Matematisk statistik
Stockholms universitet

IBNR-estimering med livförsäkringsmetoder för dödsfallsmomentet i individuell sjukförsäkring

Johan Dehaddo

Examensarbete 2008:13

Postadress:

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm
Sverige

Internet:

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

Matematisk statistik
Stockholms universitet
Examensarbete **2008:13**,
http://www.math.su.se/matstat

IBNR-estimering med livförsäkringsmetoder för dödsfallsmomentet i individuell sjukförsäkring

Johan Dehaddo*

September 2008

Sammanfattning

I Trygg-Hansas individuella sjukförsäkring ingår vanligtvis att sjukersättning utbetalas till dödsboet om den skadelidande dör efter att rätt till sjukersättning erhållits, men innan ersättningen slutreglerats. Denna förmån betraktas som livförsäkring och utbetalas inte av Trygg-Hansa utan av livförsäkringsbolaget Holmia Liv.
I dagens läge så använder sig Trygg-Hansa utav Chain ladder metoden för att beräkna IBNR (incurred but not reported) för Holmia Liv. I och med att ålder, kön och utbetalningsbelopp är kända borde en utgångspunkt i dödsfallssannolikheter fungera bättre.
Dödligheterna kommer att skattas med Makehams modell vilket är den vanligaste dödlighetsmodellen. Det visar sig att anpassningen inte blir den bästa, på grund av databrist, men accepteras ändå eftersom variansen inte kan bli högre än vad den var tidigare.
Intensiteten för slutregleringarna kommer att behandlas med variansanalytiska metoder, då vi även här har ett otillräckligt data.

*Postadress: Matematisk statistik, Stockholms universitet, 106 91, Sverige. E-post: johan.dehaddo@trygghansa.se. Handledare: Anders Martin-Löf

Abstract

In Trygg-Hansa's personal health insurance policy, the sickness allowance (payout) is usually paid to the estate in case the policy holder passes away after fulfilling the regulations regarding the insurance policy but before the compensation has been technically determined (settled). This sort of payout is regarded as a life insurance product and is paid by a separate life assurance company, Holmia Liv and not by Trygg-Hansa.

At the present time, Trygg-Hansa is using the Chain-ladder method in order to calculate the so called IBNR (incurred but not reported) for Holmia Liv. It is however known that any further knowledge of age, gender and payout are crucial in enhancing the accuracy of the mortality models.

The mortality models are estimated using Makeham's model which is the most common model in this field. Unfortunately the output is insufficient in accuracy due to a lack of input data. It is however accepted since the variance cannot exceed its earlier upper limit.

The policy settlement's intensity will be dealt with methods of analysis of variance, since we here too have acquired inadequate data.

Förord

Detta arbete utgör ett examensarbete om 30 poäng vid Stockholms universitet. Det har varit ett utmanande men roligt arbete då jag bland annat har fått syssla med omfattande datautsökning som har varit väldigt lärorikt för mig.

Jag vill först och främst tacka min chef och handledare Lars Klingberg, diplomerad chefsaktuarie på Trygg-Hansa. Han har gett mig möjligheten att få vikariera som aktuarie samtidigt som att få jobba med mitt examensarbete. Han har varit ett stort stöd då han har trott på mig. Sedan vill jag tacka Ulf Holmström, databasanalytiker på Trygg-Hansa, för all hjälp med programmering. Jag vill även tacka min andra handledare Anders Martin-Löf, professor på Stockholms universitet, som har hjälp mig med beräkningar och en del förklaringar gällande matematiska modeller. Sist men inte minst vill jag tacka min mamma, pappa, systrar och fästmö Claudia Vasquez som har trott på mig, stått ut med mig och gett mig en massa självförtroende.

Innehållsförteckning

Beskrivning av data 4
Innehåll 4
IBNR och reservering av skador 4
Chain Ladder Metoden 4
Teori, beräkning av överlevnads- och dödsfalssannolikheter 6
Livslängd och återstående livslängd 6
Dödlighetsintensitet, $\boldsymbol{\mu_x}$ 6
Överlevelsefunktion 7
Den ettåriga dödsrisken $\boldsymbol{q_x}$ 7
Förväntad återstående livslängd, $\boldsymbol{E(T_x) = e_x}$ 8
Kopplingarna mellan $\boldsymbol{\mu_x}$ och $\boldsymbol{l_x}$ respektive $\boldsymbol{q_x}$ och $\boldsymbol{l_x}$ 8
Skattningar av den ettåriga dödlighetssannolikheten $\boldsymbol{q_{x,t}}$ 10
Metod 10
Makehamanpassning 11
Skattning av sannolikheten att en skada stängs inom ett år 17
Tvåsidig indelning 17
Försäkringsbelopp 23
Ibnr-reserv, beräkning 24
Slutsats 25
Appendix 26
Referenser 30

Beskrivning av data

Innehåll

De data som har använts kommer ifrån Trygg-Hansa och består av 317743 försäkrade personers uppgifter mellan skadeåren 2000 och 2007, varav 165238 män och resterande kvinnor. De uppgifter som har erhållits från data är:

- Personnummer
- Skadedatum, då den försäkrade blev sjuk eller skadad
- Dödsdatum, då den skadelidande avled
- Slutregleringsdatum, då skadeärendet slutreglerades/stängdes utan att den skadelidande avled
- Slutdatum för öppna skador som har valts till 2007-12-31
- Försäkringsbelopp

IBNR och reservering av skador

En skadereglerare sätter reserver för skador som ännu inte är slutreglerade, men man måste även avsätta reserver för förändringar av dessa belopp eller för s.k. IBNR-skador. IBNR eller ”incurred but not reported”, som det är en förkortning för, är en reserv som skall täcka olika kostnader för de skador som utav erfarenhet bör ha inträffat men som ännu inte har rapporterats in. Det finns olika sätt att uppskatta dessa IBNR – reserver och ett sätt att skatta dem är genom Chain ladder metoden som hanterar bl.a. skadekostnader och därav IBNR – reserv.

Chain Ladder Metoden

Chain Ladder metoden bygger helt och hållet på historisk data. Datat brukar presenteras i form av skadetrianglar, där raderna representerar skadeår och kolumnerna representerar utvecklingsperiod. Om utvecklingsperioden skulle vara årsvis, så kommer diagonalerna från övre högra hörnet till det nedre vänstra hörnet representera kalenderår.
Varje cell i en sådan triangel är baserad på kumulativa data dvs. tidigare skadebelopp plus nuvarande skadebelopp. Ett skadebelopp består utav utbetalt belopp plus ersättningsreserv. Ett exempel på en skadetriangel med fem skadeår och med fem utvecklingsperioder skulle kunna se ut som följande:

Utvecklingsår (j)

Skadeår (i)		1	2	3	4	5
	2003	C_{11}	C_{12}	C_{13}	C_{14}	C_{15}
	2004	C_{21}	C_{22}	C_{23}	C_{24}	C_{25}
	2005	C_{31}	C_{32}	C_{33}	C_{34}	C_{35}
	2006	C_{41}	C_{42}	C_{43}	C_{44}	C_{45}
	2007	C_{51}	C_{52}	C_{53}	C_{54}	C_{55}

där det skuggade området innehåller det data som skall estimeras för att sedan beräkna reserverna.

Vi låter m vara den sista utvecklingsperioden som vi känner till och i detta exempel så är m lika med fem.
Vi inför följande variabler:

- $C_{i,j} =$ Kumulativ skadekostnad från skadeår *i* som är inrapporterad fram till period *j*.

- $f_j =$ Period j:s utvecklingsfaktor, där:

$$f_j = \frac{\sum_{i=1}^{m-j} C_{i,j+1}}{\sum_{i=1}^{m-j} C_{i,j}}, \qquad 1 \leq j \leq m-1$$

- $C_{i,m} =$ Slutgiltig skadekostnad, där

$$E\left(C_{i,j+1} \middle| C_{i,1}, C_{i,2}, \ldots, C_{i,j}\right) = C_{i,j} f_j , \qquad 1 \leq i \leq m, 1 \leq j \leq m-1$$

- $R_i =$ Reserv. Det belopp som ett försäkringsbolag har reserverat för kommande utbetalningar för skadeår.

$$E(R_i) = E\left(C_{i,m}\right) - C_{i,m+1-i}, \qquad 2 \leq i \leq m$$

$$Var\left(C_{i,j} \middle| C_{i,1}, C_{i,2}, \ldots, C_{i,j-1}\right) = C_{i,j-1} * \sigma_{j-1}^2$$

$$\Rightarrow Var(R_i) = \frac{1}{m-j-1} \sum_{i=1}^{m-j} C_{i,j} \left(\frac{C_{i,j+1}}{C_{i,j}} - f_j \right)^2 , j = 1, \ldots, m-2$$

Eftersom detta arbete inte riktigt handlar om denna metod, så överlämnas det åt läsaren att söka efter härledningarna i litteratur som anges i referenser.

