Matematisk statistik
Stockholms universitet

Dödlighetsundersökningar på KPA:s bestånd av förmånsbestämda pensioner

Sven-Erik Larsson

Examensarbete 2006:2

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 **2006:2**,
http://www.math.su.se/matstat

Dödlighetsundersökningar på KPA:s bestånd av förmånsbestämda pensioner

Sven-Erik Larsson*

February 2006

Abstract

KPA:s kunder har betydligt lägre dödlighet än vad deras grundantaganden anger. Kunder med låga försäkringsbelopp avlider i större omfattning än övriga. Skillnaden är rentav större än mellan könen. Kunderna har i almänhet också färre efterlevande än antagandena ager. Resultaten kommer att innebära att KPA sänker dina dödlighetsantaganden.

*Postal address: Matematisk statistik, Stockholms universitet, SE-106 91, Sweden. E-mail: svenerikevaldlarsson@hotmail.com. Supervisor: Joanna Tyrcha.

ABSTRACT

The mortality of the customers of KPA is lower than expected. The numbers of relatives is also lower. Customers with lower level of insurance have higher mortality rate. This difference is even greater than the difference between the sexes. The results will be used for decreasing the expectance of future mortality rates.

FÖRORD

Denna studie om dödlighet utgör mitt examensarbete för magisterexamen i matematisk statistik vid Stockholms universitet. Jag utförde arbetet vid KPA Pension, aktuarieavdelningen under hösten 2005.

Jag skulle vilja tacka mina handledare Tommy Kindberg vid KPA Pension och Lars Thuresson som hjälpte mig med utsökningarna, samt även aktuarierna Birgitta Alpe, Mats Bergsten, Mats Eriksson, Anna Jablonska, Tuomo Virolainen för den hjälp jag även haft av dem och för de diskussioner vi haft om försäkringar, matematik och allt annat möjligt och omöjligt under hösten.

Slutligen vill jag naturligtvis tacka min handledare vid Stockholms universitet, Joanna Tyrcha.

INNEHÅLL

1. INLEDNING.

1.1. Frågeställningar.

Syftet med detta arbete är dödlighetsundersökning på det försäkrade beståndet av förmånsbestämda pensioner under åren 1995 – 2004.

Frågor särskilt intressanta att belysa:
Vad är variansen i utfall mellan åren?
Är det skillnad i dödlighet mellan generationerna?
Är det skillnad i dödlighet beroende på pensionsnivå?
Kan något sägas om trenden?

Det är möjligt att beståndet är för litet för att dra säkra slutsatser, speciellt de första åren, men om metod- och teknikfrågorna är lösta kan allt säkrare slutsatser dras allteftersom nya år kan läggas till.

1.2. Kort om KPA.

KPA Pension, ursprungligen Kommunernas Pensionsanstalt, bildades 1922 för att hantera kommunernas pensionsåtaganden och var länge helägt av Kommunförbundet och Landstingsförbundet. Numera ägs KPA Pension till 60 % av Folksam och till 40 % av Sveriges Kommuner och Landsting. Sedan 1994 är KPA ett vanligt bolag som hanterar alla typer av livförsäkringar. Vem som helst, både arbetsgivare och privatpersoner, kan teckna försäkring, men fortfarande är kommuner och landsting de största kunderna. KPA hanterar pensioner åt drygt en miljon människor som är eller har varit anställda inom kommun eller landsting.

KPA är miljöcertifierat enligt ISO14001 1998 som första europeiska bolag inom finanssektorn.

Förutom försäkringar erbjuder KPA pensionsadministration, konsulttjänster, utbildning, pensionsvalsadministration och kapitalförvaltning.

354 personer arbetar på KPA. KPA blev utsett till Årets Livbolag 2003, 2004 och 2005 av tidningen Affärsvärlden.

1.3. Ingående variabler.

De ingående variablerna för analysen är

Variabel	**Värde**	
Födelseår		
Kön	1	Man
	0	Kvinna

Status	1	Överlevde kalenderåret
	0	Överlevde inte kalenderåret
Försäkringstyp	1	Livsfallsförsäkring
	0	Dödsfallsförsäkring
Ekonomisk klass	1	Risksumma överstigande 20 % över medel för födelseår och försäkringstyp.
	2	Risksumma överstigande 10 % under medel, men inte 20 % över medel för födelseår och försäkringstyp.
	3	Risksumma ej överstigande 10 % under medel för födelseår och försäkringstyp.
Årsrisksumma		total årsrisksumma för alla individer i cellen, d.v.s. de med likadana värden för alla ingående variabler. För dödsfallsförsäkring: tabellerad efter schablon.
Frigjord risksumma		För dödsfallsförsäkring: Beräknad på verkligt antal efterlevande.
År		Kalenderår
Antal		Antal personer i cellen
Duration		Duration i år över hela cellen avrundat till närmaste tolftedel.
Kvadratrisksumma		Kvadraten på risksummorna summerat över varje cell.

1.4. Försäkringsbeståndet.

Jag har analyserat beståndet av förmånsbestämda pensioner under åren 1994 – 2004 enligt de tre kollektivavtalen, KFS, KPA-plan och PFA, beskrivna nedan. Försäkringarna startade 1994 och beståndet har vuxit hela tiden. För 1994 fanns det 6519 poster och 947751 poster fanns för år 2004.

KFS
Avtal för fast anställda inom kommunala bolag innehållande pension från 65 år (eller lägre enl. avtal), efterlevandepension till vuxen, barnpension.

KPA-plan
Pensionsavtal tecknade av arbetsgivare ej anslutna till arbetsgivarorganisation, pensionsålder 65 år för alla.

PFA

Avtal för anställda inom kommun och landsting,

Livsfallsförsäkringar

A1	Livsvarig pension fr.o.m. 65 års ålder
A3	Tidsbegränsad pension fr.o.m. någon ålder (olika för olika delar av populationen, ofta 60 eller 62) t.o.m. 65 års ålder

Alla försäkringstagare som har A3 har även A1.

Dödsfallsförsäkringar

B1	Barnpension
E2	Efterlevandepension, betalas ut under fem år, dock längst till 65 års ålder
E3	Efterlevandepension, livsvarig

1.5. Materialet.

Antalet poster med ekonomisk klass LÅG

	Kvinnor	Män
1994	3103	3082
1995	6857	7972
1996	11057	10778
1997	12621	12538
1998	15262	15624
1999	26814	58482
2000	171036	74491
2001	226471	99893
2002	246514	79981
2003	281353	97067
2004	295987	104209

Antalet poster med ekonomisk klass MEDEL

	Kvinnor	Män
1994	781	425
1995	1022	1278
1996	2680	2427
1997	7008	7135
1998	6859	7073
1999	4290	6782
2000	8092	6238
2001	32334	14193
2002	141337	43784
2003	177447	52048
2004	155231	42679

Antalet poster med ekonomisk klass HÖG

	Kvinnor	Män
1994	829	1523
1995	3838	4660
1996	5118	9986
1997	2718	10254
1998	3238	12444
1999	11113	23309
2000	21970	37540
2001	147610	75709
2002	94718	76326
2003	138511	94853
2004	143743	98301

Antalet poster med livsfallsförsäkring

	Kvinnor	Män
1994	1139	2063
1995	2577	5121
1996	5491	8923
1997	7582	11835
1998	10548	15364
1999	20805	43419
2000	75297	58419
2001	141800	80441
2002	159915	72265
2003	205224	89742
2004	183275	81727

Antalet poster med dödsfallsförsäkring

	Kvinnor	Män
1994	3574	2967
1995	9140	8789
1996	13364	14268
1997	14765	18092
1998	14811	19777
1999	21412	45154
2000	125801	59850
2001	264615	109354
2002	322654	127826
2003	392087	154226
2004	411686	163462

2. TEORI.

2.1. Återstående livslängd och överlevelsefunktionen

Låt den återstående livslängden, T_x, för en försäkrad individ med åldern x vara en kontinuerlig stokastisk variabel med fördelningsfunktionen

$$F_x(t) = P(T_x \le t), t \ge 0$$

med täthetsfunktion

$$f_x(t) = F_x'(t), t \ge 0 .$$

Speciellt: Om $x = 0$ skriver vi

$$F(t) = P(T \le t), t \ge 0$$

respektive

$$f(t) = F'(t), t \ge 0 .$$

Sannolikheten att en x-årig individ blir högst t år till är sålunda

$$F_x(t) = \int_0^t f(x+u)du$$

Sannolikheten att en nyfödd blir högst t år är

$$F(t) = \int_0^t f(u)du$$

Ofta studerar vi överlevelsefunktionen, $l_x(t)$. Med den menas sannolikheten att en x-årig individ lever i mer än ytterligare t år.

$$l_x(t) = 1 - F_x(t)$$

Om $x = 0$ skriver vi

$$l(t) = l_0(t)$$

Vidare gäller:

$$l_x(t) = 1 - F_x(t) = P(T_x > t) = P(T > x + t \mid T > x) = \frac{P(T > x + t)}{P(T > x)} = \frac{l(x+t)}{l(x)}.$$

Av detta följer:

$$F_x(t) = 1 - l_x(t) = 1 - \frac{l(x+t)}{l(x)} \quad \text{och} \quad \frac{\partial F_x(t)}{\partial t} = \frac{-l'(x+t)}{l(x)}.$$

Då täthetsfunktionen har formen

$$f_x(t) = \frac{\partial F(t)}{\partial t} = \frac{-l'(x+t)}{l(x)} \quad \text{där i fall } x = 0 \quad f(t) = \frac{\partial F(t)}{\partial t} = \frac{-l'(t)}{l(0)} = -l'(t).$$

Vad vi fått fram är täthetsfunktionen för återstående livslängd. Den förväntade återstående livslängden $E(T_x) = e_x$ är då:

$$E(T_x) = e_x = \int_0^\infty t \frac{l'(x+t)}{l(x)} = \int_0^\infty \frac{l(x+t)}{l(x)} dt$$

Speciellt:

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

2.2. Dödlighetsintensitet och ettårig dödsrisk

Låt

$$\mu_x = \frac{f(x)}{1 - F(x)}.$$

Sannolikheten att avlida inom ett åldersintervall (x,x+dx), givet att man lever vid ålder x och att dx är kort, är approximativt lika med $\mu_x \cdot dx$. Storheten μ_x kallas dödlighetsintensiteten. Vidare gäller

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

Högerledet är lika med minus derivatan av ln(l(x)). Alltså:

$$\mu_x = -\frac{\partial \ln(l(x))}{\partial x}$$

Vidare:

$$\int_0^x \mu_s ds = \int_0^x \frac{\partial \ln(-l(s))}{\partial s} ds = \left[-\ln(l(s))\right]_0^x = -\ln(l(x)) + \ln(l(0)) = -\ln(l(x))$$

Alltså:

$$l(x) = e^{\ln(l(x))} = e^{-\int_0^x \mu_s ds}$$

2.3. Antalsdödlighet

Den ettåriga dödsrisken definieras som sannolikheten att en individ som överlevt ”förra” kalenderåret och då uppnått (x-1) års ålder inte överlever ”detta” kalenderår. Den ettåriga dödrisken betecknas som $q(x)$. Individernas dödsålder, T, skattas sålunda till kalenderåret för döden minus kalenderåret för födseln.
Ett approximativt samband mellan q_x och μ_x är

$$\mu_x = \frac{q_x}{1 - \frac{q_x}{2}}.$$

Den ettåriga dödsrisken q_x skattas enligt

$$q_x = \frac{D_x}{N_x},$$

där D_x är antalet avlidna x-åringar och N_x antalet x-åringar ö.h.t. i försäkringsbeståndet.

Dödlighetsintensiteten μ_x skattas sålunda med

$$\hat{\mu}_x = \frac{D_x}{N_x\left(1 - \frac{D_x}{2N_x}\right)} = \frac{D_x}{N_x - \frac{D_x}{2}}$$

Ett alternativt sätt att skatta μ_x är ML-skattningen

$$\hat{\mu} = \frac{D_x}{d_x},$$

där d_x är den totala durationen för cellen. (Se kap. 2.4. för definition.)

De skattade värdena minsta-kvadrat-anpassas till Makehams formel:

$$\mu_x = \alpha + \beta \cdot e^{\gamma \cdot x}, \beta > 0, \alpha + \beta > 0, \gamma > 0$$

Bivillkoren är för att alla värden skall vara positiva och ökande med x.

Den trippel (α,β,γ) som minimerar

$$Q = \sum_{x=1}^{\infty} w_x (\hat{\mu}_x - \alpha - \beta \cdot e^{\gamma \cdot x})^2,$$

givet att bivillkoren uppfylls, väljs, där w_x är en vald viktning.

För att minska variansen väljs i allmänhet $w_x = \dfrac{N_x}{\hat{\mu}_x}$. I materialet är dock många skattningar av μ_x lika med noll, och därför väljs istället $w_x = N_x$. Om γ betraktas som fixt skattas α och β genom följande formler:

$$\hat{\beta} = \frac{\left(\sum_x w_x\right)\cdot\left(\sum_x w_x \cdot e^{\gamma \cdot x} \cdot \hat{\mu}_x\right) - \left(\sum_x w_x \cdot e^{\gamma \cdot x}\right)\cdot\left(\sum_x w_x \cdot \hat{\mu}_x\right)}{\left(\sum_x w_x\right)\cdot\left(\sum_x w_x \cdot e^{2\cdot\gamma\cdot x}\right) - \left(\sum_x w_x \cdot e^{\gamma \cdot x}\right)^2}$$

$$\hat{\alpha} = \frac{\left(\sum_x w_x \cdot \hat{\mu}_x\right) - \hat{\beta}\cdot\left(\sum_x w_x \cdot e^{\gamma \cdot x}\right)}{\left(\sum_x w_x\right)}$$

Med datorns hjälp beräknas $\hat{\alpha}, \hat{\beta}$ och Q för ett större antal γ. Det lägsta Q-värdet där bivillkoren uppfylls levererar parameterskattningarna.

2.4. Duration

Begreppet duration används på två olika sätt:

1) Tiden som gått sedan försäkringen tecknades. Kan alltså vara åtskilliga år för en enskild kund.

2) Under hur långt tidsintervall försäkringen varit i kraft under studerat kalenderår uttryckt i enheten år. En kund som var med hela året får värdet 1 år och kunder som varit med del av året , d.v.s. inträtt i beståndet och/eller utträtt/dött under året bidrar får ett värde på durationen mellan 0 och 1 år. Durationen anges ofta som en summa över alla individer i en viss cell. Det här är det sätt på vilket begreppet huvudsakligen används i den här rapporten.

2.5. Grunder.

Vid beräkningen av premier använder KPA följande värden för parametrarna i dödlighetsintensiteten, $\mu_x = \alpha + \beta \cdot e^{\gamma \cdot x}$:

		α	β	γ
Livsfallsförsäkring	Män	0	0,0000154	0,103
	Kvinnor	0	0,0000089	0,103
Dödsfallsförsäkring	Män	0,0001	1,4·0,0000154	0,103
	Kvinnor	0,0001	1,4·0,0000089	0,103

2.6. Åldersförskjutning.

