräkna incidens och mortalitet, under de två antaganden, genom att betrakta antalet individer i varje åldersgrupp och utföra multiplikation med respektive ålderssannolikhet. Som exempel, om vi vet att 1000 åldringar är friska vid åldern 85 så kan vi nu beräkna hur många av dem som är friska, sjuka eller döda vid 86-års ålder. Även om detta kan vara av intresse i sig självt kan vi inte svara på frågan om hur många 86-åringar som är sjuka. Denna frågeställning för oss till nästa kapitel.

## 3.4 Prevalens

För att kunna beräkna prevalensen (antalet individer som är sjuka vid en viss tidpunkt) måste vi kunna redogöra för sannolikheten att vara sjuk en viss tidpunkt. Vi har hittills definierat ettårsövergångssannolikheterna  $P_{ij}(t)$ , men för att beräkna sannolikhetsfördelningen att vara frisk, sjuk eller död om  $x$ , ( $x = 2, 3, \dots$ ) år ska vi nu betrakta övergångssannolikheter som avser fler än ett år. Dessa sannolikheter ska skattas utifrån att man är frisk vid begynnelseåldern 75 år som vidare benämns tiden 0.

### 3.4.1 Beräkning av övergångssannolikheterna

Vi betraktar sannolikheten att en person är sjuk vid tiden  $t$  givet att personen var frisk vid tiden 0. Denna sannolikhet betecknas  $P_{12}(0, h)$ . Händelsen kan inträffa på  $h$  olika sätt<sup>11</sup>, som illustreras på nästa sida.

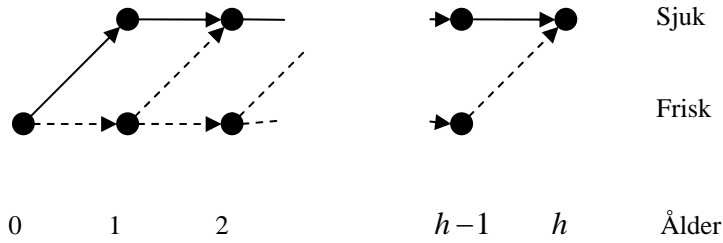
---

<sup>11</sup> Se [1] s.73

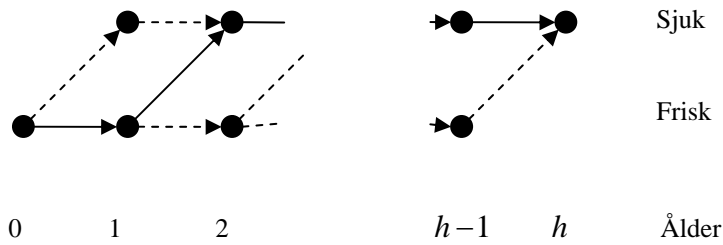
**Klargörande:**

Den hel- sträckta linjen visar vägen personen tar. Över bilden finns en verbal förklaring.

**1. Personen är sjuk under tiden 1 till  $h$ .**

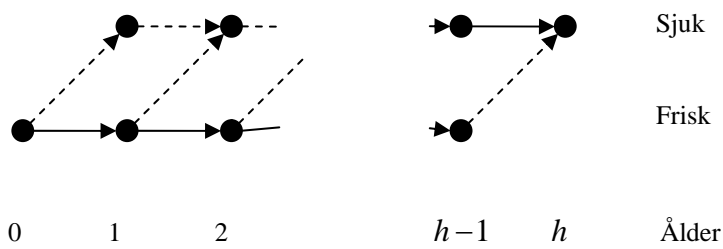


**2. Personen är sjuk under tiden 2 till  $h$ .**

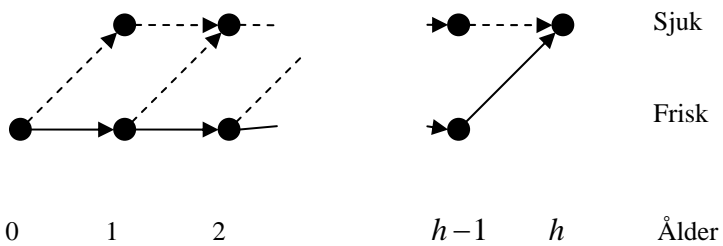


⋮                      ⋮                      ⋮                      ⋮

**$h-1$ . Personen är sjuk under tiden  $h-1$  till  $h$ .**



**$h$ . Personen är sjuk vid tiden  $h$ .**



Nu redogör vi för sannolikheten att ta de olika vägarna.