Teori, beräkning av överlevnads- och dödsfallssannolikheter

Livslängd och återstående livslängd

Vi börjar med att definiera begreppet livslängd T för en godtyckligt vald individ där T är icke-negativ kontinuerlig stokastisk variabel, där fördelningsfunktion för T är:

$$F(x) = P(T \leq x), \qquad x \geq 0$$

och täthetsfunktionen är definierat som:

$$f(x) = F^{'}(x), \qquad x \geq 0.$$

Vi tar nu hänsyn till åldern x och låter T_x vara en icke-negativ kontinuerlig stokastisk variabel som får representera den *återstående livslängden* vid åldern x för en godtycklig individ. Denna fördelningsfunktion definieras som:

$$F_x(t) = P(T_x \leq t), \qquad t \geq 0$$

och med detta kan vi nu, genom att använda begreppet livslängd T, skriva:

$$P(T_x > t) = P(T > x + t | T > x) = \frac{P(T > x + t)}{P(T > x)}.$$

Dödlighetsintensitet, $\boldsymbol{\mu_x}$

Dödlighetsintensiteten μ_x är en frekvensvariabel som talar om hur många individer som har dött per tidsenhet (vanligast per år). Sannolikheten att avlida under ett tidsintervall $(x, x + dx)$kan skrivas som $\mu_x dx$ under förutsättningarna att dx är litet och att individen lever vid x. På följande sätt kan man nu härleda dödlighetsintensiteten:

$$\mu_x dx \simeq \frac{F(x + dx) - F(x)}{1 - F(x)}$$

som också kan skrivas

$$\mu_x \simeq \frac{1}{dx} \frac{F(x + dx) - F(x)}{1 - F(x)}$$

Om man nu låter $dx \longrightarrow 0$ får man

$$\mu_x = \lim_{dx \to 0} \frac{1}{dx} \frac{F(x + dx) - F(x)}{1 - F(x)} = \frac{f(x)}{1 - F(x)}$$

Överlevelsefunktion

Vi inför nu överlevelsefunktionen $l_x(t)$, som är sannolikheten att en x-årig individ överlever i ytterligare minst t år genom:

$$l_x(t) = 1 - F_x(t) = P(T_x > t), \qquad t \geq 0.$$

$$l_0(0) = 1$$

Om man nu betraktar en individ som har uppnått åldern x år och dess återstående livslängd är T_x så kan man skriva:

$$P(T_x > t) = \frac{l_0(x+t)}{l_0(x)}$$

som även kan skrivas:

$$l_x(t) = \frac{l_0(x+t)}{l_0(x)}, \qquad t \geq 0$$

där $l_0(t)$ är överlevelsefunktion för en nyfödd. Vi kommer nu att förenkla beteckningarna $T_0, F_0(t), f_0(t)$ och $l_0(t)$ till $T, F(t), f(t)$ respektive $l(t)$.

Nu kan vi beräkna täthetsfunktionen för T_x på följande sätt:

$$P(T_x \leq t) = 1 - \frac{l(x+t)}{l(x)}$$

$$\frac{d}{dt}P(T_x \leq t) = f_x(t) = -\frac{d}{dt}\frac{l(x+t)}{l(x)}$$

och vi vet sedan tidigare att:

$$l_x(t) = 1 - F_x(t) \Rightarrow \frac{d}{dt}l(x+t) = -f(x+t)$$

$$\Rightarrow f_x(t) = \frac{f(x+t)}{l(x)} = \frac{\mu_{x+t}\big(1 - F(x+t)\big)}{l(x)} = \frac{l(x+t)}{l(x)}\mu_{x+t}, \qquad t \geq 0$$

Den ettåriga dödsrisken q_x

Den ettåriga dödsrisken är densamma som sannolikheten för en x-årig individ att dö inom ett år och definieras som:

$$q_x = P(T_x \leq 1) = 1 - P(T_x > 1) = 1 - \frac{l(x+1)}{l(x)}$$

Förväntad återstående livslängd, $E(T_x) = e_x$

Låt X vara en icke-negativ stokastisk variabel med fördelningsfunktion $F(x)$ då uttrycks, enligt definition, det förväntade värdet av X som:

$$E(X) = \int_0^\infty [1 - F(x)]dx$$

Med detta kan vi nu skriva:

$$E(T_x) = e_x = \int_0^\infty [1 - F_x(t)]dt = \int_0^\infty \frac{l(x+t)}{l(x)}dt.$$

Då $x = 0$ får vi:

$$e_0 = \int_0^\infty l(t)dt$$

vilket är den förväntade återstående livslängden för en nyfödd individ.

Vi får då att:

$$e_x = \int_0^\infty -t\frac{l'(x+t)}{l(x)}dt = \int_0^\infty t\frac{l(x+t)}{l(x)}\mu_{x+t}dt =$$
$$= -\frac{1}{l(x)}\left([tl(x+t)]_{t=0}^{t=\infty} - \int_0^\infty l(x+s)ds\right).$$

Kopplingarna mellan μ_x och $l(x)$ respektive q_x och $l(x)$

Kopplingen mellan μ_x och $l(x)$ är:

$$\mu_x dx = \frac{f(x)dx}{1 - F(x)} = \frac{-l'(x)dx}{l(x)}$$

$$\mu_x = \frac{-l'(x)}{l(x)} = \frac{-d(\ln l(x))}{dx}$$

$$\int_0^x \mu_t dt = \int_0^x \frac{d}{dt}(-\ln l(t))dt = [-\ln l(t)]_{t=0}^{t=x} = \ln l(0) - \ln l(x) = -\ln l(x)$$

$$\Rightarrow l(x) = exp\left(-\int_0^x \mu_t dt\right)$$

Kopplingen mellan q_x och $l(x)$ är:

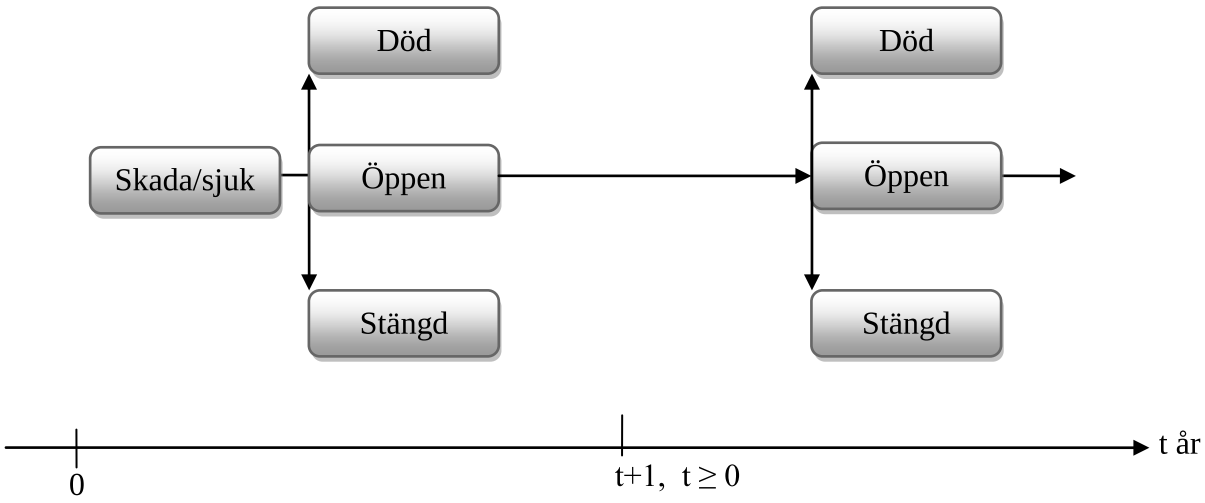
$$q_x = P(T_x \leq 1) = 1 - P(T_x > 1) = 1 - \frac{l(x+1)}{l(x)}$$

$$\Longrightarrow q_x = 1 - exp\left(-\int_x^{x+1} \mu_t dt\right).$$

Skattningar av den ettåriga dödlighetssannolikheten $q_{x,t}$

Metod

En skadelidande kan alltså hamna i tre olika tillstånd och för att få en begriplig uppfattning om detta så kan man med en förenklad bild nedan se hur det ser ut.



Låt nu N(x,t) vara antalet skadade vid en viss ålder x samt duration t och $D(x,t)$ vara antalet skadelidande som dör under år t. Åldern x är den ålder som skadan inträffade plus durationen. Vi kan skatta den ettåriga dödsrisken $q_{x,t}$ med:

$$\hat{q}_{x,t} = \frac{D(x,t)}{N(x,t)}$$

där den betingade fördelningen för $D(x,t)$, givet alla tidigare $N(x,s)$ och $D(x,s)$ för $s \le t$, är binomialfördelad med parametrarna $(N(x,t), q_{x,t})$. Variansen och väntevärdet för dödsrisken ges då av:

Ty $E(D) = N * q$

$$var(D) = Nq(1-q)$$

och

$$E(\hat{q}) = E\left(\frac{D}{N}\right) = \frac{N * q}{N} = q$$

$$var(\hat{q}) = var\left(\frac{D}{N}\right) = \frac{q(1-q)}{N} \approx \frac{q}{N}$$

Makehamanpassning

För att uppnå en mer utjämnad skattning av dödligheten så anpassar vi en Makehamfunktion till observerad dödlighet. Den parametriska dödlighetsmodellen som vi skall arbeta med är:

$$\mu_x = \alpha + \beta e^{\gamma x}$$

där $\alpha + \beta > 0, \beta > 0, \gamma > 0$ och $x \geq 0$

För att skatta parametrarna α, β och γ så använder vi den modifierade χ^2-metoden, vilket innebär att man bildar:

$$Q = \sum_{i=1}^{n} \omega_{x_i} \left(\hat{\mu}_{x_i} - \alpha - \beta e^{\gamma x_i}\right)^2$$

där n är antalet individer och x_i är åldern för individ i. Därefter väljer man de parametrar som gör så att Q blir så litet som möjligt. Det görs enkelt med excelverktyget problemlösaren. Vikterna ω_{x_i} väljs som:

$$\omega_{x_i} = \frac{N_{x_i}}{\hat{\mu}_{x_i}}$$

där N_{x_i} är antalet individer i början av ett kalenderår. Vi löser nu ekvationssystemet:

$$\frac{\partial Q}{\partial \alpha} = 0$$

$$\frac{\partial Q}{\partial \beta} = 0$$

och man betraktar då γ som fixt. Därefter skriver vi skattningarna som:

$$\hat{\alpha} = \frac{m_{01} - \hat{\beta} m_{10}}{\omega}$$

och

$$\hat{\beta} = \frac{\omega m_{11} - m_{10} m_{01}}{\omega m_{20} - m_{10}^2}$$

där

$$\omega = \sum_{i=1}^{n} \omega_{x_i},$$
$$m_{10} = \sum_{i=1}^{n} \omega_{x_i} e^{\gamma x_i},$$
$$m_{01} = \sum_{i=1}^{n} \omega_{x_i} \hat{\mu}_{x_i},$$

$$m_{20} = \sum_{i=1}^{n} \omega_{x_i} e^{2\gamma x_i},$$
$$m_{11} = \sum_{i=1}^{n} \omega_{x_i} e^{x_i} \hat{\mu}_{x_i}.$$