Ett sätt att avgöra skillnad i dödlighet mellan två delpopulationer är att studera åldersförskjutning mellan två Makeham-kurvor, d.v.s. ett viktat medelvärde för skillnaden i ålder för samma dödsintensitet. Låt $\mu_1 = \alpha_1 + \beta_1 \cdot e^{\gamma_1 x_{1i}}, \mu_2 = \alpha \;\; + \beta_2 \cdot e^{\gamma_2 x_{2j}}$ vara Makehamskattningar för två delpopulationer, för olika åldrar i = 1,… och j = 1,… . För ett givet x_{1i} -värde är det värde på x_{2j} som ger samma μ-värde lika med

$$x_{2j} = \frac{\ln\left(\frac{\alpha_1 - \alpha_2 + \beta_1 \cdot e^{\gamma_1 x_{1i}}}{\beta_2}\right)}{\gamma_2}$$. Differensen mellan x-värdena är $\frac{\ln\left(\frac{\alpha_1 - \alpha_2 + \beta_1 \cdot e^{\gamma_1 x_{1i}}}{\beta_2}\right)}{\gamma_2} - x_{1i}$.

Om man viktar differenserna med riskpremien vid åldern x_{1i}, $r(x_{1i}) \cdot \left(\alpha + \beta e^{\gamma x_{1i}}\right)$ blir

medelvärdet lika med
$$f_1 = \frac{\sum_{i=1}^{\infty} \left(\alpha_1 + \beta_1 e^{\gamma_1 x_{1i}}\right) r(x_{1i}) \left(\frac{\ln\left(\frac{\alpha_1 - \alpha_2 + \beta_1 \cdot e^{\gamma_1 x_{1i}}}{\beta_2}\right)}{\gamma_2} - x_{1i} \right)}{\sum_{i=1}^{\infty} \left(\alpha_1 + \beta_1 e^{\gamma_1 x_{1i}}\right) r(x_{1i})}.$$

Utgår man istället från x_{2j} blir det viktade medelvärdet

$$f_2 = -\frac{\sum_{j=1}^{\infty} \left(\alpha_2 + \beta_2 e^{\gamma_2 x_{2j}}\right) r(x_{2j}) \left(\frac{\ln\left(\frac{\alpha_2 - \alpha_1 + \beta_2 \cdot e^{\gamma_2 x_{2j}}}{\beta_1}\right)}{\gamma_1} - x_{2j} \right)}{\sum_{j=1}^{\infty} \left(\alpha_2 + \beta_2 e^{\gamma_2 x_{2j}}\right) r(x_{2j})}.$$

Minustecknet här är till för att tecknet för åldersförskjutningen skall bli detsamma i båda fallen. Åldersförskjutningen f definieras här som $f = \frac{f_1 + f_2}{2}$.

2.7. Ekonomisk dödlighet.

För att analysera hur dödligheten påverkar ett försäkringsbestånd ekonomiskt är man inte intresserad av att antalsdödligheten utan viktar varje individs påverkan på μ med individens risksumma. Om en person med stor försäkring avlider eller ej påverkar värdet på μ mer är om en person med liten försäkring avlider eller ej. Då jag beräknar dödligheten använder jag alltså

$$\hat{q}_x = \frac{\sum_{\text{avlidna}} R(t)}{\sum_{alla} R(t) \cdot dt}$$

där täljaren är summan av de avlidna x-åringarnas frigjorda risksumma (d.v.s. den risksumma som gällde vid, eller snarare precis före, dödsfallet) och nämnaren är summan av x-åringarnas årsrisksumma.

2.8. Risksumma och riskpremie.

Låt A(t) = kapitalvärdet av försäkringsgivarens framtida förpliktelser vid tidpunkten t. Med kapitalvärdet menas de förväntade förpliktelserna diskonterade tillbaka till tidpunkt t. Låt vidare B(t)=kapitalvärdet av försäkringstagarens framtida förpliktelser vid tidpunkten t. Låt V(t)=A(t)-B(t). V(t) kallas värdefunktionen. $V(t) \geq 0$ då durationen $t > 0$, d.v.s. efter tecknandet. Låt nu S(t) vara det belopp som betalas ut vid dödsfall. Sätt

R(t) = S(t) – V(t)

R(t) kallas försäkringens risksumma och är ett mått på försäkringsavtalets storlek. Om $R(t) > 0$ kallas försäkringen dödsfallsförsäkring, annars livsfallsförsäkring. Dödsfallsförsäkringar ger utbetalning under förutsättning att någon individ är avliden vid någon tidpunkt eller några tidpunkter. Livsfallsförsäkringar ger utbetalning under förutsättning att någon individ lever vid någon tidpunkt eller några tidpunkter. Med årsrisksumma menas R(t)/12 uträknad för varje månad och sedan hopsummerad. Årsrisksumman kan alltså ses som ett genomsnitt av risksumman över året. För en individ som träder in i försäkringsbeståndet under kalenderåret är R(t)/12 = 0 innan avtalet sluts. Med frigjord risksumma avses vanligen den risksumma som förelåg vid tidpunkten för dödsfallet, eller snarare strax före.
Med riskpremie i tidsintervallet (t,t+dt) menas

$$P_r = \mu_{x+t} \cdot \big(S(t) - V(t)\big) \cdot dt = \mu_{x+t} \cdot R(t) \cdot dt\,,$$

där x är åldern då t=0, d.v.s. då avtalet ingås och t durationen. dt förutsätts här vara tämligen litet – vanligtvis ett år eller en månad. I alla tillämpningar i denna rapport är dt = 1år.

2.9. Relativ ekonomisk dödlighet.

Antalsdödlighetens dödlighetsintensitet μ_x skattas som vi tidigare sagt med ML-skattningen

$$\hat{\mu} = \frac{D_x}{d_x},$$

där d_x är den totala durationen för cellen och D_x antalet döda.
Risksumman (innan dödsfall) för efterlevandeförsäkring (dödsfallsförsäkring) beräknar företaget inte efter individernas faktiska familjeomständigheter utan efter schablon. En 39-årig man kanske antas ha 0,6 fru och 1,5 barn o.s.v . Den frigjorda risksumman däremot beräknas efter den avlidnas faktiska förhållanden och ger viss utbetalning per efterlevande. En person med många efterlevande får mycket i utbetalning och en person utan efterlevande får ingen utbetalning. Den frigjorda risksumman kan då variera och vara både positiv och negativ (och inte som i läroböckerna lika med risksumman strax innan dödsfallet och alltid positiv för dödsfallsförsäkring). En ML-skattning av antalsdödligheten är att man tar antalet döda delat med durationen för en viss åldersgrupp. Om man vikter denna med risksumman får man ekonomisk dödlighet lika med total (de facto) frigjord risksumma delat med total (tabellerad) årsrisksumma. Tar man denna dödlighetsintensitet och delar med bolagets Makeham-antaganden för dödlighetsförsäkringar så får man en kvot. Är den större än 1 är utfallet ekonomiskt värre för företaget än väntat. Mindre än 1 bättre än väntat.

2.10. Generationsdödlighet.

Generationsdödligheten beskriver dödligheten för en generation (personer födda under samma 5- eller 10-årsintervall). För att kunna fullfölja en sådan analys krävs att man studerar värden från ca 110 kalenderår.
Mitt materiel består av människor födda 1904 och senare och vad som hände dem 1994 – 2004. Av uppenbara skäl kan jag endast studera ett begränsat åldersintervall för varje generation. För 1910-talisterna t.ex. intervallet 75 år – 94 år. Urvalet är som störst mitt i åldersintervallet och minskar i storlek mot ändpunkterna. Generationerna överlappar inte varandra åldersmässigt under den här perioden, så en generationsanalys är inte relevant. När flera år läggs till materialet kan variabeln generation behandlas i en analys på samma sätt som variabeln trend görs.

2.11. Logistisk regression.

Multipel logistisk regression definieras av n_i $(i = 1,\ldots k)$ oberoende stokastiska variabler $n_i \in Bin(N_i, p(x_i))$ och

$$\ln\left(\frac{p(x_i)}{1 - p(x_i)}\right) = \hat{a}_0 + \hat{a}_1 x_{1i} + \hat{a}_2 x_{2i} + \ldots + \hat{a}_r x_{ri}$$

där $x_{i,}$ $i = 1,\cdots k$, är de ingående *r*-dimensionella oberoende variablerna. Detta är ekvivalent med

$$p(x_i) = \frac{e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_r x_{ri}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_r x_{ri}}}$$

β_j är logoddskvoten för att individerna skall avlida om x_j ökar från 0 till 1 och de övriga *x*-variablerna hålls oförändrade.

Vektorn $\boldsymbol{\beta}$ av β_j-parametrar skattas medelst Newton-Raphsons metod. Ett iterationssteg ser ut som

$$\boldsymbol{\beta}^{[t+1]} = \boldsymbol{\beta}^{[t]} + [\mathbf{I}(\boldsymbol{\beta}^{[t]})]^{-1} (\mathbf{X'n} - \boldsymbol{\mu}'\mathbf{n})$$

Här står **X** för designmatrisen, **n** för vektorn av antalet döda och $\boldsymbol{\mu}$ är väntevärdesvektorn enligt modellen. $\boldsymbol{\mu} = \mathbf{N'p}$, där **N** är vektorn av antalet personer i cellerna. Vi sätter $\boldsymbol{\beta}^{[0]} = \mathbf{0}$. Matrisen **I** är informationsmatrisen (minus andraderivatorna av likelihood-funktionen med avseende på β_j-variablerna) som i det här fallet är $\mathbf{I} = (\mathbf{X'DX})^{-1}$, där D är en diagonalmatris med $N_i \cdot p(\mathbf{x}_i)(1-p(\mathbf{x}_i))$ Beräkningarna utförs i Mathcad.

Skattningarna är approximativt normalfördelade enligt

$$\hat{\boldsymbol{\beta}} \approx N\left(\boldsymbol{\beta}, (X'DX)^{-1}\right)$$

(Se Ohlsson)

Modellval

Vi analyserar tre grupper av modeller, en där åldrarna delas in i ettårsintervall, ett med femårsintervall och en med tioårsintervall.
I alla modeller inför vi en egen dummy-variabel för varje representerat åldersintervall, i, där $x_i = 1$ för en individer i åldersintervall i och $x_1 = 0$ för övriga. Eftersom varje individ tillhör exakt ett åldersintervall, varken mer eller mindre, kan man utesluta β_0-parametern. Detta för att undvika överparametrisering. Förutom åldersvariablerna finns ingår en variabel för försäkringstyp, en för kön, en för trend (d.v.s. vi jämför åren 1994 – 1999 med 2000 – 2004), samt två dummy-variabler för de tre nivåerna på risksumman. I modellerna kan x-värdena alltså bara anta värdena 0 eller 1. När ytterligare kalenderår läggs till och generationerna mera överlappar varandra åldersmässigt kan man ersätta trend med generation (födelseårtionde).

Diskrepans

Diskrepansen vid logistisk regression definieras som

$$D = 2\sum_i n_i \cdot \ln\left(\frac{p^{(1)}(x_i)}{p^{(0)}(x_I)}\right) + 2\sum_i (N_i - n_i) \cdot \ln\left(\frac{1 - p^{(1)}(x_i)}{1 - p^{(0)}(x_I)}\right) ,$$

där $p^{(1)}(x_i)$ och $p^{(0)}(x_i)$ är ML-skattningar i hypotesmodellen resp. grundmodellen. Om antalet observationer är stort är så fäller vanligen $D \approx \chi^2(fg)$, där fg är lika med differensen mellan antalet parametrar i de olika modellerna.

Vidare:
I binomialfördelningen gäller att

$$P(Y = n) = \binom{N}{n} p^n (1-p)^{N-n} = \frac{\Gamma(N+1)}{\Gamma(n+1) \cdot \Gamma(N-n+1)} p^n (1-p)^{N-n} ,$$

där Γ är Eulers gammafunktion. För att få modellen att passa våra behov generaliserar vi ovanstående till att gälla även då N_i ej är heltal. För små värden på p, vilket vi har här, är Y approximativt Po(Np)-fördelat (se Gut) och vidare kan dödsfallen ses som en poissonprocess, med tidshorisont maximalt lika med 1 år. Alltså är modellen logistisk regression även tillämplig för N som ej är heltal. **N** i den logistiska modellen blir alltså vektorn av durationer och **n** vektorn av antalet döda.

Modeller

För alla tre grupper av modeller testas följande hypoteser. Observera att ålder och kön ingår i alla modeller.

	Ålder	Kön	Försäkrings-typ	Ekonomisk klass	Trend
H1	Ja	Ja	Nej	Nej	Nej
H2	Ja	Ja	Ja	Nej	Nej
H3	Ja	Ja	Nej	Ja	Nej
H4	Ja	Ja	Nej	Nej	Ja
H5	Ja	Ja	Ja	Nej	Ja
H6	Ja	Ja	Ja	Ja	Nej
H7	Ja	Ja	Nej	Ja	Ja
H8	Ja	Ja	Ja	Ja	Ja

2.12. Skadeprocessen

Låt antalet dödsfall, N, under ett år vara en Poisson-fördelad stokastisk variabel med väntevärde och varians n. Låt vidare varje frigjord risksumma, Z_i, tillhöra samma fördelning, och vara sinsemellan oberoende. Sätt vidare $E(Z_i) = m$ och $Var(Z_i) = s^2$. Totala frigjorda risksumman X har då väntevärdet E(X) = nm och varians Var(X) = E Var (X|N) + Var E(X|N) = $E(Ns^2)$ + Var Nm = $ns^2 + nm^2 = n(s^2 + m^2)$. Då gäller

$$F_X(x) \xrightarrow[n\to\infty]{} N(nm, n(s^2 + m^2)) = N(nm, nE(X^2)) \qquad \text{(Se Blom och Bowers)}$$

Låt varje enskilt dödsfall, Z_i, i inträffa med sannolikheten $\lambda_i = \mu_i \cdot d_i$, där μ_i är dödlighetsintensiteten och d_i durationen under kalenderåret. Eftersom varje $\mu_i \cdot d_i$ är så litet och vi bara studerar ett kalenderår i taget, kan man skatta varje skada som vore det en Poisson-process med intensiteten $\mu_i \cdot d_i$. Väntevärdet n kan då skattas enligt $\hat{n} = \sum_i \mu_i \cdot d_i$. Vidare skattas m med $\hat{m} = \frac{\sum_i \mu_i \cdot d_i \cdot R_i}{\sum_i \mu_i \cdot d_i}$, där R_i är risksummorna. $d_i \cdot R_i$ blir då individens årsrisksumma. Väntevärdet E(X) = nm skattas då $\left(E(\hat{X})\right) = \sum_i \mu_i \cdot d_i \cdot R_i$ Vidare kan variansen för totalt skadebelopp skattas enligt

$$\left(Var\hat{(}X)\right) = \left(\sum_i \mu_i \cdot d_i\right) \frac{\sum_i \mu_i \cdot d_i \cdot R_i^2}{\sum_i \mu_i \cdot d_i} = \sum_i \mu_i \cdot d_i \cdot R_i^2 \approx \sum_j \mu_j \left(\frac{d_j}{a_j} \sum_{k=1}^{a_j} R_k^2 \right)$$, där a_j är antalet personer i cell j och d_j hela cellens sammanlagda duration.