På grund av den Markovska egenskapen får vi:

Sannlikheten att personen tar väg 1:

$$P_{12}(\text{tar väg 1}) = P_{12}(0,1)P_{22}(1,2)P_{22}(2,3)\cdots P_{22}(h-2,h-1)P_{22}(h-1,h)$$

Sannlikheten att personen tar väg 2:

$$P_{12}(\text{tar väg 2}) = P_{11}(0,1)P_{12}(1,2)P_{22}(2,3)\cdots P_{22}(h-2,h-1)P_{22}(h-1,h)$$

•                    •                    •                    •  
 •                    •                    •                    •  
 •                    •                    •                    •

Sannlikheten att personen tar väg h-1 är:

$$P_{12}(\text{tar väg } h-1) = P_{11}(0,1)P_{11}(1,2)P_{11}(2,3)\cdots P_{12}(h-2,h-1)P_{22}(h-1,h)$$

Sannlikheten att personen tar väg h är:

$$P_{12}(\text{tar väg } h) = P_{11}(0,1)P_{11}(1,2)P_{11}(2,3)\cdots P_{11}(h-2,h-1)P_{12}(h-1,h)$$

Summan av dessa sannolikheter är ekvivalent med den sannolikhet vi söker, nämligen

$P_{12}(0,h)$ . (Se [1] s.73):

$$P_{12}(0,h) = \sum_{j=0}^{h-1} P_{11}(0,j)P_{12}(j,j+1)P_{22}(j+1,h)$$

För övertygelse ges en härledning av  $P_{12}(0,h)$  med hjälp av Chapman-Kolmogorovs ekvationer i Appendix 2.

På motsvarande sätt kan vi visualisera de andra relevanta övergångarna,  $1 \rightarrow 1$ ,  $1 \rightarrow 3$  och frambringa övergångssannolikheterna  $P_{11}(0,h)$  och  $P_{13}(0,h)$

$$P_{11}(0,h) = \prod_{j=0}^{h-1} P_{11}(j,j+1)$$

$$P_{13}(0,h) = \sum_{j=0}^{h-1} P_{11}(0,j)P_{13}(j,j+1)$$

Vi väljer att beräkna sannolikheten  $P_{13}(0,h)$  på ovanstående sätt men det ska poängteras att då vi betraktar en person som är frisk vid tidpunkten 0 och död vid tidpunkten  $h$  finns det en

möjlighet att personen varit sjuk däremellan. Därför är den nedanstående formeln för  $P_{13}(0, h)$  korrekt:

$$P_{13}(0, h) = \sum_{j=0}^{h-1} P_{11}(0, h) P_{13}(j, j+1) + \sum_{j=0}^{h-2} P_{11}(0, j) P_{12}(j, j+1) \sum_{i=j+1}^{h-1} P_{22}(j+1, i) P_{23}(i, i+1)$$

I följande kapitel redovisas resultaten av våra beräkningar av övergångssannolikheterna.

### 3.4.2 Resultat

Efter den givna teorin i föregående avsnitt kan vi nu presentera resultatet av våra beräkningar. Vi väljer att visa resultaten i bildform men kommer även hänvisa till appendix där resultaten även presenteras i tabellform.