För kvinnor får vi parametrarna, där t är som nämnt ovan antalet år efter skadetillfället:

	α	**B**	**γ**	**Q**
t=0	-0,001160	0,002043	0,156750	55,288000
t=1	0,000341	0,000413	0,323281	13,131000
t=2	0,000001	0,000239	0,348288	11,334000
t≥3	-0,001373	0,001328	0,076122	6,724000

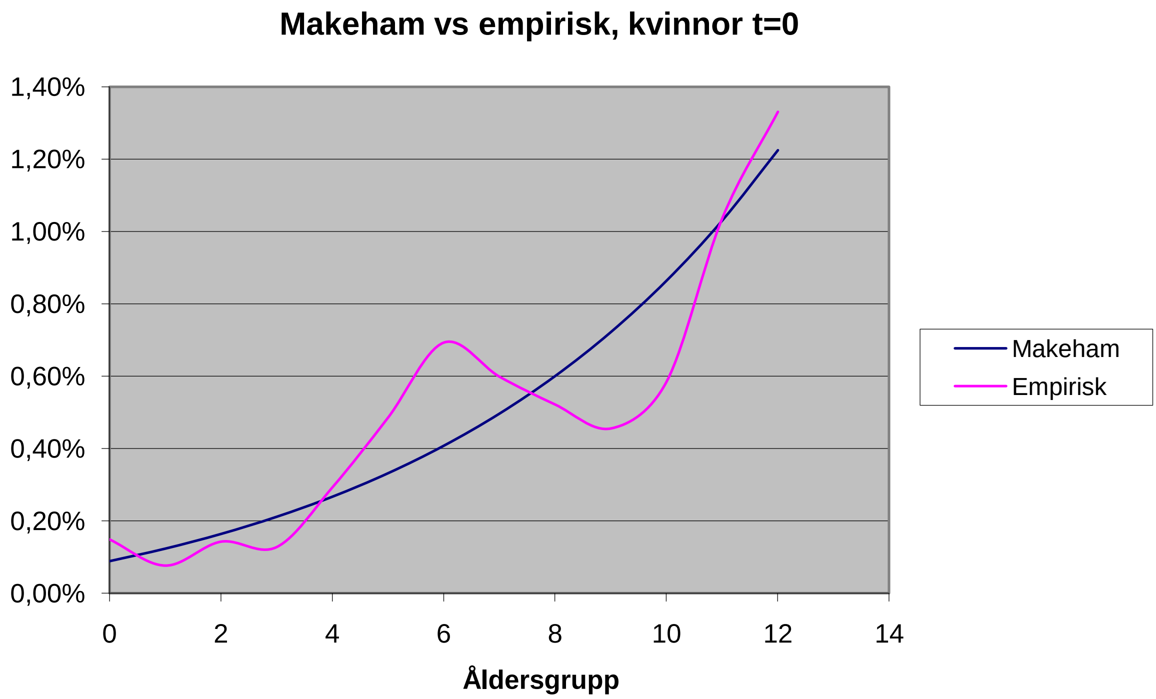
och för män så får vi:

	α	**B**	**γ**	**Q**
t=0	0,000957	0,000283	0,328901	21,878000
t=1	0,000712	0,000099	0,452317	5,224400
t≥2	0,000109	0,000038	0,457612	9,046200

Eftersom vi har en sådan databrist så har vi summerat antalet individer som avled efter tre år för kvinnor och två år för män. Vi har också här grupperat åldrarna i femårsintervaller av samma anledning, så att åldergrupp 0 innebär åldrarna 0 – 4 osv.

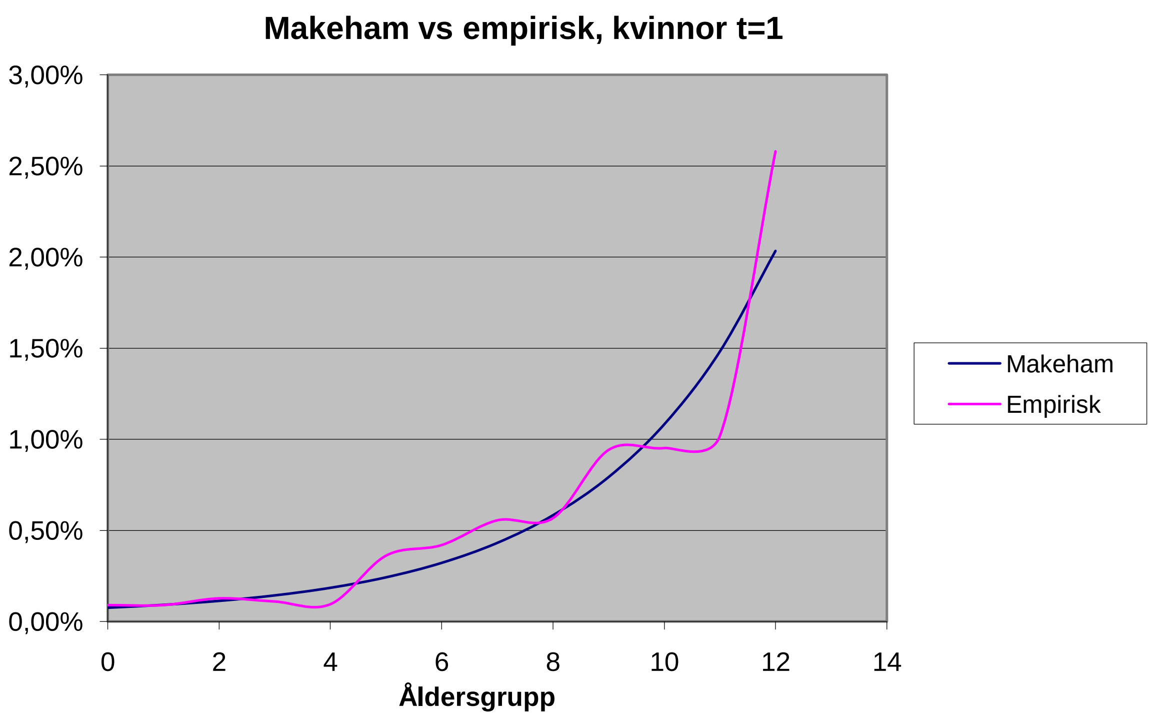
Nedan kan vi se resultaten av de ettåriga dödsriskerna som Makehams modell medför.

I figur 1 så kan vi se en Makehamanpassning för kvinnor, som avled samma år som skadeåret, jämfört med det empiriska datat. Den ser ut att följa relativt bra trots att det svänger rejält överallt, speciellt mellan grupperna 4 – 10 där det skiljer sig med upp till ca 0,3 %. Den lilla förhöjningen vid åldersgrupp 6 beror troligtvis på att de har relativt svårare sjukdomar eller mer psykisk ohälsa än de övriga åldersgrupperna.



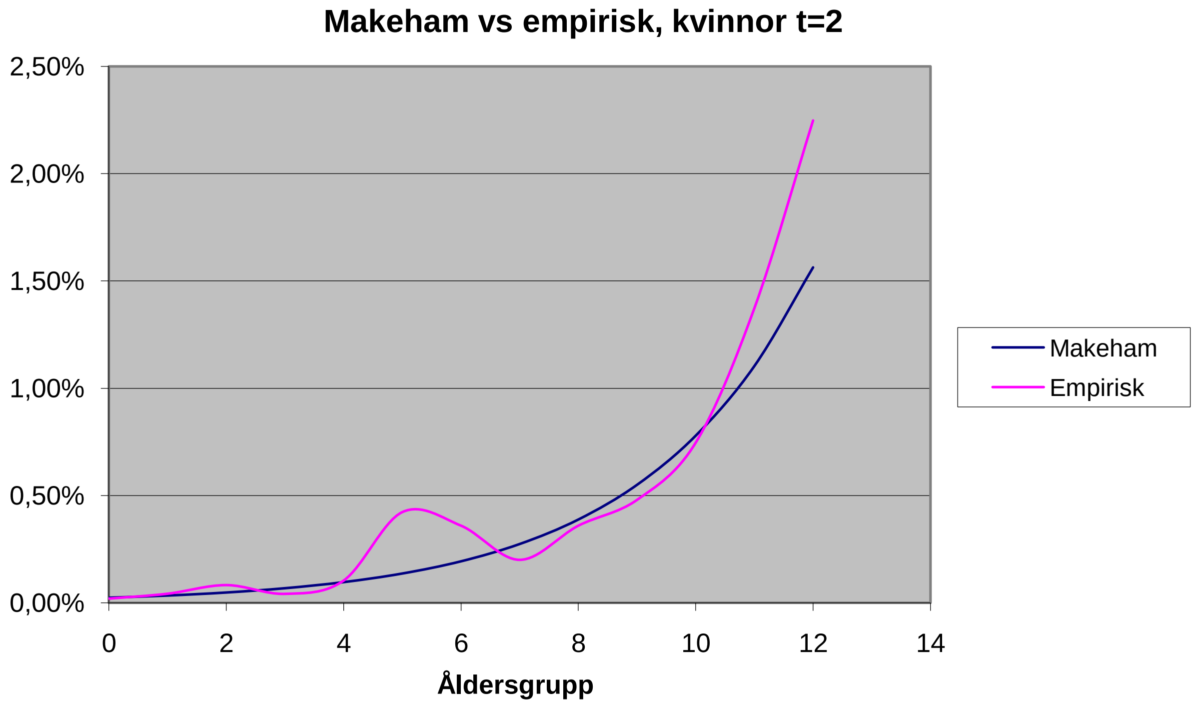
Figur 1

I figur 2 har vi en någon bättre anpassning än ovan. Det är kvinnor som avled ett år efter skadetillfället. Anpassningen ser speciellt bra ut för åldersgrupperna 0 – 3 och sedan börjar det även här svänga, men inte lika kraftigt. Vi har inte samma förhöjda dödlighet för åldersgrupp 0 som i figur 1.