3. RESULTAT.

3.1. Makehamskattningar

Ekonomisk dödlighet

			α	β	γ
Män	Livsfallsförsäkring	Hög risksumma	-1,055e-5	1,285e-5	0,1086
		Medel-risksumma	-2,026e-5	5,548e-5	0,1021
		Låg risksumma	-7,609e-7	5,483e-6	0,1282
		Alla risksummor	7.7500e-6	1,1689e-5	0,1123
Kvinnor	Livsfallsförsäkring	Hög risksumma	8,479e-5	1,261e-5	0,1025
		Medel-risksumma	0,0001503	0,0001642	0,0796
		Låg risksumma	0,001660	5,672e-6	0,1227
		Alla risksummor	-1,1498e-5	1,1626e-5	0,1077

Antalsdödlighet

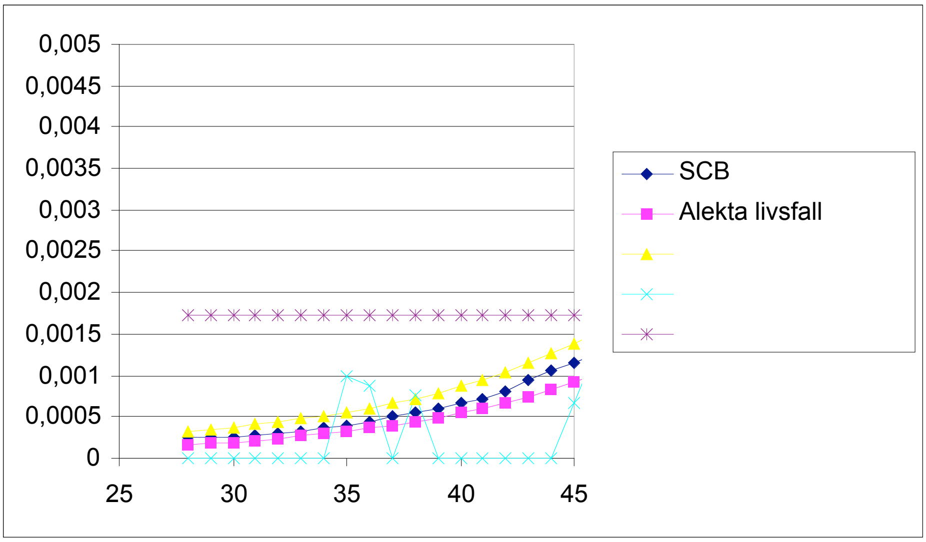
			α	β	γ
Män	Livsfallsförsäkring	Hög risksumma	0,0006977	9,719e-7	0,1262
		Mede-lrisksumma	0,0004570	2,111e-6	0,1264
		Låg risksumma	3,173e-6	3,190e-6	0,1305
		Alla risksummor	5,404e-6	6,083e-6	0,1151
	Dödsfallsförsäkring	Hög risksumma	2,910e-5	1,516e-8	0,1912
		Mede-lrisksumma	0,0003048	3,323e-7	0,1469
		Låg	-6,031e-5	6,712e-5	0,0884

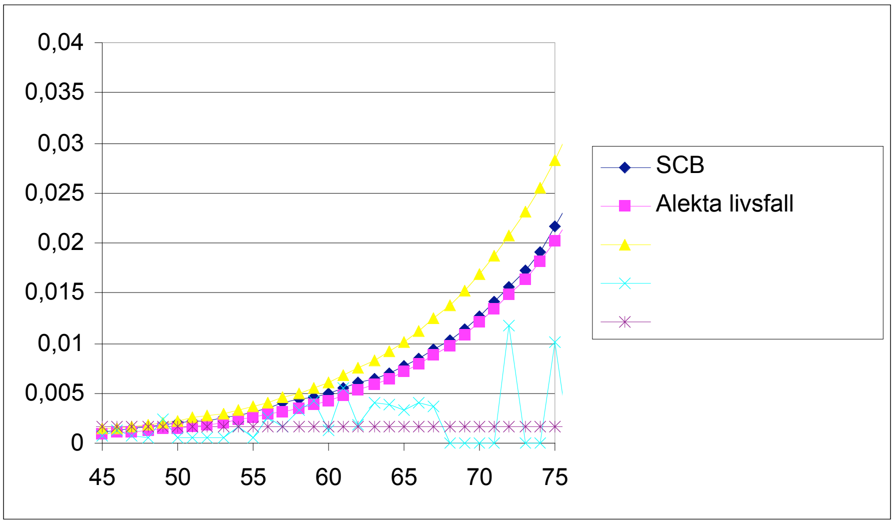
		risksumma			
		Alla risksummor	-8,939e-7	4,592e-6	0,1193
Kvinnor	Livsfallsförsäkring	Hög risksumma	9,586e-8	5,213e-8	0,1400
		Medel-risksumma	-1,748e-5	1,786e-5	0,0900
		Låg risksumma	0,0001496	6,144e-7	0,1447
		Alla risksummor	5,480e-5	2,474e-6	0,1206
	Dödsfallsförsäkring	Hög risksumma	0,0001393	3,775e-12	0,2752
		Medel-risksumma	2,344e-5	9,842e-8	0,1473
		Låg risksumma	6,848e-5	1,288e-5	0,1027
		Alla risksummor	2,510e-5	1,343e-6	0,1273

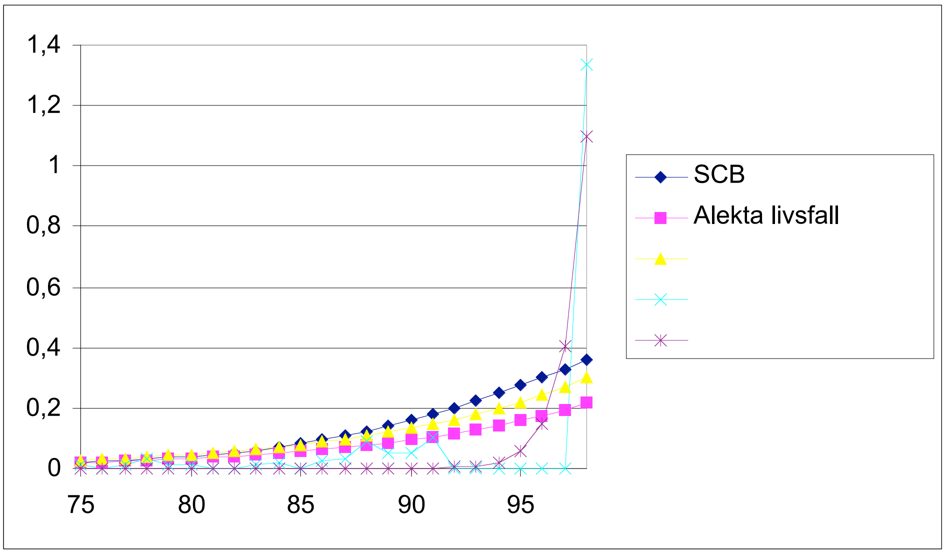
3.2. Grafer över dödlighetsintensiteter.

3.2.1. Dödlighetsintensitet, antalsdödlighet.

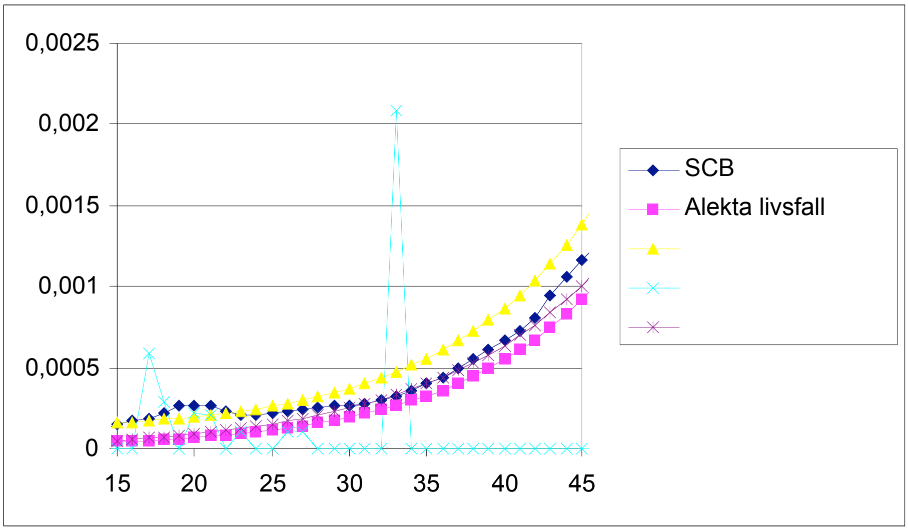
Kvinnor, livsfallsförsäkring, hög ekonomisk nivå

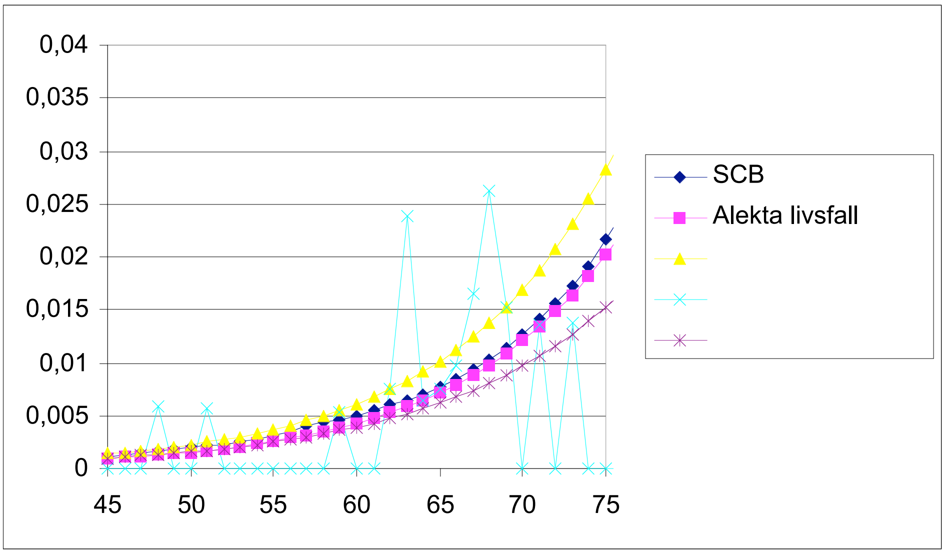

0,005
0,0045
0,004
0,0035
0,003
0,0025
0,002
0,0015
0,001
0,0005
0
25
30
35
40
45
SCB
Alekta livsfall

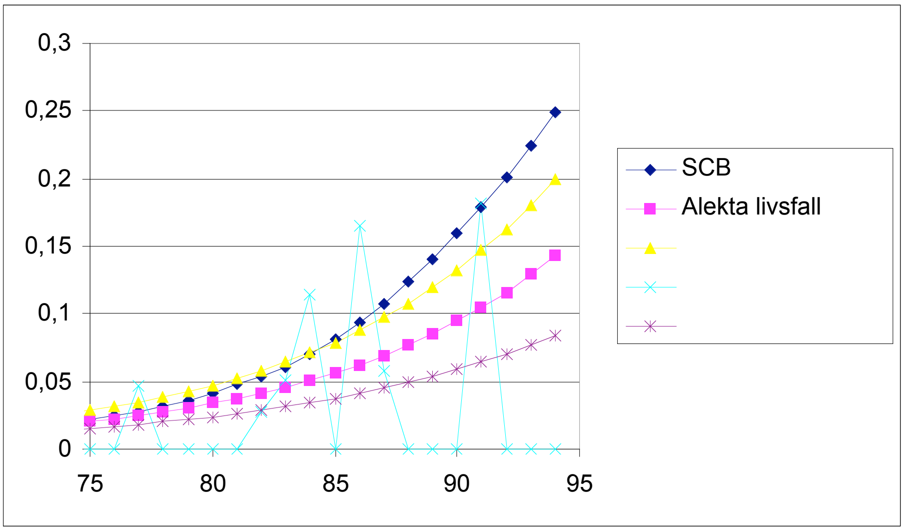

0,04
0,035
0,03
0,025
0,02
0,015
0,01
0,005
0
45
50
55
60
65
70
75
SCB
Alekta livsfall


1,4
1,2
1
0,8
0,6
0,4
0,2
0
75
80
85
90
95
SCB
Alekta livsfall

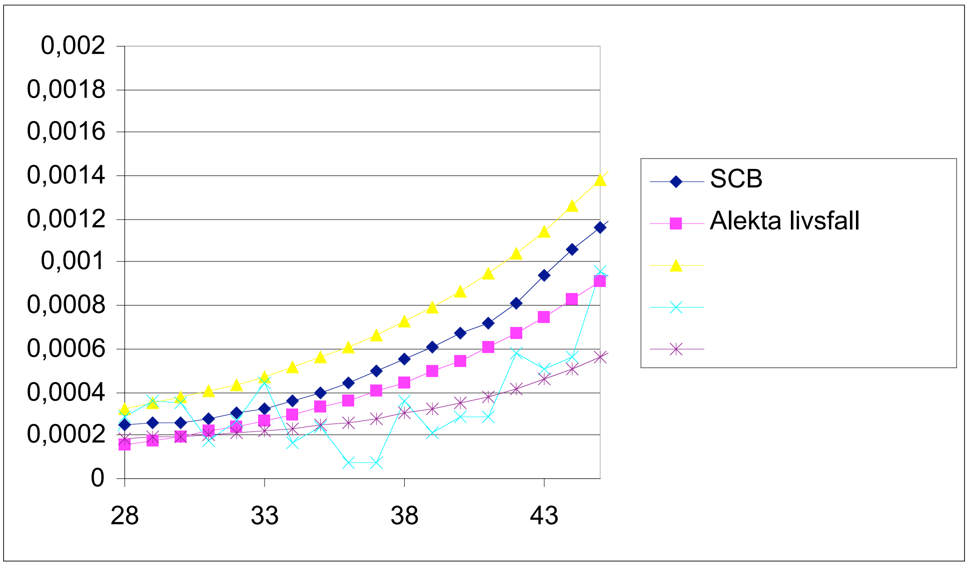
Kvinnor, livsfallsförsäkring, ekonomisk medelnivå