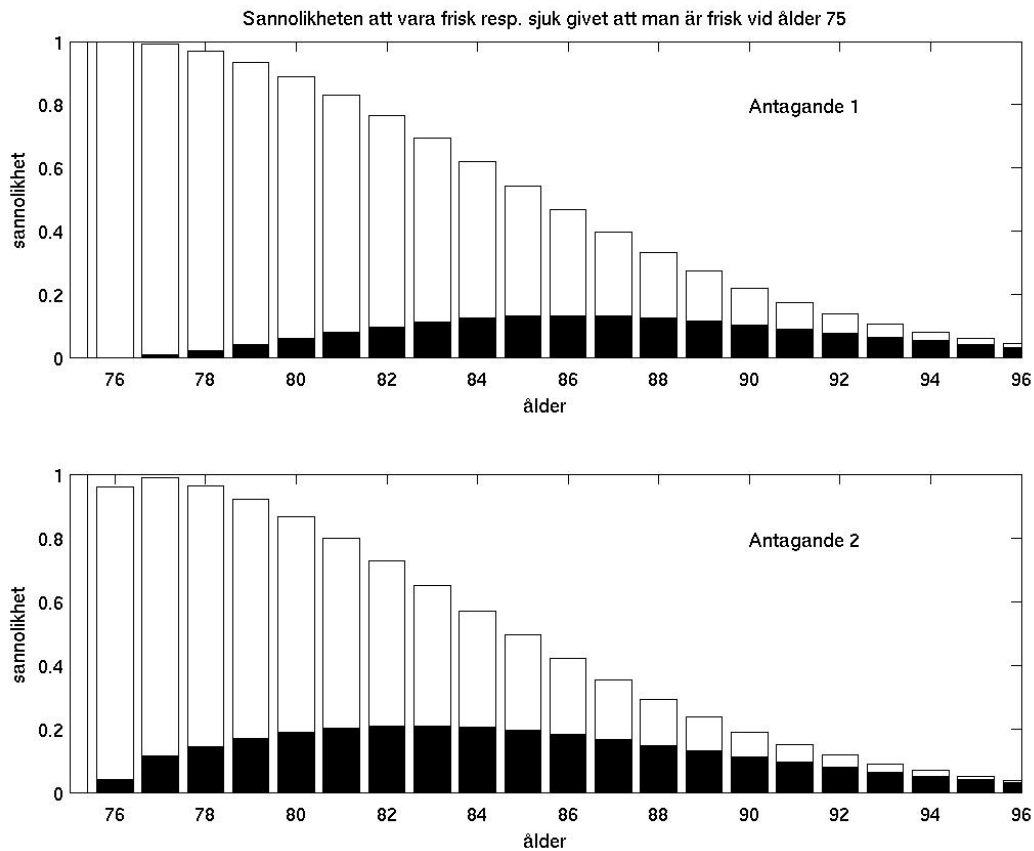
Vi repeterar betydelsen av de olika övergångssannolikheterna:

- $P_{11}(0, h)$  Sannolikheten att man är **frisk** vid åldern  $h$  ( $h = 76, 77, \dots, 97$ ) givet att man är frisk vid åldern 75.
- $P_{12}(0, h)$  Sannolikheten att man är **sjuk** vid åldern  $h$  ( $h = 76, 77, \dots, 97$ ) givet att man är frisk vid åldern 75.
- $P_{13}(0, h)$  Sannolikheten att man är **död** vid åldern  $h$  ( $h = 76, 77, \dots, 97$ ) givet att man är frisk vid åldern 75.

De två övergångssannolikheterna,  $P_{11}(0, h)$  och  $P_{12}(0, h)$ ;  $h = 76, 77, \dots, 97$ , som redovisas i bild längre ner, kommer bestå av staplar uppstaplade på varandra.  $P_{11}(0, h)$  har vita staplar och  $P_{12}(0, h)$  har svarta. Resterande utrymme upp till sannolikhet 1 består av övergångssannolikheten  $P_{13}(0, h)$ .

Med anledning av att vi beräknat övergångssannolikheterna med under två antaganden presenteras två bilder. En bild med stöd av antagande 1 och en bild med stöd av antagande 2.