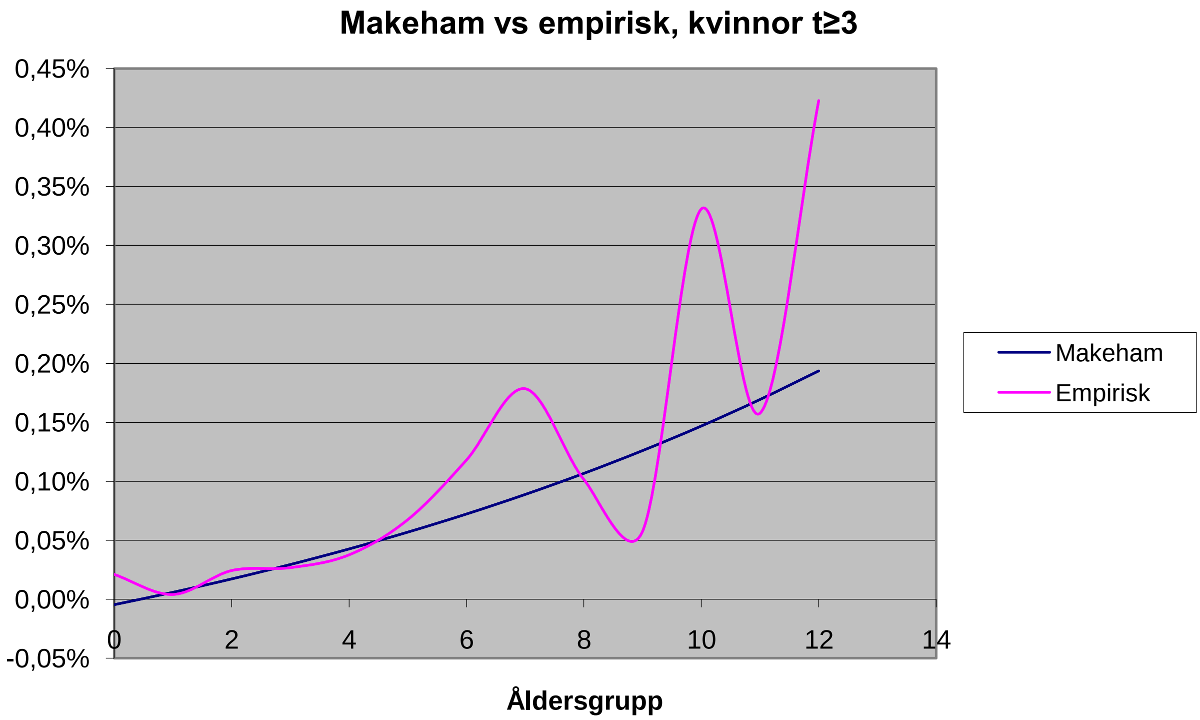
Figur 2

I figur 3 är anpassningen även här relativt bra. Vi underskattar dödligheten mellan åldersgrupp 4 – 7 med upp till 0,3 %, men som mest för åldersgrupp 12 där det blir en underskattnig på ca 0,65 %. Det är även här en förhöjd dödlighet vid åldersgrupp 5 – 6 och jämn dödlighet för åldersgrupp 0.



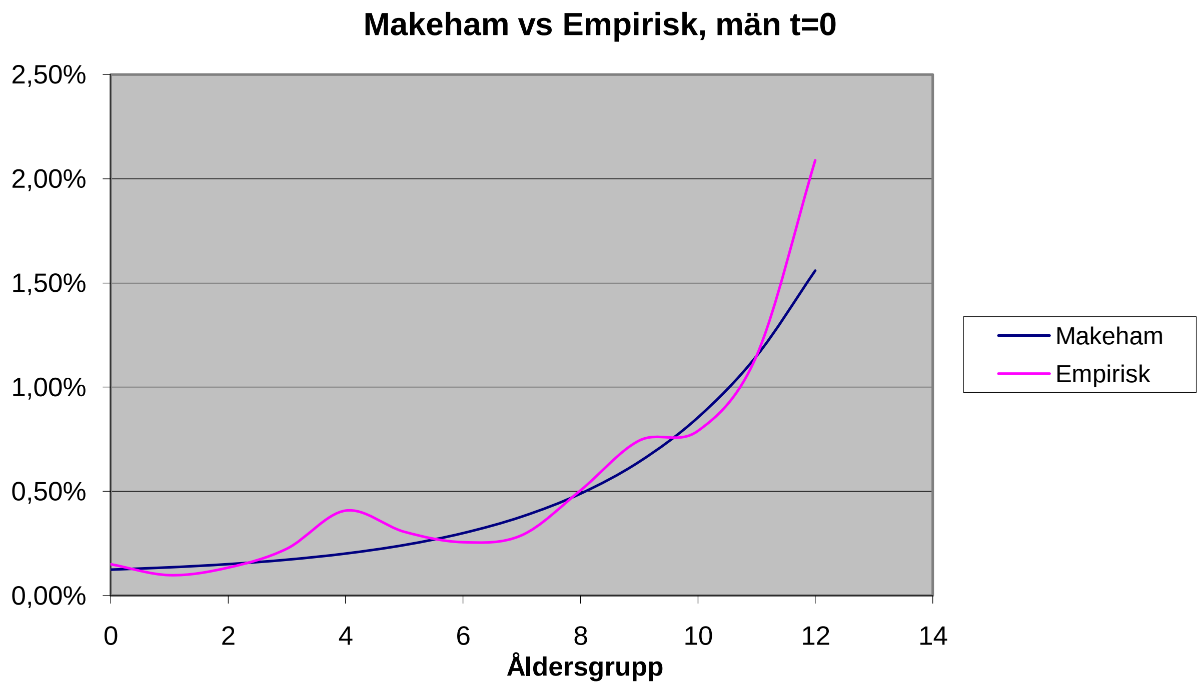
Figur 3

I figur 4 ser vi den ettåriga dödligheten för kvinnor som avled efter 3 år eller senare. Här ser anpassningen relativt dåligt ut då vi bl.a. får ett negativt värde för åldersgrupp 0 och att det svänger väldigt kraftigt efter åldersgrupp 5, men ändå inte så farligt då den största avvikelsen är vid åldersgrupp 12 och är bara 0,2 %. Eftersom åldrarna är återigen åldern vid skadetillfället plus duration så finns här inga barn som är yngre än 3 år och detta leder till att kurvan bestäms mer av de äldre barnen. Att barngrupper dominerar beror på att barnskador tar längre tid att slutreglera än vuxenskador, dessutom är Trygg-Hansa marknadsledande inom detta segment. En annan orsak till det kan vara att vi har bättre data för de låga åldrarna än för de höga som gör att den skeva tyngden påverkar anpassningen på detta sätt. Det vi gör här är att vi väljer det empiriska datat för åldersgrupp 0.
Man kan också tycka att anpassningen här borde ”lyftas upp lite” för de högre åldrarna. Man kan även argumentera för att summera ihop alla dödsfall som skedde efter 2 år och senare istället för 3 år och senare.



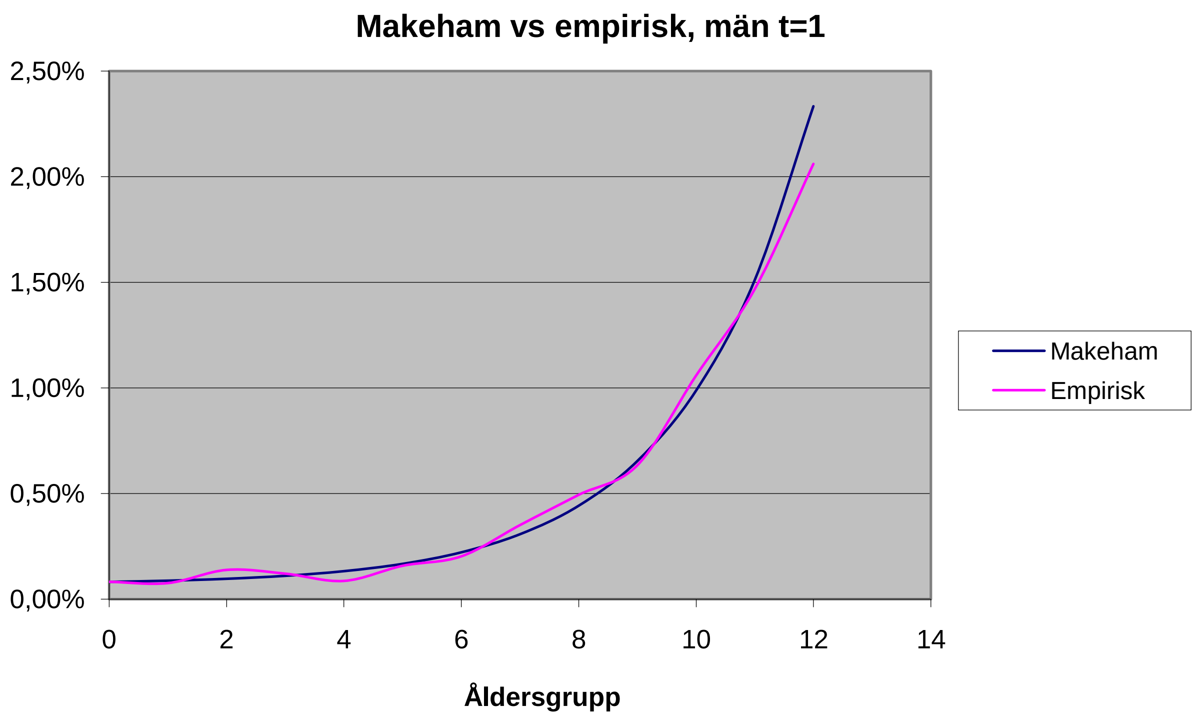
Figur 4

I figur 5 nedan har vi dödligheten för män som avled samma år som skadetillfället. Anpassningen här ser relativt bra ut. Det blir en liten underskattning av dödligheten vid åldersgrupp 4 respektive 12 och en del överskattningar för övrigt, men i små marginaler.