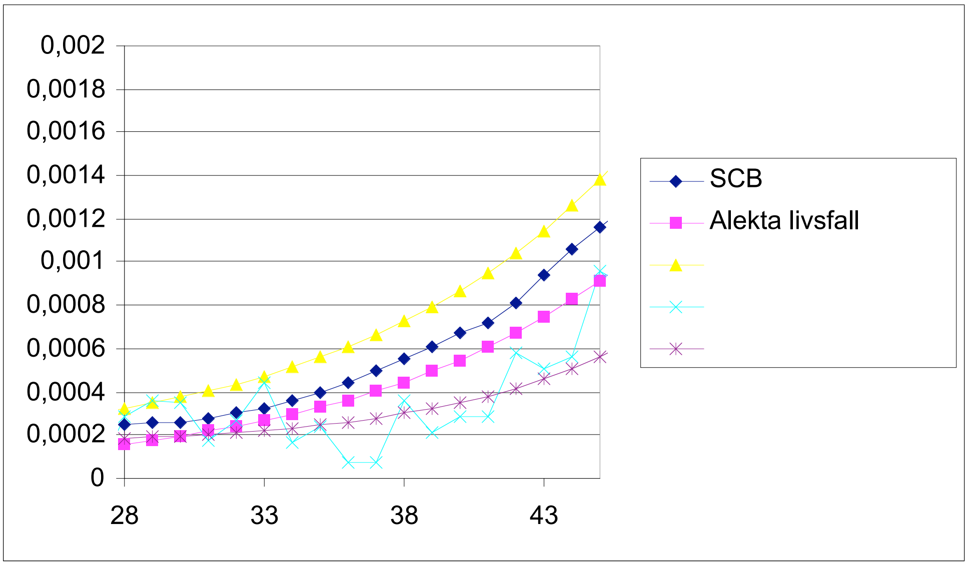


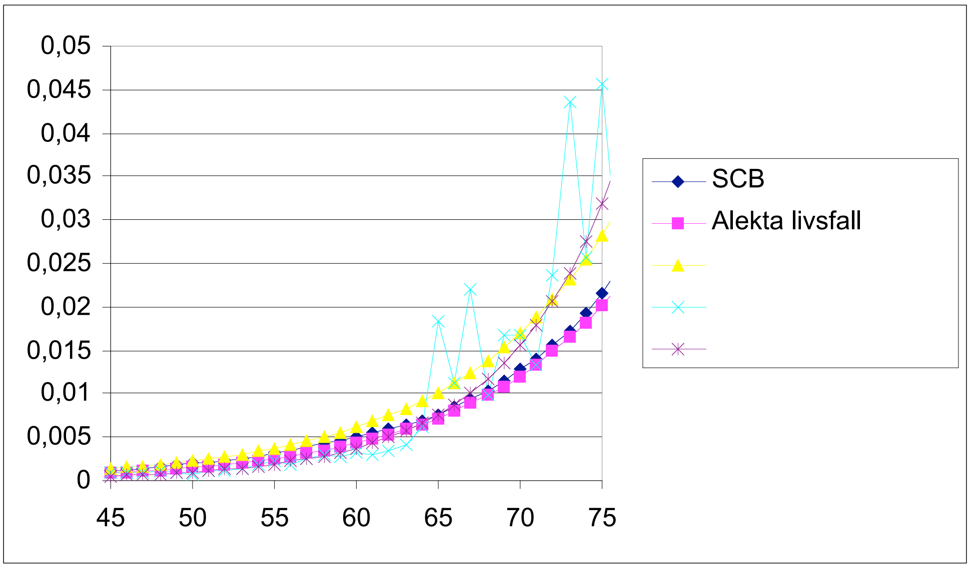


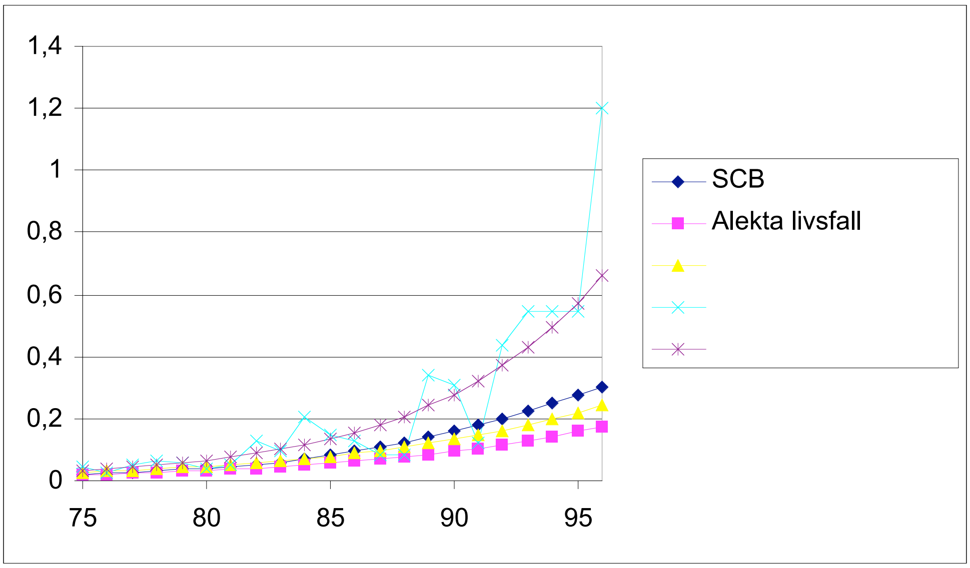


Kvinnor, livsfallsförsäkring, låg ekonomisk nivå.

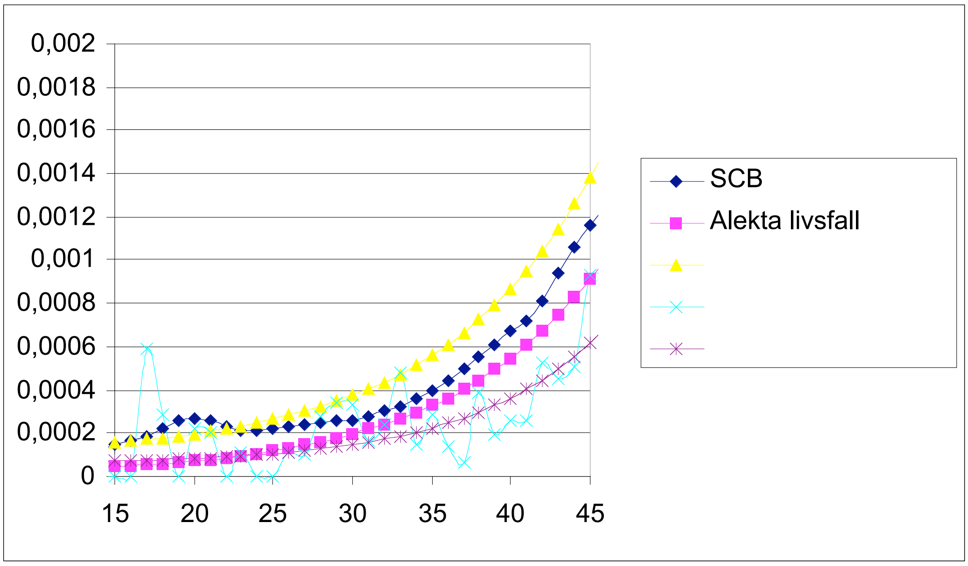


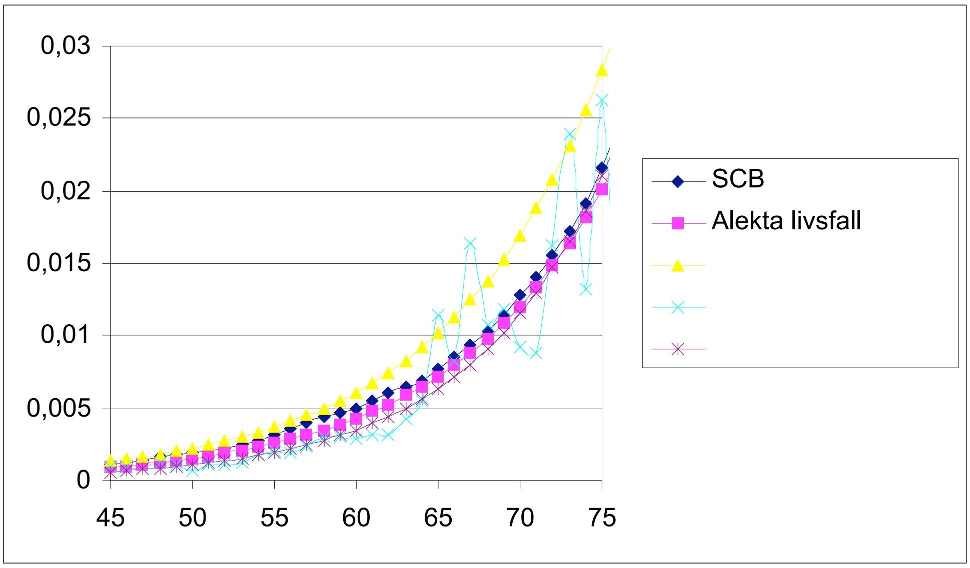

0,002
0,0018
0,0016
0,0014
0,0012
0,001
0,0008
0,0006
0,0004
0,0002
0
28
33
38
43
SCB
Alekta livsfall

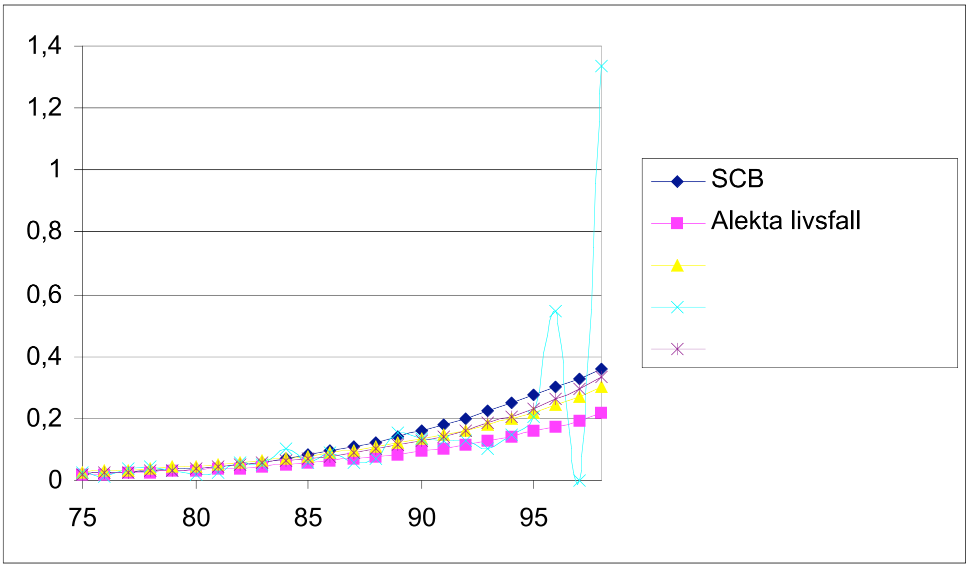

0,05
0,045
0,04
0,035
0,03
0,025
0,02
0,015
0,01
0,005
0
45
50
55
60
65
70
75
SCB
Alekta livsfall


1,4
1,2
1
0,8
0,6
0,4
0,2
0
75
80
85
90
95
SCB
Alekta livsfall

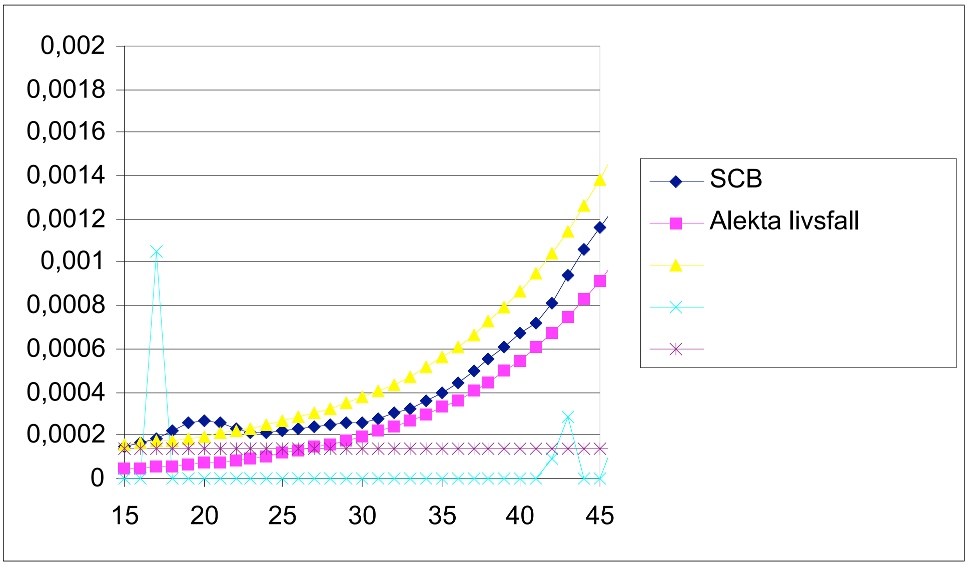
Kvinnor, livsfallsförsäkring, alla

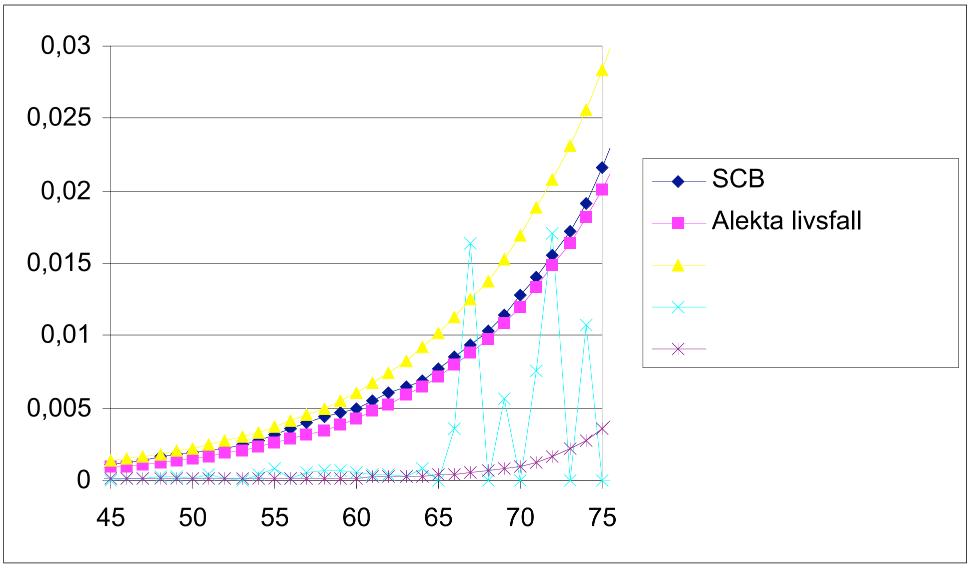

0,002
0,0018
0,0016
0,0014
0,0012
0,001
0,0008
0,0006
0,0004
0,0002
0
15
20
25
30
35
40
45
SCB
Alekta livsfall