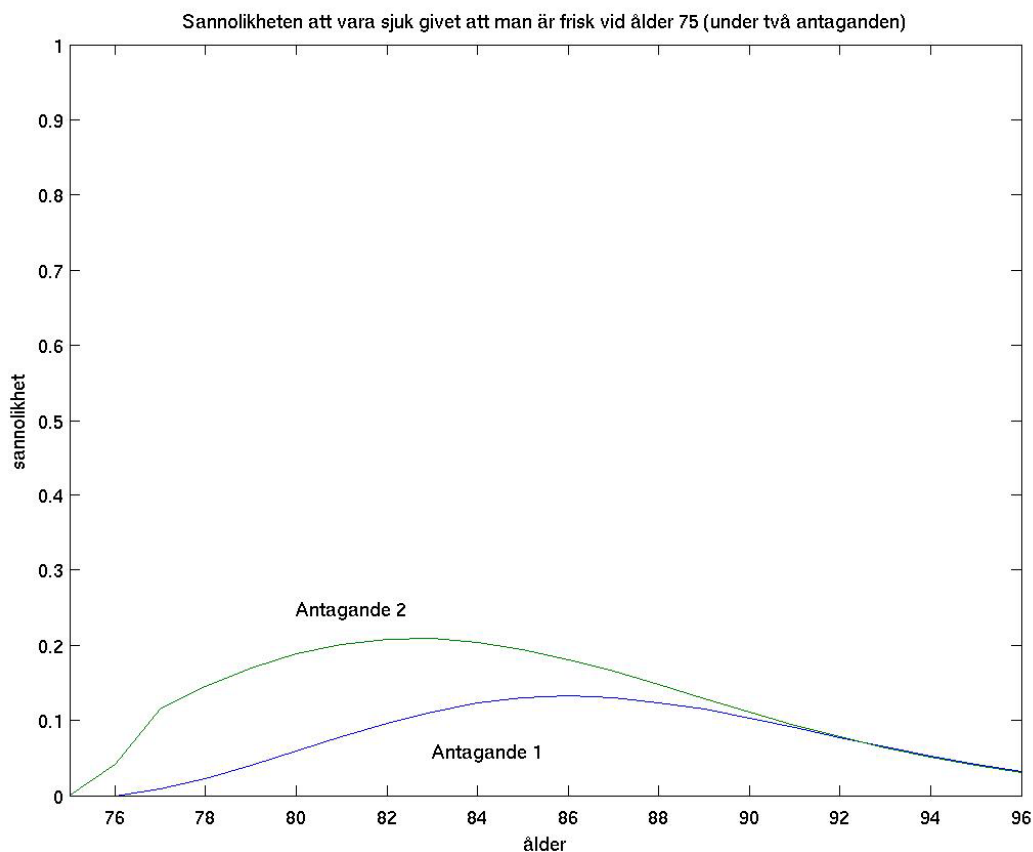
**Bild 11:** Övergångssannolikheterna  $P_{11}(0, h)$  och  $P_{12}(0, h)$ ;  $h = 76, 77, \dots, 97$  under antagande 1 och antagande 2.



För att se resultaten i tabellform hänvisas läsaren till Appendix 3.

För att förtydliga teckningen visar vi övergångssannolikheterna  $P_{12}(0, h)$  med stöd av respektive antagande i samma bild. Denna gång presenteras resultaten som två kurvor för att förenkla jämförelsen av de två antaganden.

**Bild 12:** Övergångssannolikheterna  $P_{12}(0, h)$ ;  $h = 76, 77, \dots, 97$  under antagande 1 och antagande 2.



Bilden ovan representerar resultatet av detta arbete. Vi har bildat en övre och undre gräns för sannolikheten att vara sjuk vid en viss ålder och kan med dessa gränser skapa ett intervall som beskriver prevalensen. Vi kan med bilden (och Appendix 3) uppskatta hur många av 1000 friska 75-åringar som kommer vara friska respektive sjuka vid en viss ålder  $x$ ,  $x = 76, 77, \dots, 97$ . T.ex. kan vi skatta att av de 1000 friska 75-åringar så kommer, vid åldern 86, antalet friska finns i intervallet 239 till 337 stycken och antalet sjuka i intervallet 132 till 181 stycken. Denna uppskattning förändras då vi ändrar incidensen och prevalensen det vill säga när vi ändrar övergångssannolikheterna i kapitel 3.3.3. Såldes har vi lyckats med vad vi ville åstadkomma; en modell som beskriver prevalens som resultat av incidens och mortalitet.



## 4. DISKUSSION

Syftet med detta examensarbete var att skapa en modell som beskriver prevalens som resultat av incidens och mortalitet. Först skapades en modellframställning med begränsningen att sjukdomen är oåterkallelig. Incidens och mortalitet beror av ålder men inte kön. Avsaknaden av könsberoendet inrättades med anledning av det begränsade antalet observerade män. En ytterligare begränsning i materialet var att uppföljningsundersökningarna inte genomfördes årligen. Av denna anledning skattades två uppsättningar av de ettåriga övergångssannolikheterna (Bild 1 till 10). Utjämningsmetoden sker enligt minsta kvadrat metoden. Därefter skattades övergångsintensiteterna under antagandet att alla 75-åringar var friska och slutresultatet frambringade ett konfidensintervall för prevalensen (Bild 12).