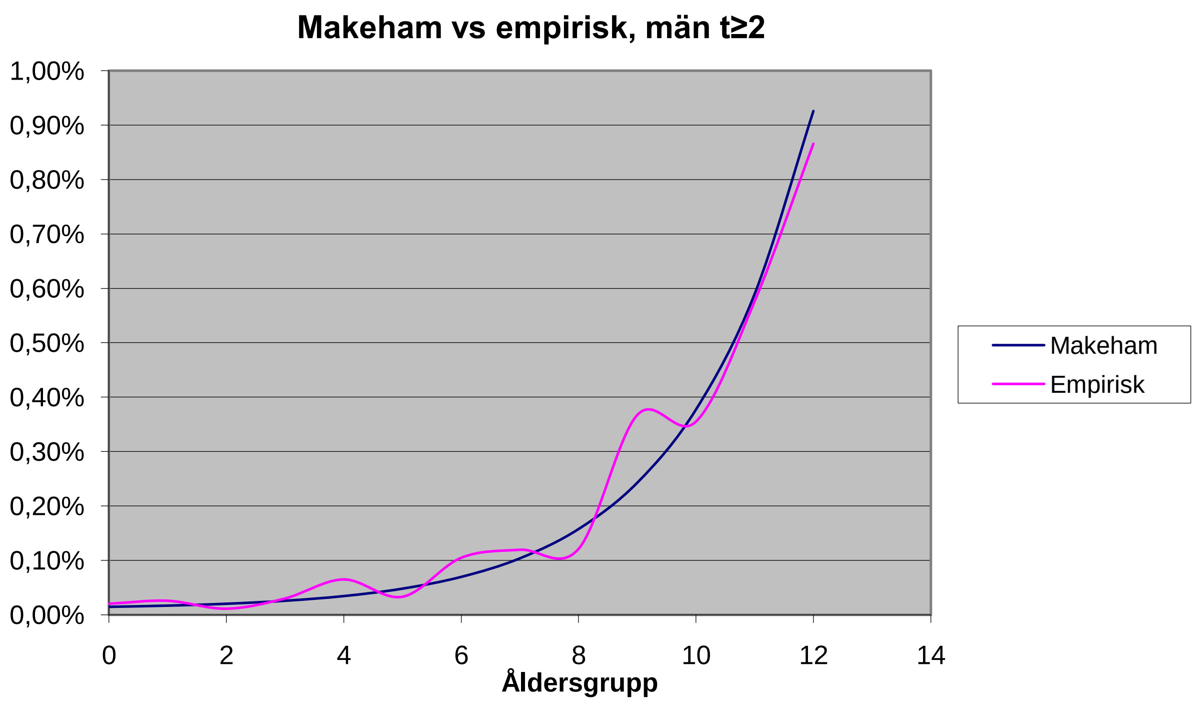
Figur 5

I figur 6 ser vi dödligheten för män som avled ett år efter skadetillfället. Det är en väldigt bra anpassning till det empiriska datat. Visst finns det en del avvikelser, men inget märkvärdigt att kommentera. Vi ser dock inte en förhöjd dödlighet för åldersgrupp 0 som i figur 5.



Figur 6

I figur 7 ser vi dödligheten för män som avled efter två år eller senare. Den påminner lite om figur 6, men i en mindre skala. Vi får en del svängningar mellan åldersgrupp 3 – 10, men ser för övrigt bra ut.

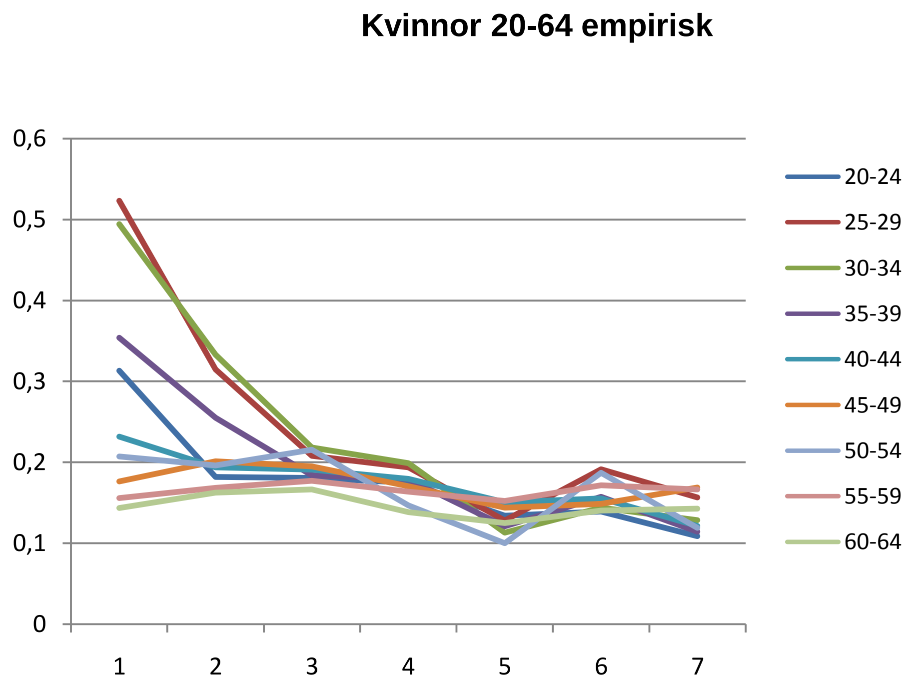


Figur 7

Skattning av sannolikheten att en skada stängs inom ett år

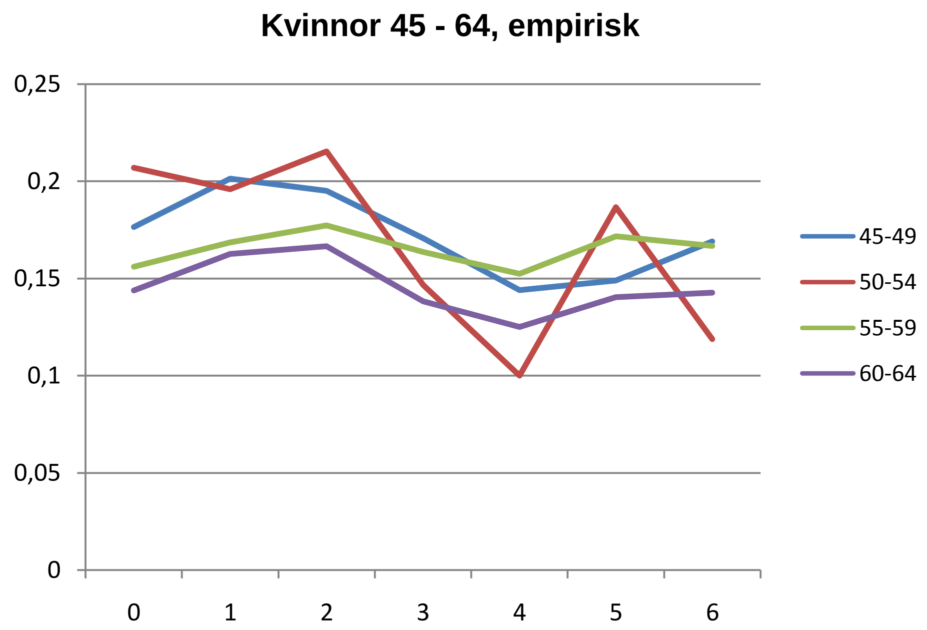
Tvåsidig indelning

I figur 1 så kan vi se sannolikheten att en skada stängs inom ett år för kvinnor i åldrarna 20 – 64. Man kan ju se att kurvorna konvergerar mot 10 – 20 % nivå. Tiden i x-led är i kalenderår.



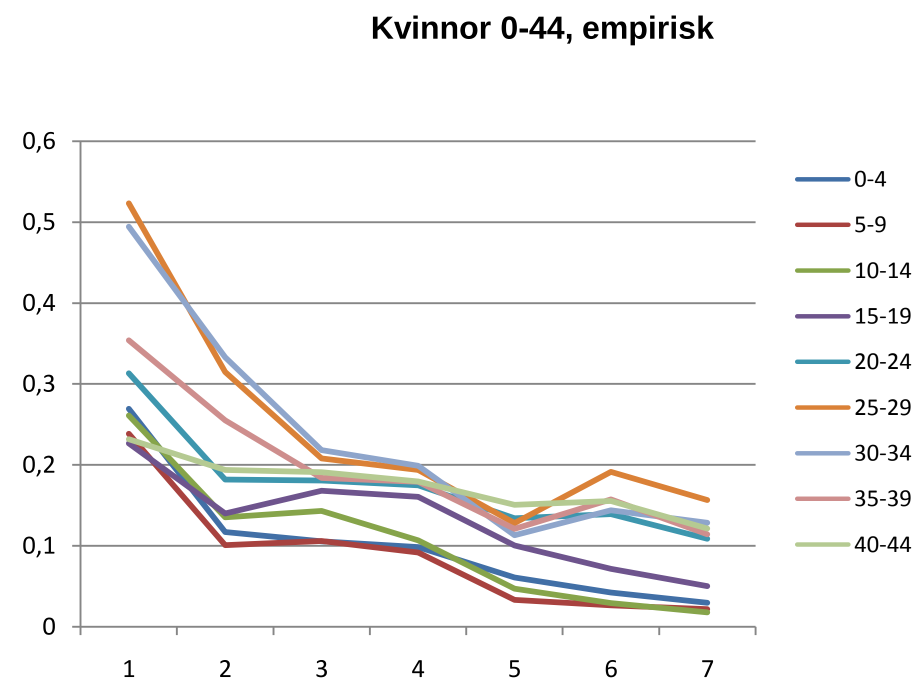
Figur 1

I figur 2 nedan så kan vi se utvecklingen för kvinnor i åldrarna 45 – 64 och att sannolikheten för att en skada skall stängas är högst då det har gått ca 1 – 2 år efter skadetillfället och lägst efter ca 4 år för att sedan gå upp igen. För åldrarna 50 – 54 så ser man att det nästan är samma effekter men alldeles för stora avvikelser. Detta beror på, med största sannolikhet, att datat inte är tillräckligt.