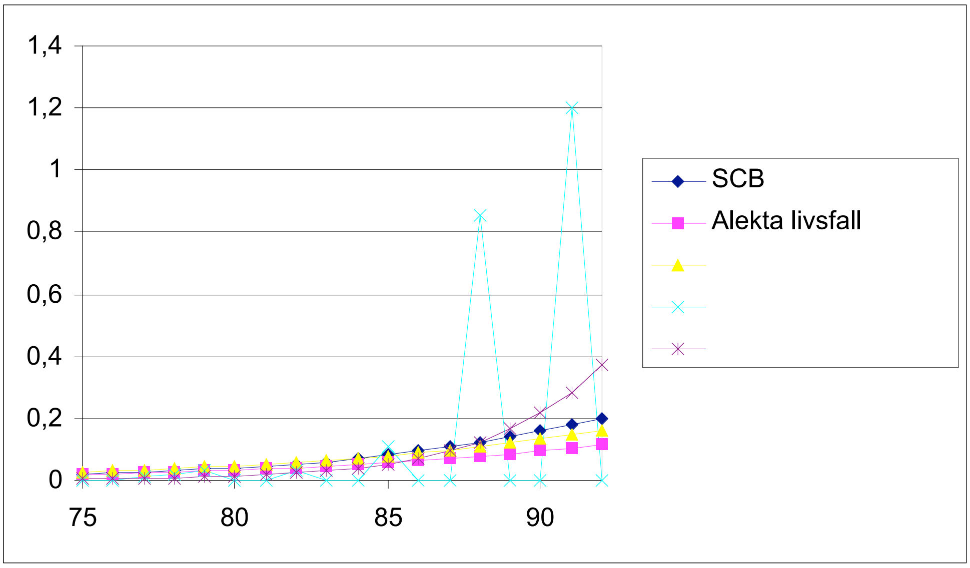

0,03
0,025
0,02
0,015
0,01
0,005
0
45
50
55
60
65
70
75
SCB
Alekta livsfall


1,4
1,2
1
0,8
0,6
0,4
0,2
0
75
80
85
90
95
SCB
Alekta livsfall

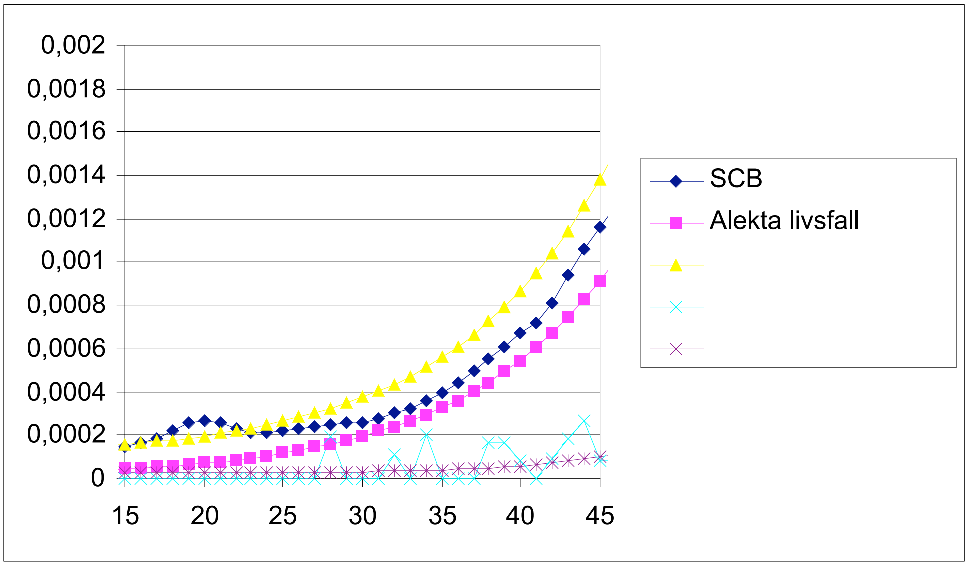
Kvinnor, dödsfallsförsäkring, hög ekonomisk nivå.

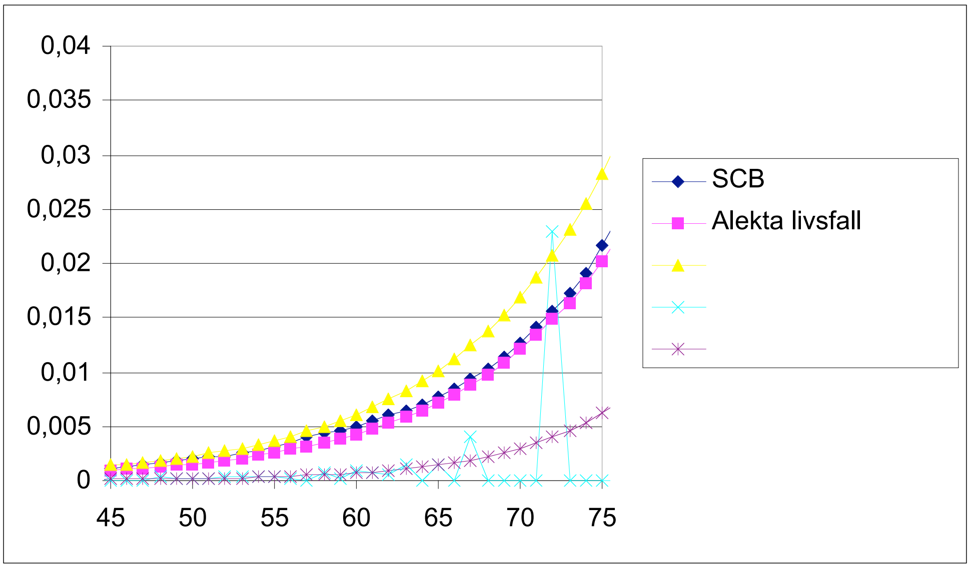


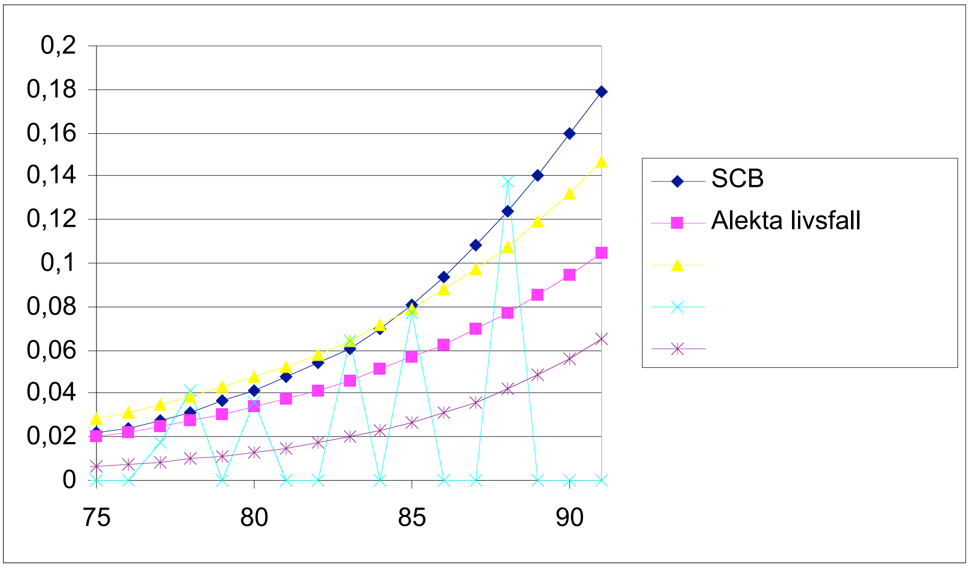




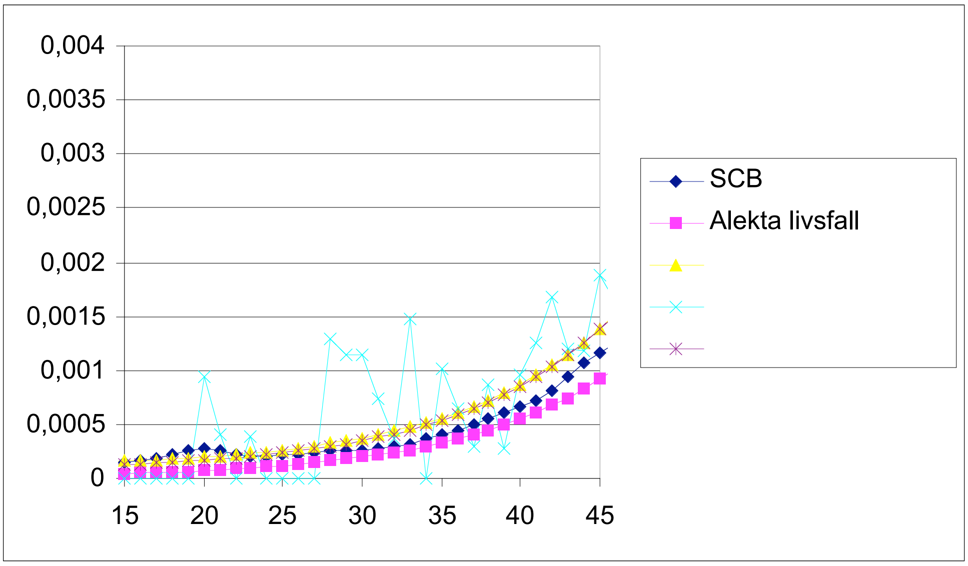
Kvinnor, dödsfallsförsäkring, ekonomisk medelnivå.



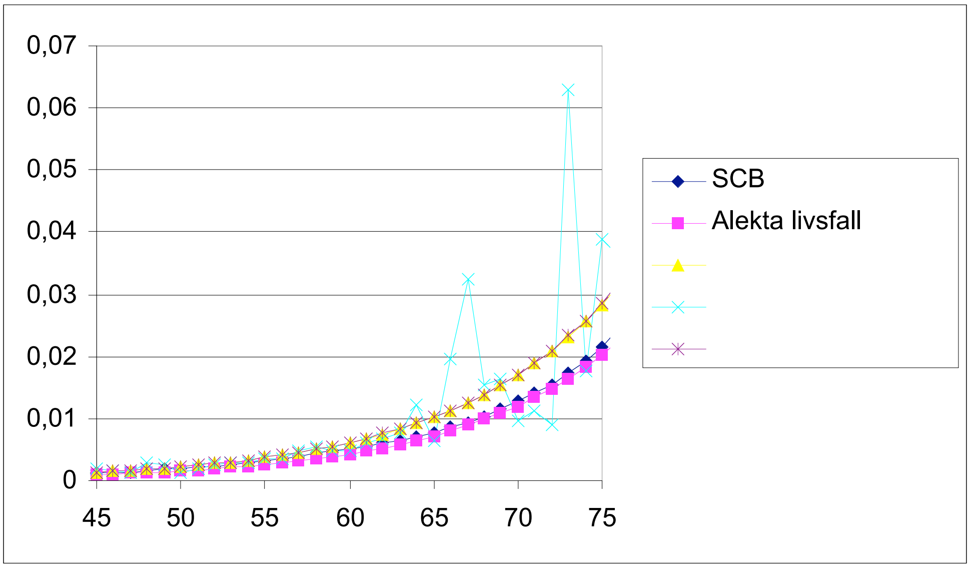




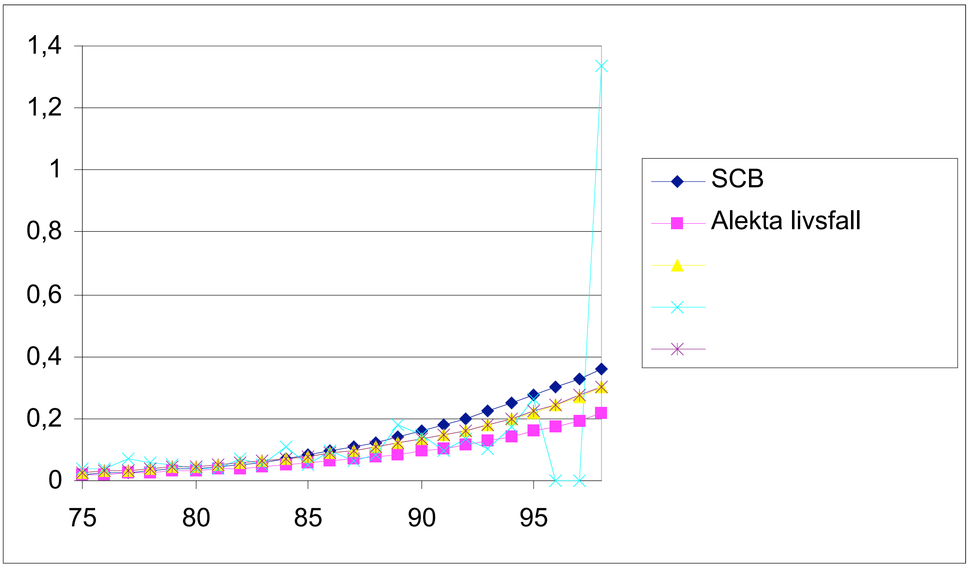
Kvinnor, dödsfallsförsäkring, låg ekonomisk nivå.