Föreliggande rapport är ett led i ett projekt som forskningsprogrammet Tredjeliv har initierat. Projektet fullföljs av Malin Düring, student på Stockholms Universitet, som i ljuset av ovan nämnda modell ska verkställa förändringar i de ettåriga övergångsintensiteterna för att observera effekten på intervallet för prevalensen. Förutom att undersöka samspelet mellan de ingående faktorerna i modellen så finns en förhoppning att i och med detta finna svar på vilken av de tre hälsoteorier (Se sidan 7) som bäst beskriver framtiden, alternativt svara på frågan om vilka förändringar i de ettåriga övergångssannolikheterna som krävs för att respektive teori kan infrias.

## 5. REFERENSER

- [1] Haberman, S och Pitacco, E (1999): *Actuarial Models for Disability Insurance*, Chapman&Hall
- [2] Hallberg, Daniel och Lagergren, Mårten (2004): "Transitions between modes of public care services and mortality among the very old in Sweden", Stiftelsen Stockholms läns äldrecentrum.
- [3] Lagergren, Mårten (2004): "Utvecklingen av de äldres hälsa och levnadslängd", Stiftelsen Stockholms läns äldrecentrum.
- [4] Ross, Sheldon M. (2003): *Introduction to Probability models*, Academic Press.
- [5] [www.ida.liu.se/labs/eis/people/petah\\_files/Forsknings-%20och%20utvecklingsprogrammet%20TREDJELIVET.pdf](http://www.ida.liu.se/labs/eis/people/petah_files/Forsknings-%20och%20utvecklingsprogrammet%20TREDJELIVET.pdf)
- [6] [www.aldrecentrum.se/kungs.html](http://www.aldrecentrum.se/kungs.html)
- [7] [www.math.su.se/~teke/undervisning/linjanextramaterial.pdf](http://www.math.su.se/~teke/undervisning/linjanextramaterial.pdf)
- [8] Öberg, Svante och Isralesson, Torbjörn (2003): "Sveriges framtida befolkning, befolkningsframskridning för åren 2003-2050", SCB, Programmet Demografisk analys och jämställdhet

## 6. Appendix

### Appendix 1

Antal personer som vid åldern  $x-1$  var i tillstånd **1** och som i åldern  $x$  befann sig i något av tillstånden **1, 2 eller 3**, där  $x = 76, 77, \dots, 103$ .

Ålder	1 till 1	1 till 2	1 till 3	Summa	Ålder	1 till 1	1 till 2	1 till 3	Summa
76	118	0	3	121	90	237	26	50	313
77	215	0	7	222	91	185	30	28	243
78	301	2	8	311	92	130	22	33	185
79	392	7	17	416	93	101	12	30	143
80	473	6	24	503	94	65	11	26	102
81	528	8	27	563	95	48	7	11	66
82	564	24	31	619	96	30	4	14	48
83	586	24	44	654	97	16	4	10	30
84	589	21	41	651	98	7	3	6	16
85	544	43	47	634	99	4	0	3	7
86	523	27	29	579	100	1	1	2	4
87	459	34	61	554	101	1	0	0	1
88	375	42	44	461	102	0	0	1	1
89	305	33	31	369	103	0	0	0	0

Antal personer som vid åldern  $x-1$  var i tillstånd **2** och som i åldern  $x$  befann sig i något av tillstånden **2 eller 3**, där  $x = 76, 77, \dots, 103$ .

Ålder	2 till 2	2 till 3	Summa	Ålder	2 till 2	2 till 3	Summa
76	11	1	12	90	199	40	239
77	30	5	35	91	177	52	229
78	43	7	50	92	163	47	210
79	56	8	64	93	125	50	175
80	64	12	76	94	105	31	136
81	78	14	92	95	79	35	114
82	104	10	114	96	63	29	92
83	134	29	163	97	48	20	68
84	139	32	171	98	38	12	50
85	174	21	195	99	30	8	38
86	212	38	250	100	16	14	30
87	222	38	260	101	11	4	15
88	228	38	266	102	7	4	11
89	211	47	258	103	1	5	6

## Appendix 2

I det följande härleds  $P_{12}(0, h)$  med hjälp av Chapman-Kolmogorovs ekvationer. Vi börjar med att definiera Chapman-Kolmogorovs ekvationer.