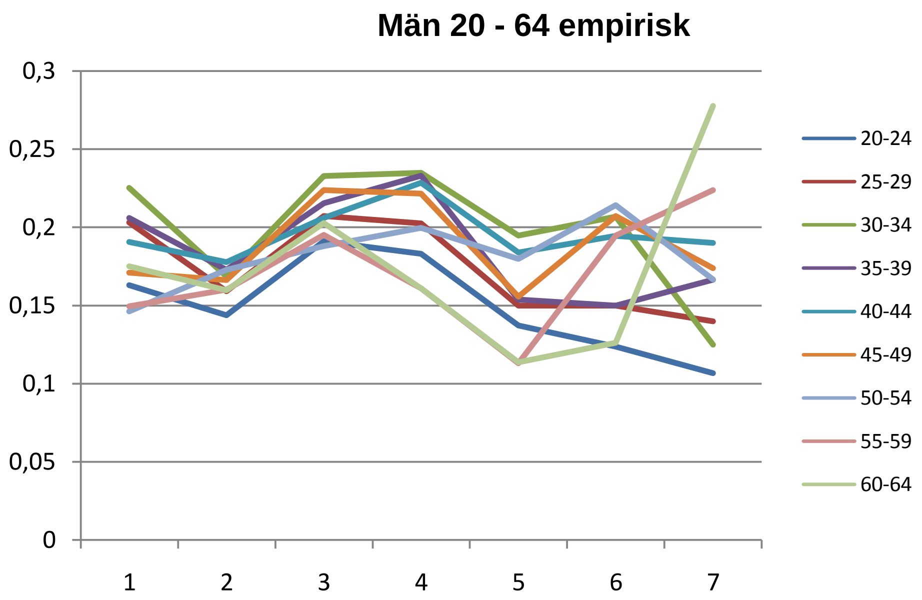
Figur 2

I figur 3 så kan vi se kvinnor mellan åldrarna 0 – 19. Här ser man ett tydligt mönster mellan dessa åldersgrupper. Det är lite högre sannolikhet att en skada stängs för den äldsta gruppen här och det beror nog på att det är svårare att bedöma vilken slags sjukdom de yngre har etc



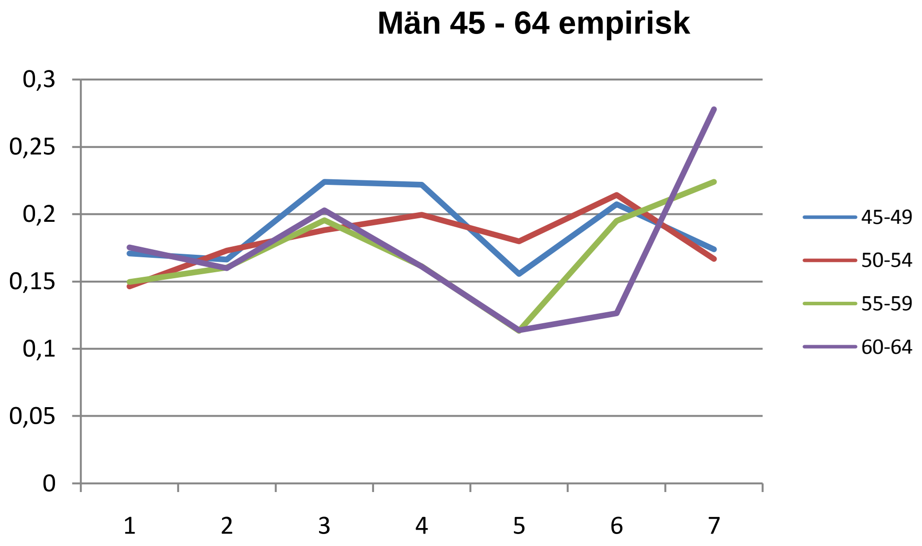
Figur 3

I figur 4 så kan vi se sannolikheten att en skada stängs inom ett år för män i åldrarna 20 – 64. Här ser man ett mönster i de första åren som förflutit men sedan börjar spridningen bli alltför stor för att kunna se ett mönster.



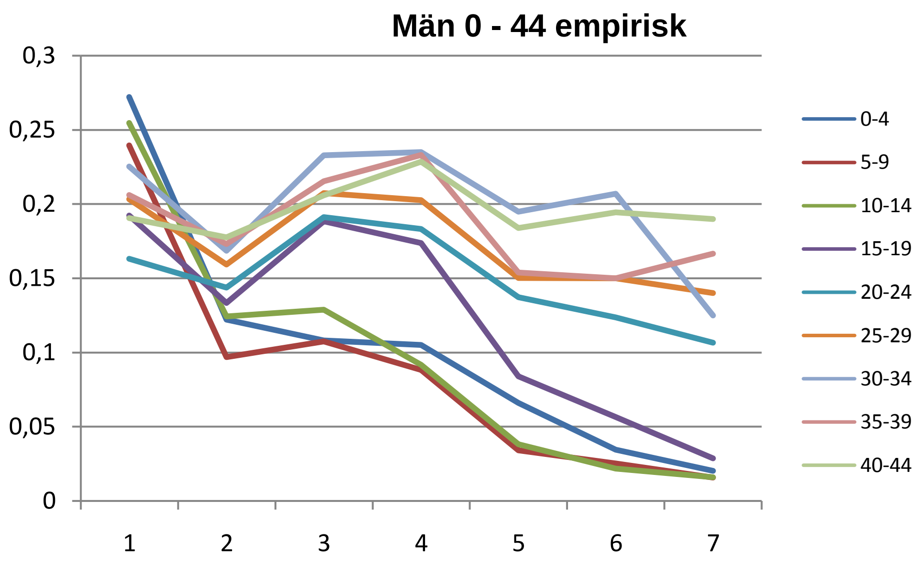
Figur 4

I figur 5 kan man se utvecklingen av män i åldrarna mellan 45 och 64 år. För åldrarna 55 – 64 kan man se ett väldigt tydligt mönster mellan åren 1 – 4. Utöver det finns knappt något mönster man kan följa förutom att kurvorna ligger i snitt på 17 till 18,5 procents nivå. Detta är sannolikt data som inte är tillräckligt och måste därför behandlas.



Figur 5

I figur 6 så har vi män i åldrarna 0 – 44. Det ser ut att vara ganska stora glapp mellan de olika åldersgrupperna men de följer ändå ett mönster och att de lägre åldrarna har en lägre sannolikhet att stängas. Förklaringen till det är även här att det tar längre tid att bedöma vilken slags sjukdom barnen har och man väntar tills de blir äldre för att ta dessa beslut.



Figur 6

Eftersom vi har två faktorer som har inverkan på försöksutfallet dvs. ålder och duration, så kan vi använda oss av en variansanalytisk modell där Y_{ij} = antalet stängda fall i cell (i, j) med i = ålder och j = duration. Detta gör vi för åldrarna 45-64. Anledningen till detta är att sannolikheten att en skada stängs för dessa åldrar bör ha ett mer gemensamt mönster, men på grund av databrist så blir fallet inte så. Denna metod kommer leda till att avstånden mellan kurvorna minimeras och får ett gemensamt mönster. Detta görs med följande teori, där vår grundmodell är:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \qquad i = 1, \dots . r; \;\; j = 1, \dots . , s;$$

$$\varepsilon_{ij} \sim N(0, \sigma^2) \text{ och sinsemellan oberoende.}$$

och parametrarna kommer att uppfylla:

$$\sum_{i=1}^{r} \alpha_i = \sum_{j=1}^{s} \beta_j = 0$$

Skattningarna av parametrarna beräknas på följande sätt:

$$\begin{aligned} \hat{\mu} &= \bar{y}_{..} \\ \hat{\alpha}_i &= \bar{y}_{i.} - \bar{y}_{..} \;\;, \qquad i = 1, \dots . , r \\ \hat{\beta}_j &= \bar{y}_{.j} - \bar{y}_{..} \;\;, \qquad j = 1, \dots . , s \end{aligned}$$

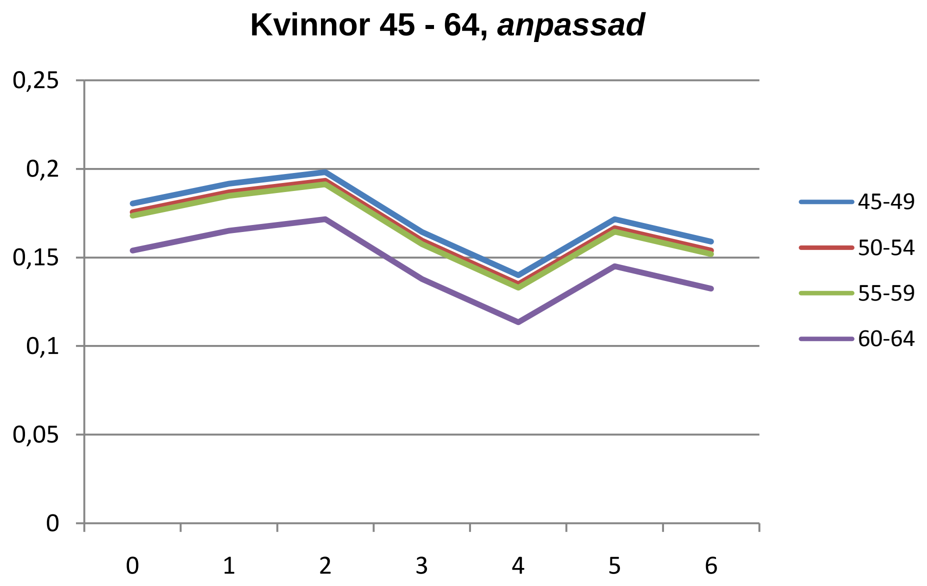
där

$$\begin{aligned} \bar{y}_{..} &= \frac{1}{r} \sum_{i=1}^{r} \bar{y}_{i.} = \frac{1}{s} \sum_{j=1}^{s} \bar{y}_{.j} \\ \bar{y}_{i.} &= \frac{1}{s} \sum_{j=1}^{s} y_{ij} \\ \bar{y}_{.j} &= \frac{1}{r} \sum_{i=1}^{r} y_{ij} \end{aligned}$$