0,004
0,0035
0,003
0,0025
0,002
0,0015
0,001
0,0005
0
15
20
25
30
35
40
45
SCB
Alekta livsfall

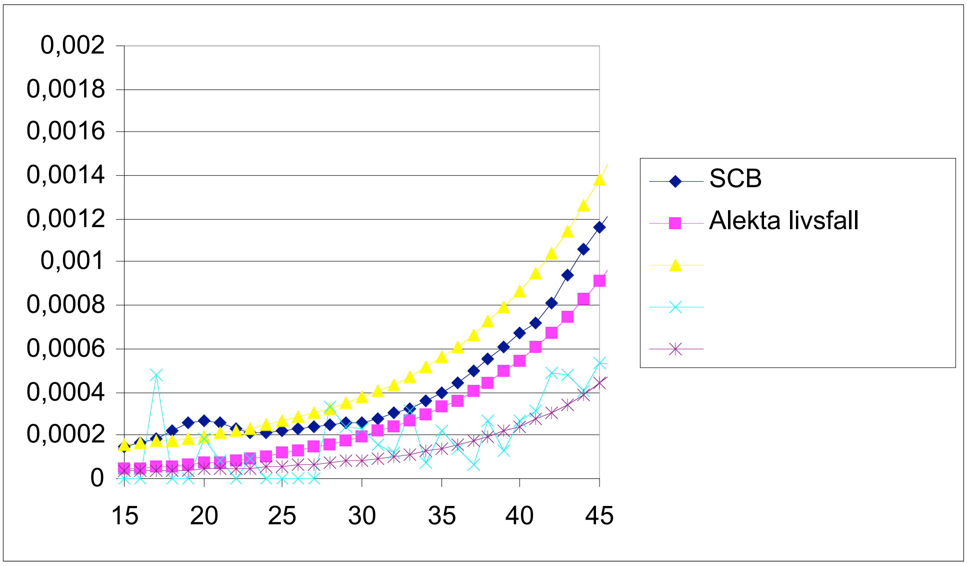


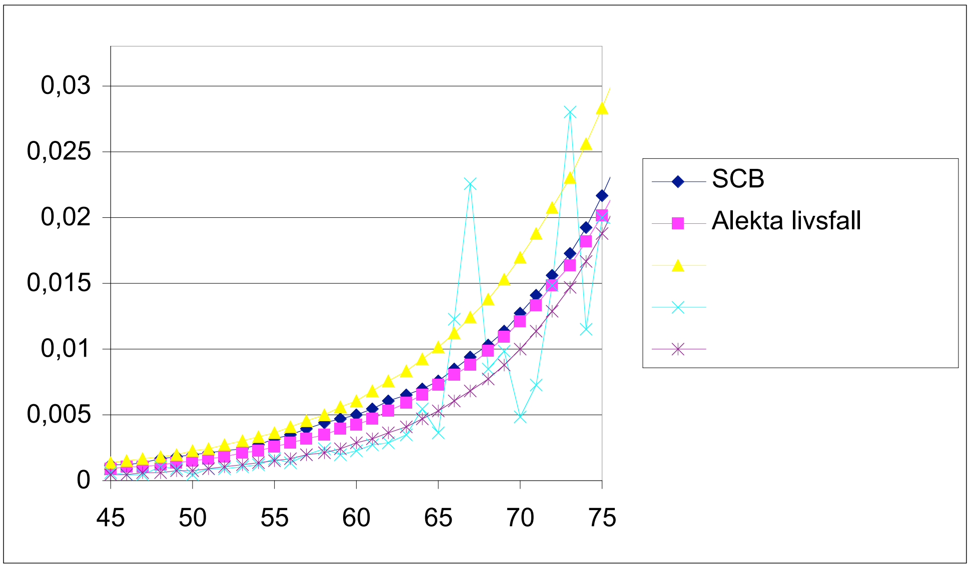
0,07
0,06
0,05
0,04
0,03
0,02
0,01
0
45
50
55
60
65
70
75
SCB
Alekta livsfall

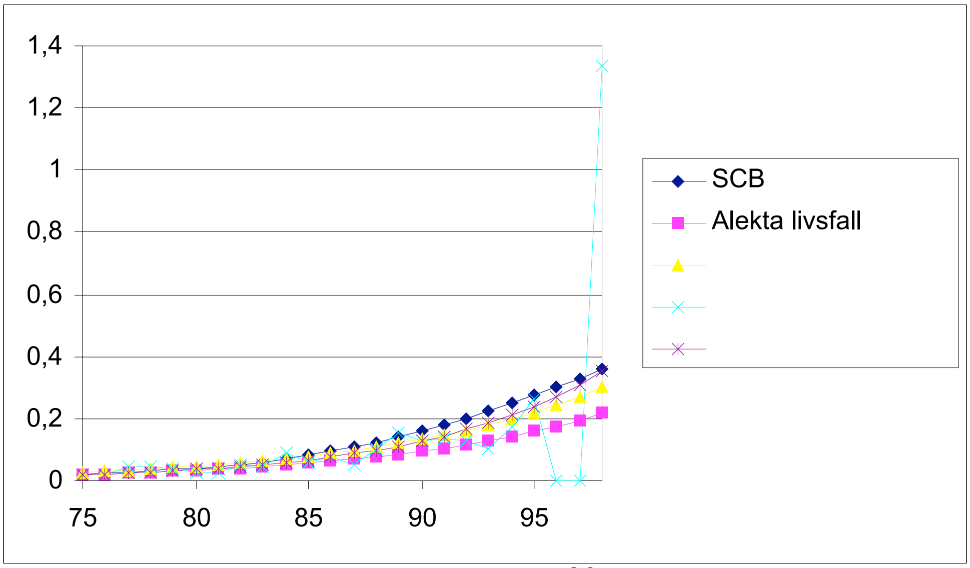


1,4
1,2
1
0,8
0,6
0,4
0,2
0
75
80
85
90
95
SCB
Alekta livsfall

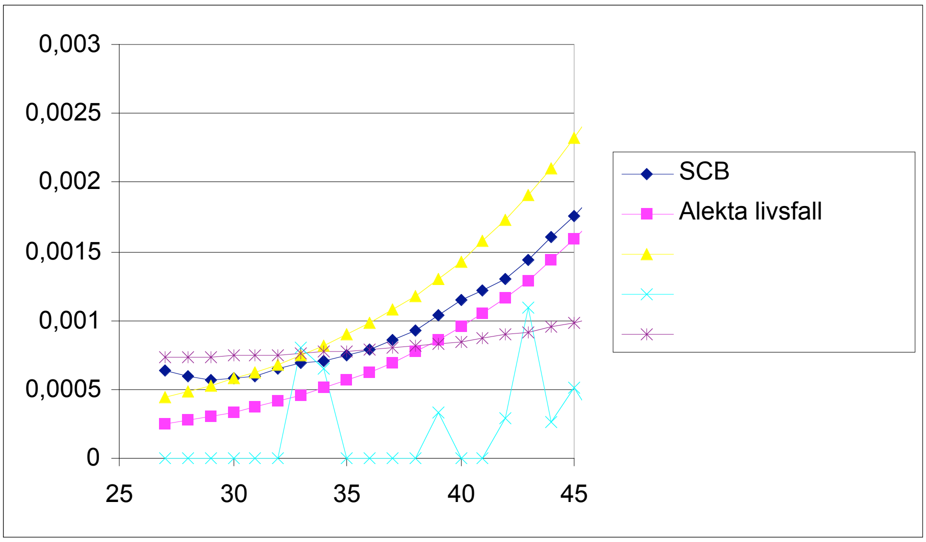
Kvinnor, dödsfallsförsäkring, alla

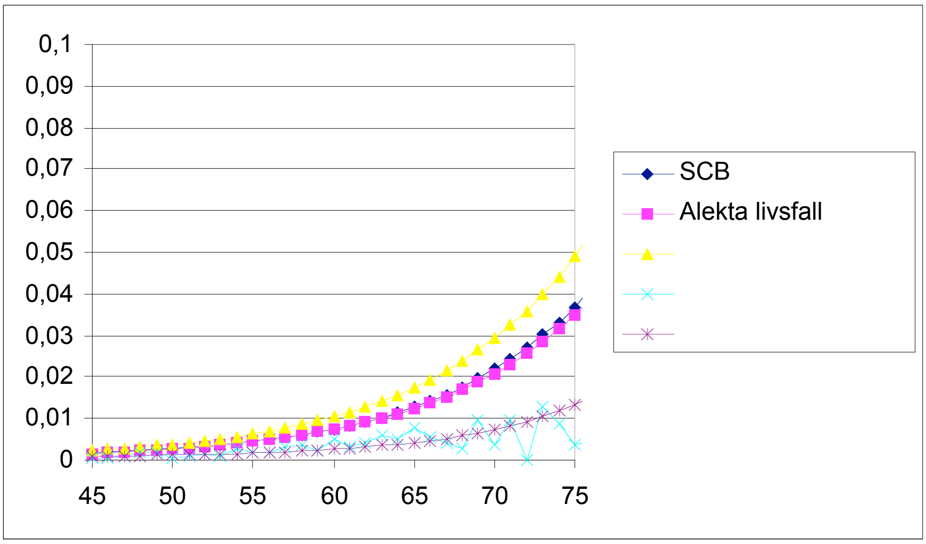




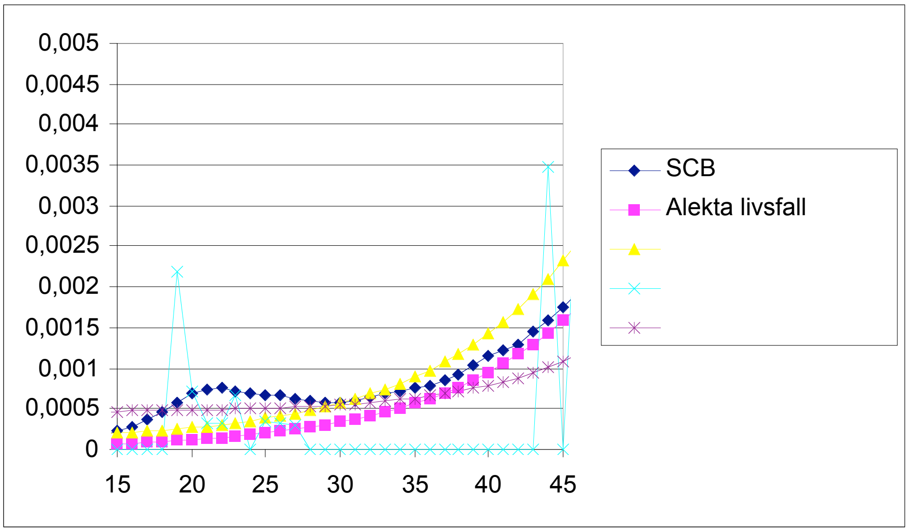


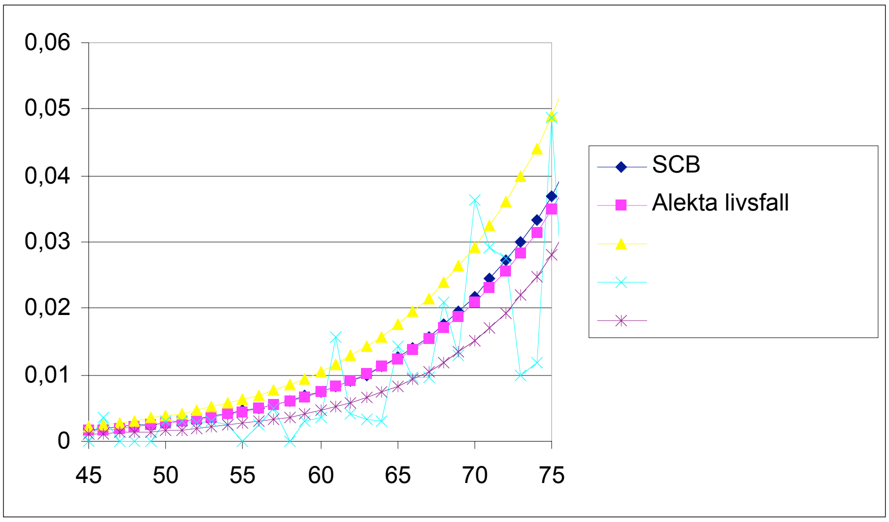
Män, livsfallsförsäkring, hög ekonomisk nivå.



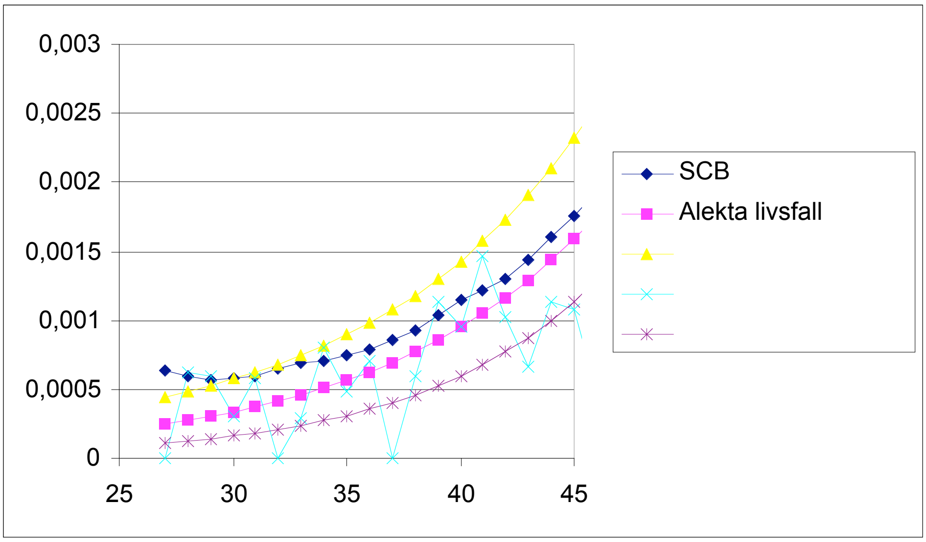


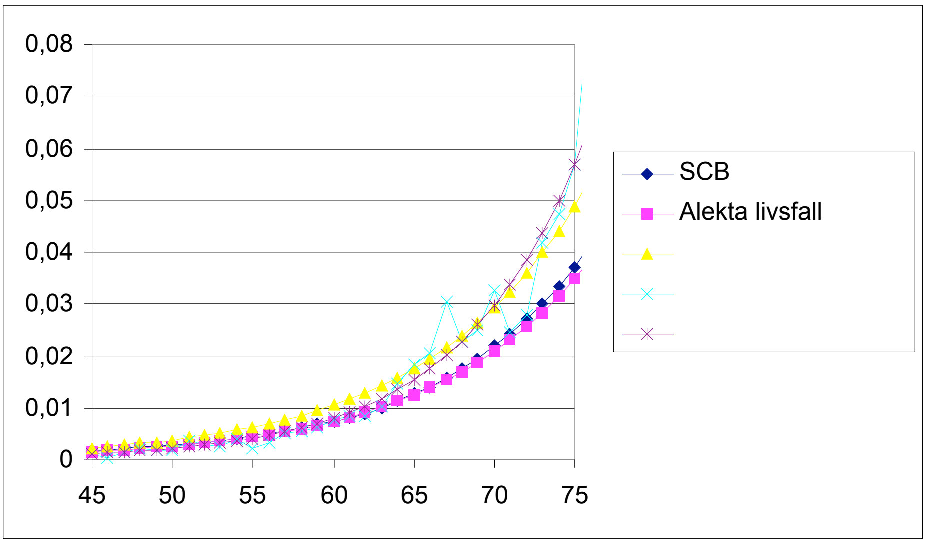
Män, livsfallsförsäkring, ekonomisk medelnivå.