DEFINITION (Chapman-Kolmogorovs ekvationer)

$$P_{ij}(t, u) = \sum_k P_{ik}(t, w)P_{kj}(w, u) \quad \text{där } t, w, u \text{ är heltal} \quad (t \leq w \leq u)^{12}$$

Enligt Chapman-Kolmogorovs ekvationer gäller följande

$$P_{12}(0, h) = P_{11}(0, h-1)P_{12}(h-1, h) + P_{12}(0, h-1)P_{22}(h-1, h)$$

Låt vara att

$$A_0 = P_{11}(0, h-1)P_{12}(h-1, h)$$

Då  $P_{22}(h, h) = 1$  har vi möjlighet att uttrycka  $A_0$  såsom

$$A_0 = P_{11}(0, h-1)P_{12}(h-1, h)P_{22}(h, h)$$

Vilket längre fram ska visa sig som lämpligt. Följaktligen är

$$P_{12}(0, h) = A_0 + P_{12}(0, h-1)P_{22}(h-1, h)$$

Vidare utvecklar vi uttrycket för  $P_{12}(0, h)$  genom att utveckla  $P_{12}(0, h-1)$  med hjälp av Chapman-Kolmogorovs ekvationer.

$$\begin{aligned} P_{12}(0, h) &= A_0 + (P_{11}(0, h-2)P_{12}(h-2, h-1) + P_{12}(0, h-2)P_{22}(h-2, h-1))P_{22}(h-1, h) \\ &= A_0 + P_{11}(0, h-2)P_{12}(h-2, h-1)P_{22}(h-1, h) + P_{12}(0, h-2)P_{22}(h-2, h) \end{aligned}$$

Där sista stycket kom ut av att

$$P_{22}(h-2, h-1)P_{22}(h-1, h) = P_{22}(h-2, h)$$

Låt vara att

$$A_1 = P_{11}(0, h-2)P_{12}(h-2, h-1)P_{22}(h-1, h)$$

I så fall har vi att

$$P_{12}(0, h) = A_0 + A_1 + P_{12}(0, h-2)P_{22}(h-2, h)$$

---

<sup>12</sup> Se [4] s. 185

Fortsättningsvis utvecklar vi  $P_{12}(0, h-2)$  och får att

$$P_{12}(0, h) = A_0 + A_1 + A_2 + P_{12}(0, h-3)P_{22}(h-3, h)$$

Där

$$A_2 = P_{11}(0, h-3)P_{12}(h-3, h-2)P_{22}(h-2, h)$$

Slutför vi algoritmen så får vi till sist att

$$P_{12}(0, h) = \sum_{i=0}^{h-1} A_i$$

Som även kan skrivas

$$P_{12}(0, h) = \sum_{j=0}^{h-1} P_{11}(0, j)P_{12}(j, j+1)P_{22}(j+1, h)$$

I och med det är härledningen av övergångssannolikheten att man är i tillstånd 2 vid tidpunkten  $h$  givet att man är i tillstånd 1 vid tidpunkten 0 fullbordad.

## Appendix 3

*Sannolikheten att vara frisk respektive sjuk vid ålder  $x$ , där  $x = 76, 77, \dots, 103$ , givet att man var frisk vid ålder 75. Antagande 1 och Antagande 2 presenteras.*

Ålder	Frisk Ant. 1	Frisk Ant. 2	Sjuk Ant. 1	Sjuk Ant. 2
76	1	0,9172	0	0,0416
77	0,9823	0,8733	0,0088	0,1146
78	0,9466	0,8167	0,0225	0,1448
79	0,8946	0,7499	0,0397	0,1698
80	0,8287	0,6758	0,0589	0,1888
81	0,7522	0,5975	0,0783	0,2015
82	0,6688	0,5182	0,0963	0,2080
83	0,5821	0,4405	0,1116	0,2085
84	0,4959	0,3671	0,1230	0,2035
85	0,4131	0,2996	0,1300	0,1940
86	0,3365	0,2394	0,1323	0,1808
87	0,2678	0,1873	0,1301	0,1649
88	0,2081	0,1433	0,1240	0,1475
89	0,1579	0,1072	0,1147	0,1293
90	0,1168	0,0784	0,1032	0,1112
91	0,0843	0,0560	0,0904	0,0939
92	0,0592	0,0390	0,0772	0,0778
93	0,0405	0,0265	0,0643	0,0634
94	0,0269	0,0176	0,0522	0,0508
95	0,0174	0,0114	0,0415	0,0400
96	0,0109	0,0072	0,0322	0,0310