Den skattade variansen fås då enligt nedan

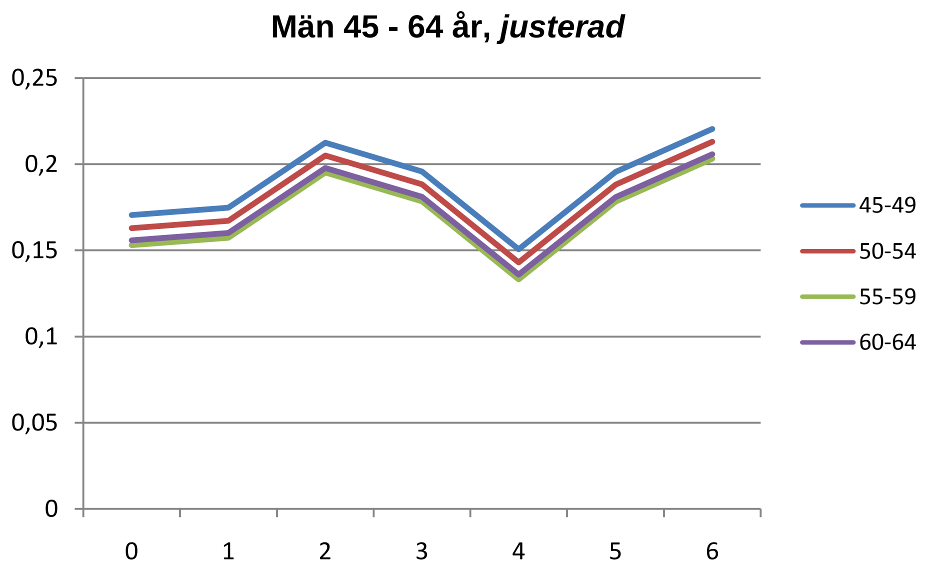
$$\hat{\sigma}^2 = \frac{1}{(r-1)(s-1)} \sum_{i=1}^{r} \sum_{j=1}^{s} \left(y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..} \right)^2$$

I figur 7 nedan ser vi samma modell som i figur 2 ovan, men här är datat justerat och det ser lite mer pålitligt ut. Sannolikheten att skadan stängs är fortfarande högst då det har gått ca 2 år efter skadetillfället och lägst då det har gått ca 4 år. Nu kan man även se att kurvan för åldrarna 50 – 54 ligger mittemellan åldrarna 45 – 49 respektive 55 – 59.



Figur 7

Nedan, i figur 8, ser vi en anpassning av figur 5 ovan. Detta mönster påminner väldigt mycket om kvinnor mellan 45 och 64 år. De startar på 15 till 20-procentsnivån och går sedan upp de två första åren med ca 5 procentenheter. Till skillnad från kvinnor så ökar sannolikheterna från fyra till sex år efter skadeåret.



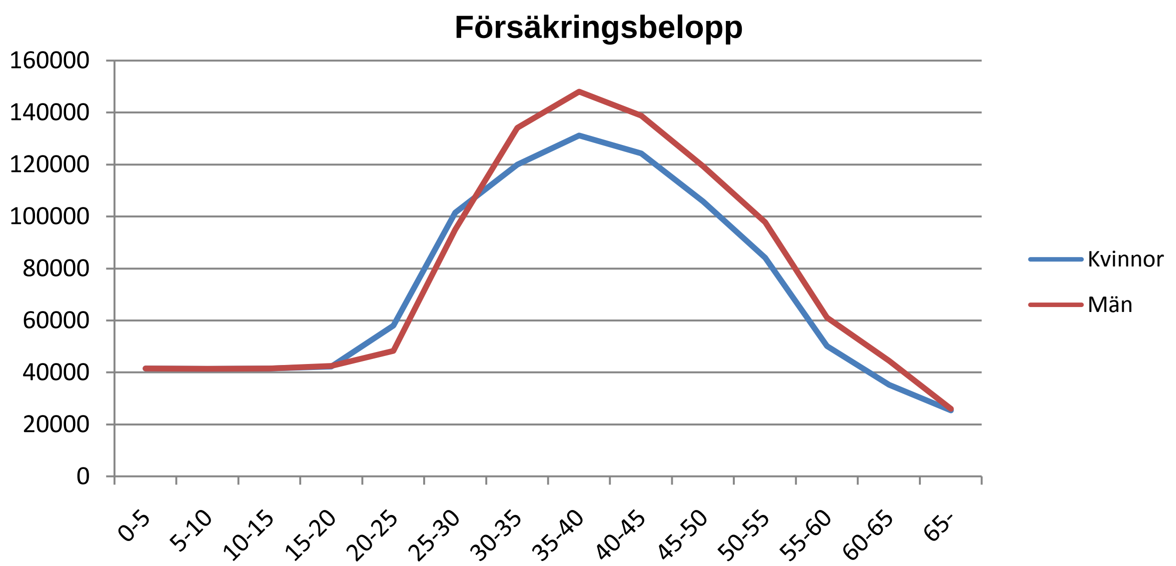
Figur 8

Försäkringsbelopp

Försäkringsbelopp är, i detta fall, det belopp som man vill att ens familj eller nära och kära skall få i ersättning av försäkringsbolaget om den försäkrade avlider. Den försäkrade väljer själv hur mycket som skall betalas ut och det bestäms av hur mycket premier denne är beredd att betala. Försäkringsbeloppet kan variera kraftigt mellan olika kön och åldrar.

Nedan kan vi se hur dessa försäkringsbelopp kan se ut. För de låga åldrarna ligger försäkringsbeloppen på ca 40 000 kr och ökar sedan, med en faktor 3 – 4, från 20 års ålder till och med 40 års ålder. En förklaring till detta är att det oftast är föräldrar som tecknar dessa försäkringar till barn/ungdomar mellan 0 och 20 år, och att det anses oetiskt att en förälder skall kunna få ut ett större försäkringsbelopp vid ett barns dödsfall. Efter 40 år blir det en relativt linjär minskning av beloppen till och med 65 års ålder.

Man kan också se en klar skillnad mellan de båda könen med. Vid 35 till 40 års ålder har männen ett försäkringsbelopp som överstiger kvinnornas med ett gap på ca 16 000 kr. Mellan åldrarna 0 och 30 är det relativt jämt mellan kvinnor och män, men därefter har männen ständigt högre belopp. En hypotes till detta fenomen är att män har högre löner än kvinnor, speciellt efter 30 års ålder och en annan är att män är något mer själviska än kvinnor och spenderar mer pengar på sig själva medans kvinnor även försäkrar sina barn.



Ibnr-reserv, beräkning

Eftersom detta påminner om en markovmodell, så beräknar vi ibnr-reserverna genom att använda oss av en modifierad variant av retrospektivreserven/värdefunktionen, som är det förväntade nuvärdet av framtida betalningar fram till sluttiden T.

$$V_{x,t} = q_{x,t} * F_x + \left(1 - q_{x,t} - P_{x,t}\right) * V_{x,t+1} * d$$

där $q_{x,t}$ är dödsrisken för en individ i ålder x vars tid som förflutit sedan skadetillfället är t år och F_x är det försäkringsbelopp som den försäkrades dödsbo får om man avlider i ålder x. Sannolikheten att en skada kvarstår öppen beräknas genom $\left(1 - q_{x,t} - P_{x,t}\right)$ där $P_{x,t}$ står sannolikheten att en skada stängs. Variabeln d är en diskonteringsfaktor och bestäms av förmånens indexering, som vi för enkelhetens skull låter vara konstant, och sätter den till d = 1/1,02.

Då får vi till slut den totala ibnr-reserven genom att summera ihop den med avseende på ålder och duration på följande sätt:

$$\sum_{x=0}^{X} \sum_{t=0}^{T} V_{x,t}$$

Slutsats

En Makehamanpassning till dödligheten är en metod som används i de allra flesta fall för att beräkna den ettåriga dödsrisken. Man tar då hänsyn till åldrar, kön och duration till skillnad från chainladder-metoden som endast tar hänsyn till trend. Den här omvandlingen kommer att med stor sannolikhet att minska variansen för beräkning av ibnr-reserverna. I sakförsäkringsmatematisk teori står det klart och tydligt att det inte är lämpligt att använda chainladder-metoden för livförsäkring, vilket detta faktiskt är.

Anledningen till att jag grupperade åldrar i femårsgrupper var att det inte gick att göra någon analys samt att det inte gick att förstå sig på resultat på grund av att det har varit ganska dåligt med data. Vi grupperade även ihop alla som var äldre än 64 år till en och samma grupp av samma anledning.

Det som nu kvarstår är att implementera detta i Trygg-Hansas datamiljöer inför den sista kvartalsanalysen 2008. Innan detta görs måste en del personer, som berörs av denna omvandling, övertygas om att detta är en bättre metod.

Denna uppsats är en grund som Trygg-Hansa sedan får bygga vidare på när de har bättre statistik.

Appendix

	Emirisk qx				Makeham qx			
Ålder\duration	t=0	t=1	t=2	t≥3	t=0	t=1	t=2	t≥3
0 – 4	0,15%	0,09%	0,02%	0,02%	0,09%	0,08%	0,02%	-0,004%
5 – 9	0,08%	0,09%	0,04%	0,00%	0,12%	0,09%	0,03%	0,01%
10 – 14	0,14%	0,13%	0,08%	0,02%	0,16%	0,11%	0,05%	0,02%
15 – 19	0,13%	0,11%	0,04%	0,03%	0,21%	0,14%	0,07%	0,03%
20 – 24	0,29%	0,09%	0,10%	0,04%	0,27%	0,18%	0,10%	0,04%
25 – 29	0,48%	0,36%	0,42%	0,07%	0,33%	0,24%	0,14%	0,06%
30 – 34	0,69%	0,42%	0,36%	0,12%	0,41%	0,32%	0,19%	0,07%
35 – 39	0,60%	0,55%	0,20%	0,18%	0,50%	0,43%	0,27%	0,09%
40 – 44	0,52%	0,56%	0,36%	0,10%	0,60%	0,58%	0,39%	0,11%
45 – 49	0,45%	0,94%	0,48%	0,06%	0,72%	0,79%	0,55%	0,13%
50 – 54	0,58%	0,95%	0,74%	0,33%	0,86%	1,08%	0,78%	0,15%
55 – 59	1,03%	1,01%	1,36%	0,16%	1,03%	1,48%	1,10%	0,17%
60 – 64	1,32%	2,55%	2,22%	0,42%	1,22%	2,03%	1,56%	0,19%

Figur 8, Ettåriga dödsrisker för kvinnor, uppdelade i femårsgrupper. Den markerade cellen högst upp till höger har vi, som tidigare nämnt, justerat upp till det empiriska datat.