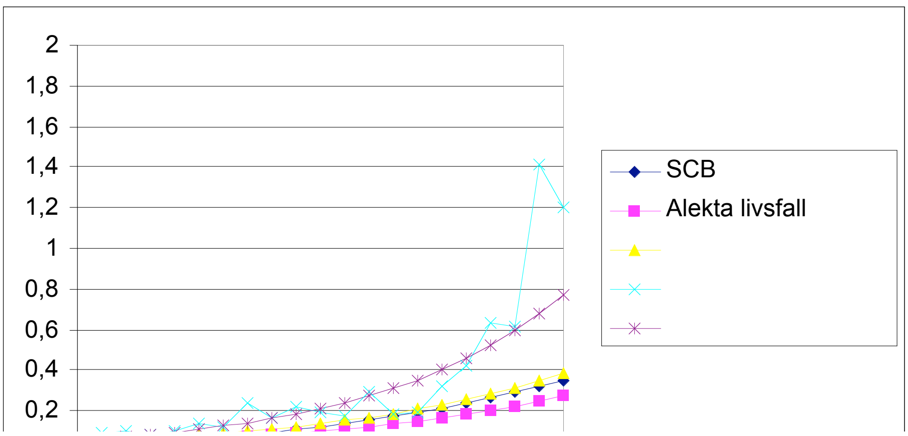




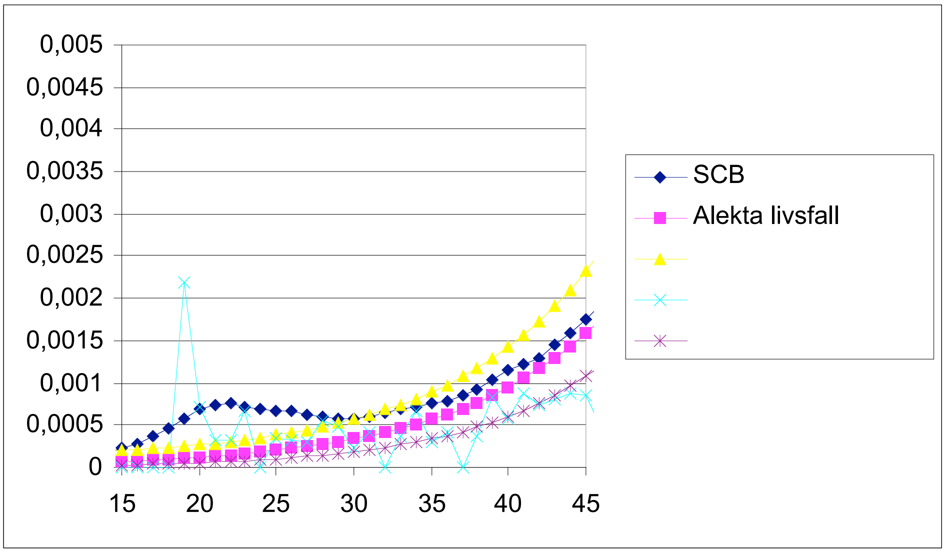
Män, livsfallsförsäkring, låg ekonomisk nivå.

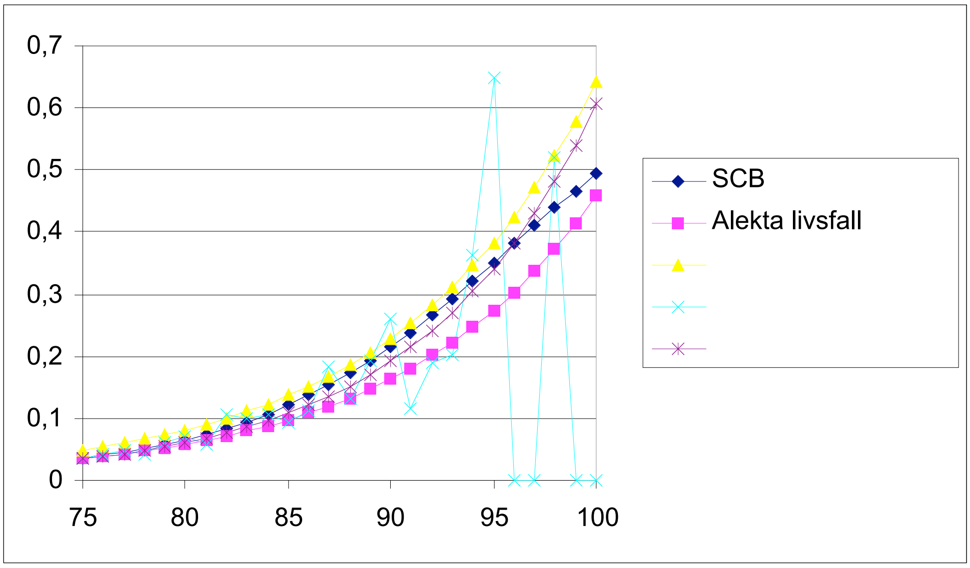

0,003
0,0025
0,002
0,0015
0,001
0,0005
0
25
30
35
40
45
SCB
Alekta livsfall

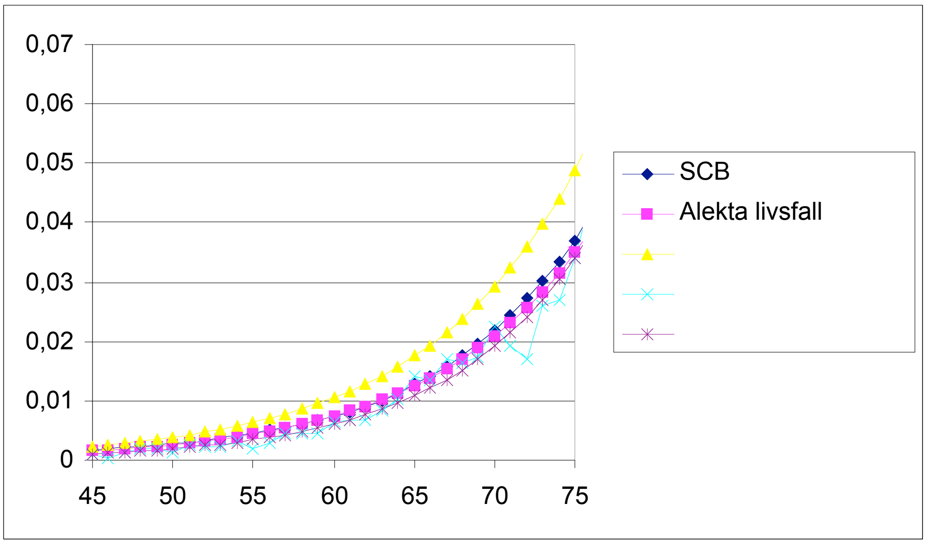

0,08
0,07
0,06
0,05
0,04
0,03
0,02
0,01
0
45
50
55
60
65
70
75
SCB
Alekta livsfall