	Emirisk qx			Makeham qx		
Ålder\duration	t=0	t=1	t≥2	t=0	t=1	t≥2
0 – 4	0,15%	0,08%	0,02%	0,12%	0,08%	0,01%
5 – 9	0,10%	0,08%	0,03%	0,14%	0,09%	0,02%
10 – 14	0,13%	0,14%	0,01%	0,15%	0,10%	0,02%
15 – 19	0,22%	0,12%	0,03%	0,17%	0,11%	0,03%
20 – 24	0,41%	0,09%	0,06%	0,20%	0,13%	0,03%
25 – 29	0,30%	0,16%	0,03%	0,24%	0,17%	0,05%
30 – 34	0,25%	0,20%	0,10%	0,30%	0,22%	0,07%
35 – 39	0,29%	0,35%	0,12%	0,38%	0,31%	0,10%
40 – 44	0,50%	0,49%	0,12%	0,49%	0,44%	0,16%
45 – 49	0,74%	0,63%	0,37%	0,64%	0,65%	0,24%
50 – 54	0,79%	1,05%	0,35%	0,85%	0,99%	0,38%
55 – 59	1,14%	1,46%	0,57%	1,15%	1,51%	0,59%
60 – 64	2,07%	2,04%	0,86%	1,56%	2,33%	0,93%

Figur 2, Ettåriga dödsrisker för män, uppdelade i femårsgrupper.

Ålder\duration	t=0	t=1	t=2	t=3	t=4	t=5	t≥6
0-4	27,22%	12,20%	10,80%	10,52%	6,60%	3,46%	2,02%
5-9	23,95%	9,69%	10,76%	8,84%	3,41%	2,53%	1,59%
10-14	25,47%	12,44%	12,89%	9,18%	3,82%	2,18%	1,60%
15-19	19,24%	13,35%	18,84%	17,39%	8,42%	5,66%	2,88%
20-24	16,32%	14,38%	19,12%	18,32%	13,73%	12,38%	10,67%
25-29	20,32%	15,92%	20,74%	20,27%	15,01%	22,50%	18,68%
30-34	22,54%	16,86%	23,29%	23,50%	19,49%	20,69%	12,50%
35-39	20,60%	17,30%	21,55%	23,30%	15,40%	22,68%	16,67%
40-44	19,07%	17,78%	20,60%	22,86%	18,39%	19,46%	20,74%
45-49	17,09%	16,62%	22,40%	22,18%	15,56%	20,73%	17,39%
50-54	14,62%	17,28%	18,81%	19,96%	17,99%	21,43%	16,67%
55-59	14,95%	16,02%	19,53%	16,11%	11,32%	19,49%	22,39%
60-64	17,53%	15,98%	20,28%	16,11%	11,38%	12,63%	27,78%

Figur 3, Den ettåriga sannolikheten, för män, att en skada stängs innan den skadelidande dör. Empirisk data Px.

.

Ålder\duration	t=0	t=1	t=2	t=3	t=4	t=5	t≥6
45-49	17,04%	17,46%	21,24%	19,58%	15,05%	19,56%	22,04%
50-54	16,29%	16,72%	20,50%	18,83%	14,31%	18,81%	21,30%
55-59	15,30%	15,73%	19,50%	17,84%	13,31%	17,82%	20,31%
60-64	15,57%	16,00%	19,77%	18,11%	13,58%	18,09%	20,58%

Figur 4, Justerad data med en variansanalytisk modell för män och för åldrarna 45-64. Px.

Ålder\duration	t=0	t=1	t=2	t=3	t=4	t=5	t≥6
0-4	26,94%	11,73%	10,59%	9,87%	6,09%	4,23%	2,98%
5-9	23,84%	10,06%	10,58%	9,15%	3,32%	2,61%	2,16%
10-14	26,10%	13,50%	14,30%	10,66%	4,70%	2,92%	1,77%
15-19	22,67%	14,01%	16,82%	16,05%	10,03%	7,16%	5,00%
20-24	31,31%	18,19%	18,08%	17,47%	13,42%	13,94%	10,88%
25-29	52,33%	31,48%	20,80%	19,37%	12,81%	19,16%	15,64%
30-34	49,46%	33,29%	21,81%	19,87%	11,29%	14,37%	12,83%
35-39	35,39%	25,51%	18,36%	17,91%	12,09%	15,73%	11,41%
40-44	23,16%	19,38%	19,11%	17,97%	15,07%	15,55%	12,15%
45-49	17,65%	20,14%	19,51%	17,06%	14,40%	14,89%	16,91%
50-54	20,70%	19,60%	21,55%	14,67%	10,00%	18,68%	11,88%
55-59	15,61%	16,86%	17,73%	16,37%	15,24%	17,17%	16,67%
60-64	14,38%	16,27%	16,67%	13,83%	12,50%	14,04%	14,27%

Figur 5, Den ettåriga sannolikheten, för kvinnor, att en skada stängs innan den skadelidande dör. Empirisk data Px.

Ålder\duration	t=0	t=1	t=2	t=3	t=4	t=5	t≥6
45-49	18,05%	19,18%	19,83%	16,45%	14,00%	17,16%	15,90%
50-54	17,55%	18,68%	19,33%	15,95%	13,50%	16,66%	15,40%
55-59	17,35%	18,48%	19,12%	15,75%	13,30%	16,46%	15,20%
60-64	15,39%	16,52%	17,17%	13,79%	11,34%	14,50%	13,24%

Figur 6, Justerad data med en variansanalytisk modell för kvinnor och för åldrarna 45-64 Px.

Ålder\duration	0	1	2	3	4	5	6
0 – 4	72,66%	87,72%	89,18%	89,47%	93,39%	96,53%	97,96%
5 – 9	75,91%	90,22%	89,22%	91,14%	96,57%	97,45%	98,40%
10 – 14	74,38%	87,47%	87,09%	90,80%	96,16%	97,80%	98,37%
15 – 19	80,59%	86,54%	81,13%	82,58%	91,56%	94,31%	97,09%
20 – 24	83,48%	85,49%	80,85%	81,65%	86,24%	87,58%	89,29%
25 – 29	79,44%	83,92%	79,22%	79,68%	84,95%	77,45%	81,27%
30 – 34	77,16%	82,91%	76,64%	76,43%	80,44%	79,24%	87,43%
35 – 39	79,02%	82,39%	78,35%	76,59%	84,50%	77,22%	83,23%
40 – 44	80,44%	81,78%	79,24%	76,98%	81,45%	80,39%	79,10%
45 – 49	82,32%	81,88%	78,52%	80,18%	84,71%	80,20%	77,71%
50 – 54	82,85%	82,29%	79,13%	80,79%	85,32%	80,81%	78,32%
55 – 59	83,55%	82,76%	79,91%	81,57%	86,10%	81,59%	79,10%
60 – 64	82,87%	81,67%	79,30%	80,96%	85,49%	80,99%	78,50%

Figur 7, den ettåriga sannolikheten, för män, att en skada är öppen, som vi tar fram genom att beräkna: P(öppen)=1-qx-Px, där vi tar de justerade datan för de äldre individerna från figur 4 ovan.

Ålder\duration	0	1	2	3	4	5	6
0 – 4	72,97%	88,19%	89,39%	90,11%	93,89%	95,75%	97,00%
5 – 9	76,03%	89,85%	89,39%	90,84%	96,67%	97,38%	97,84%
10 – 14	73,73%	86,39%	85,65%	89,33%	95,29%	97,06%	98,22%
15 – 19	77,12%	85,85%	83,12%	83,92%	89,94%	92,81%	94,97%
20 – 24	68,42%	81,62%	81,82%	82,49%	86,54%	86,02%	89,08%
25 – 29	47,34%	68,28%	79,06%	80,57%	87,13%	80,78%	84,30%
30 – 34	50,14%	66,38%	77,99%	80,06%	88,64%	85,56%	87,09%
35 – 39	64,12%	74,06%	81,37%	82,00%	87,82%	84,19%	88,50%
40 – 44	76,24%	80,04%	80,51%	81,93%	84,82%	84,34%	87,74%
45 – 49	81,23%	80,03%	79,62%	83,43%	85,87%	82,71%	83,98%
50 – 54	81,58%	80,24%	79,89%	83,90%	86,35%	83,19%	84,45%
55 – 59	81,62%	80,04%	79,77%	84,08%	86,53%	83,37%	84,63%
60 – 64	83,38%	81,44%	81,27%	86,02%	88,46%	85,31%	86,57%

Figur 8, den ettåriga sannolikheten, för kvinnor, att en skada är öppen, som vi tar fram genom att beräkna: P(öppen)=1-qx-Px, där vi tar de justerade datan för de äldre individerna från figur 6 ovan.

Referenser

Meaza Yirga. Modeller för trend i dödlighet hos befolkningen och de försäkrade (2006).

Jörgen Olsén. Modeller och projektioner för dödlighetsintensitet (2005).

Patrik Dahl. Introduction to reserving (2003).

Niels Rietdorf. Chain Ladder Reserving (2008)

Erik Alm, Gunnar Andersson, Bengt von Bahr, Ander Martin-Löf. Livförsäkringsmatematik (2006).

Gunnar Andersson. Livförsäkringsmatematik (2004).