2
1,8
1,6
1,4
1,2
1
0,8
0,6
0,4
0,2
SCB
Alekta livsfall

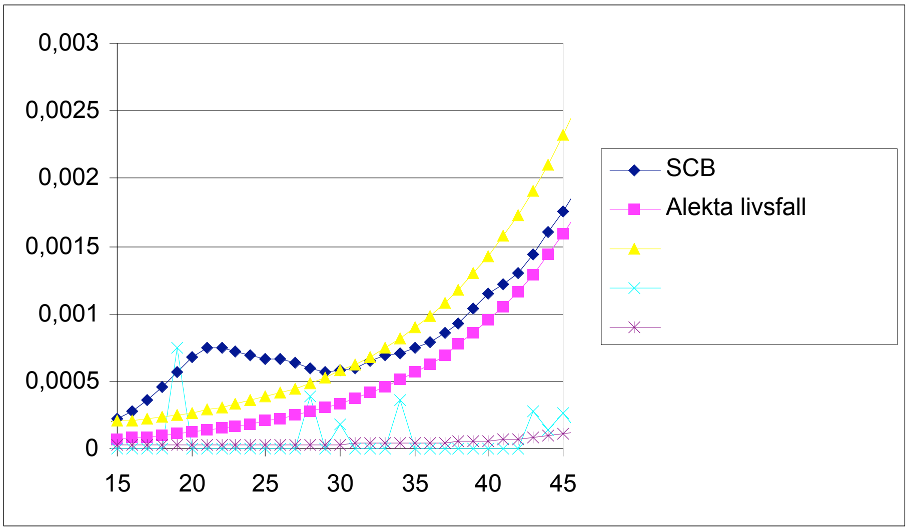
Män, livsfallsförsäkring, alla

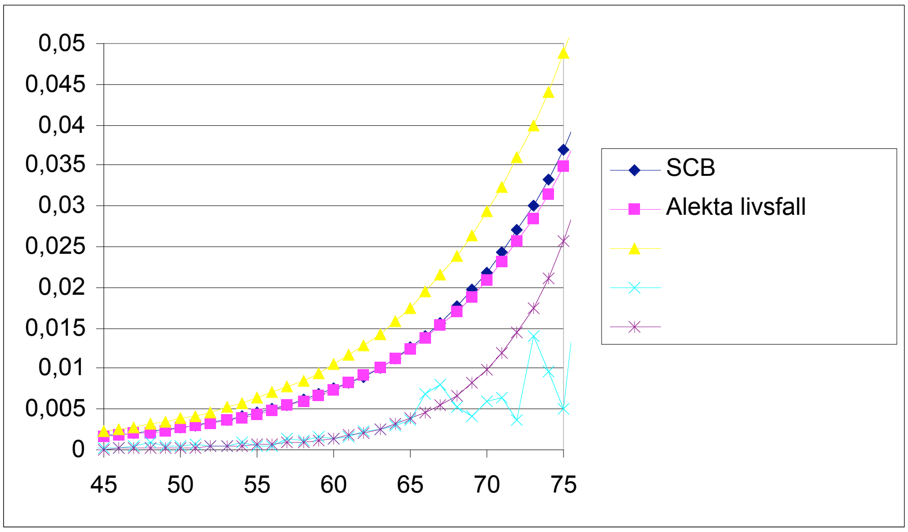


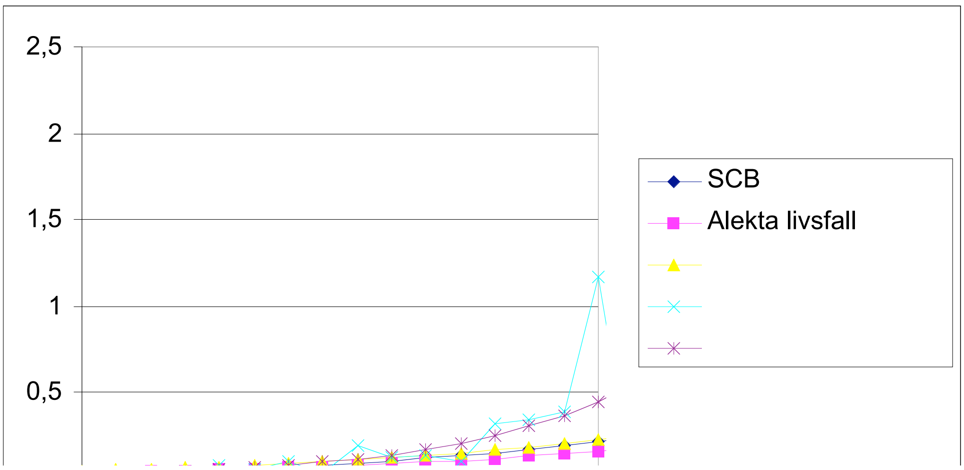




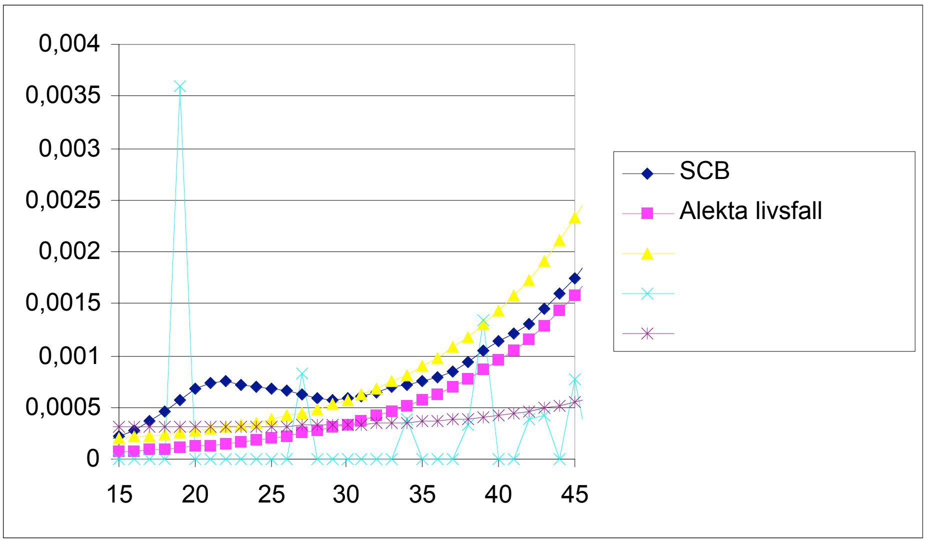
Män, dödsfallsförsäkring, hög

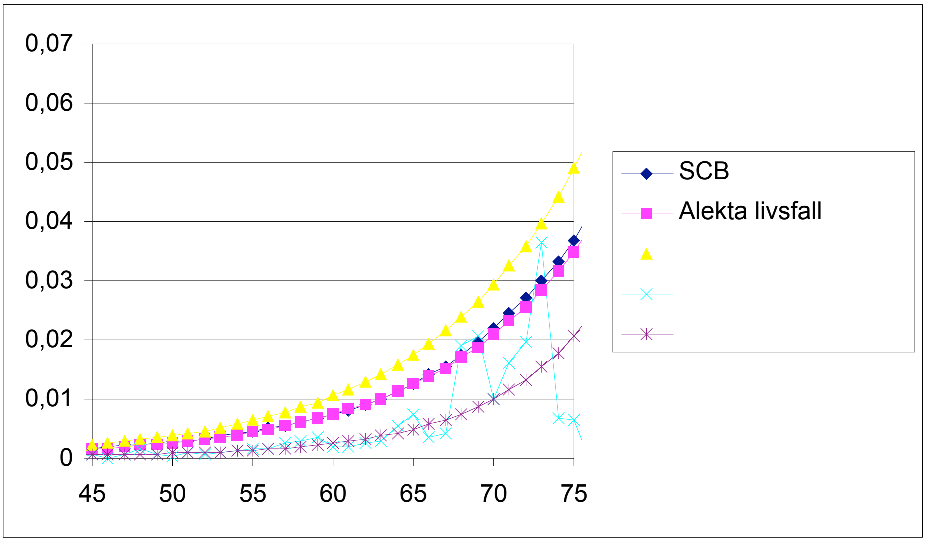


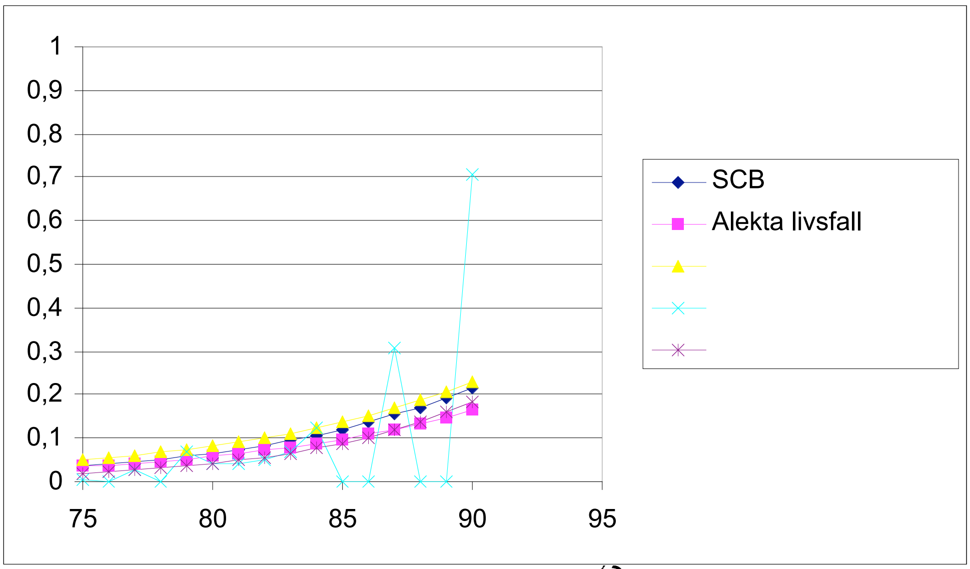




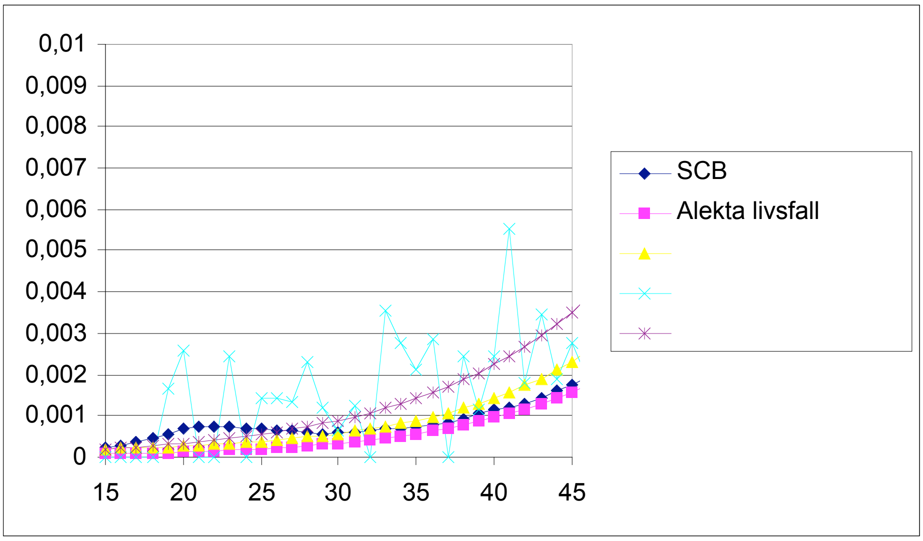
Män, dödsfallsförsäkring, ekonomisk medelnivå.

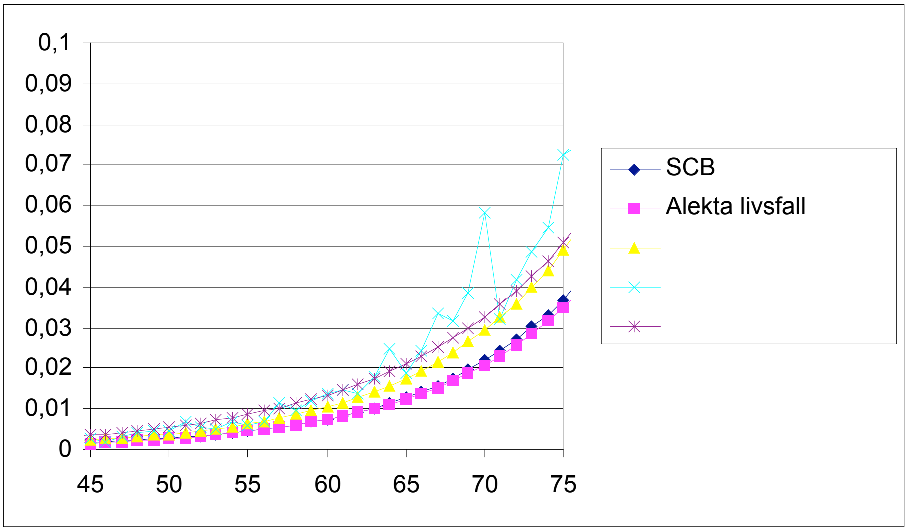


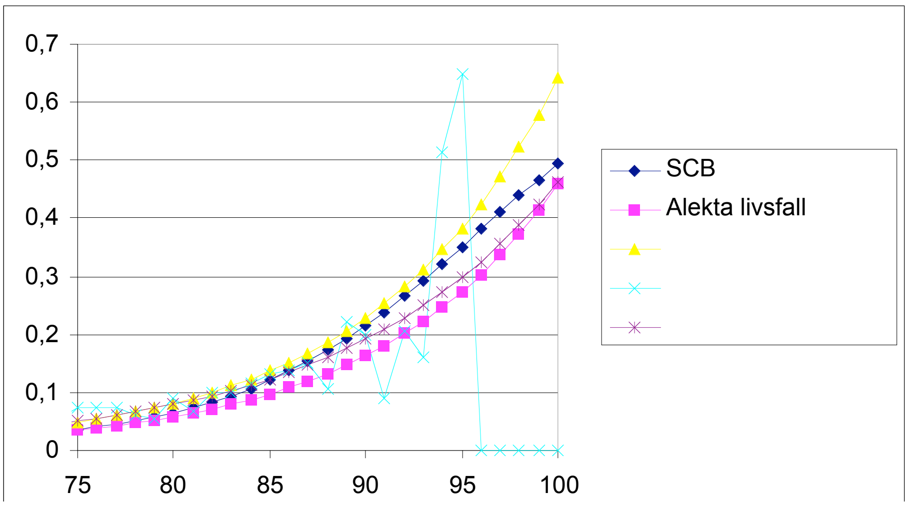




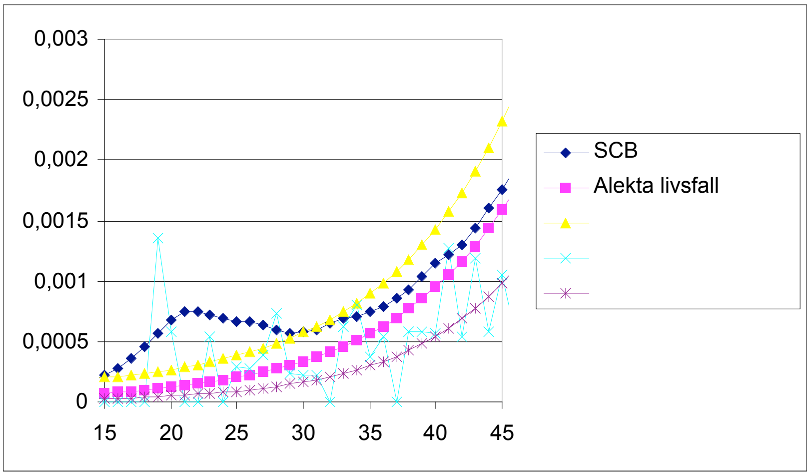
Män, dödsfallsförsäkring, låg ekonomisk nivå.

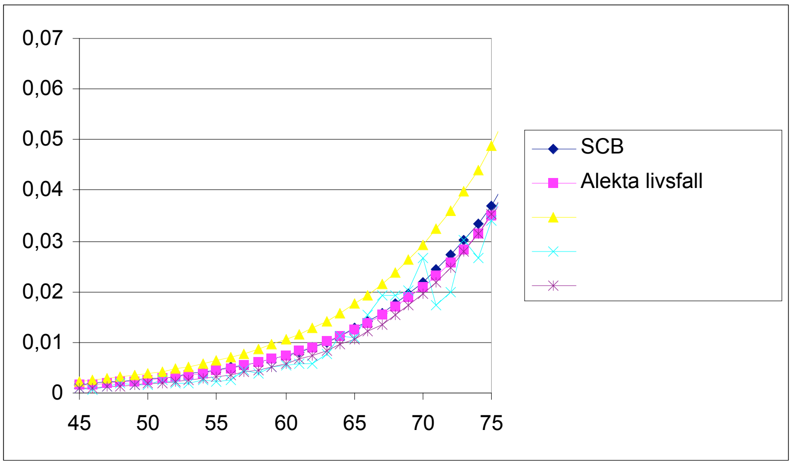


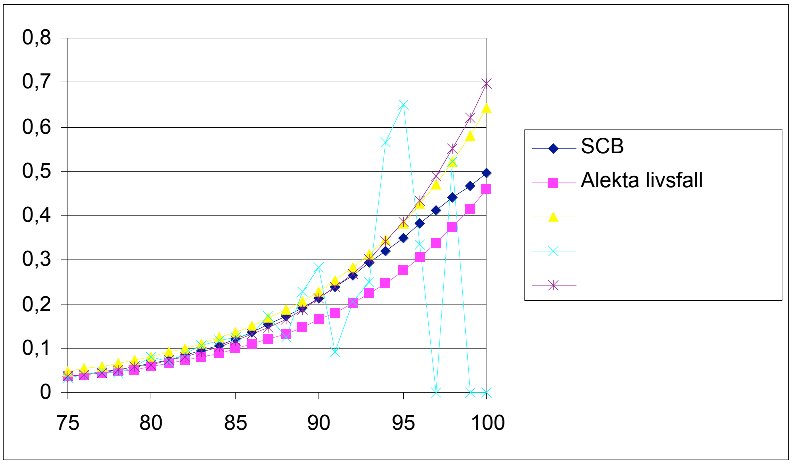




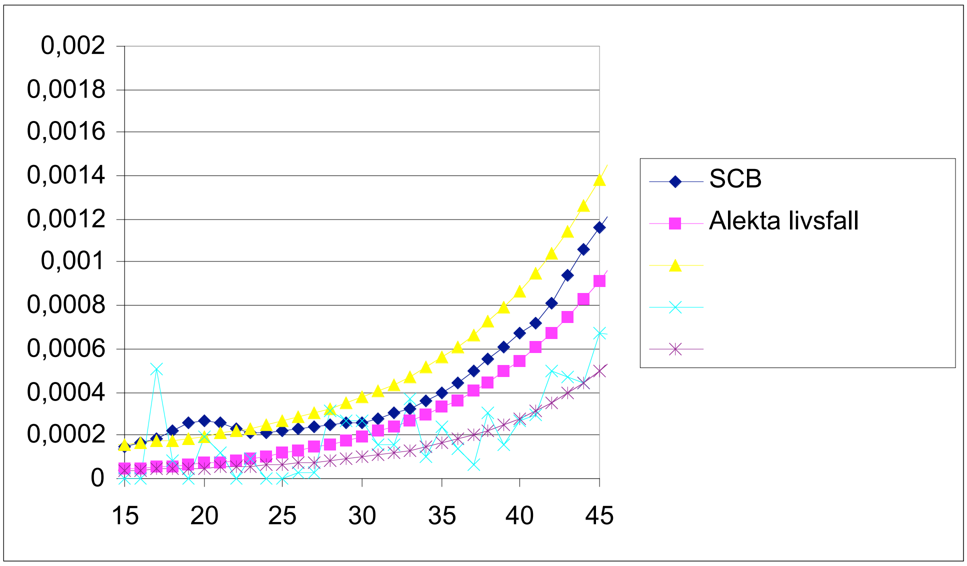
Män, dödsfallsförsäkring, alla

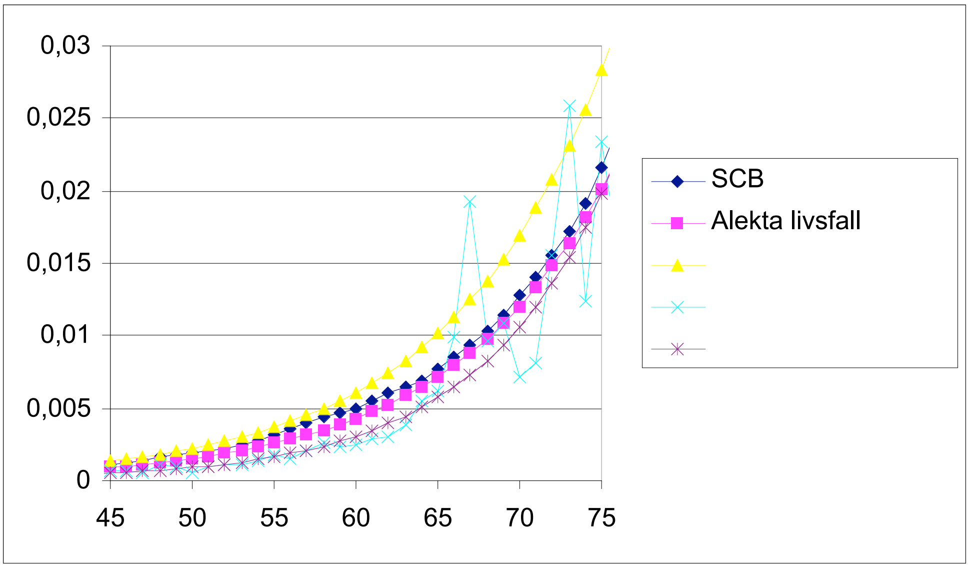


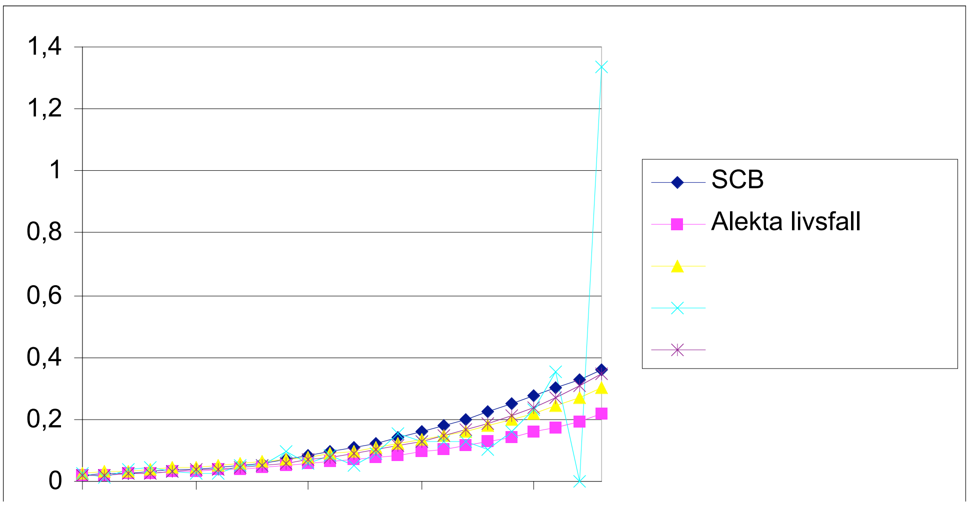




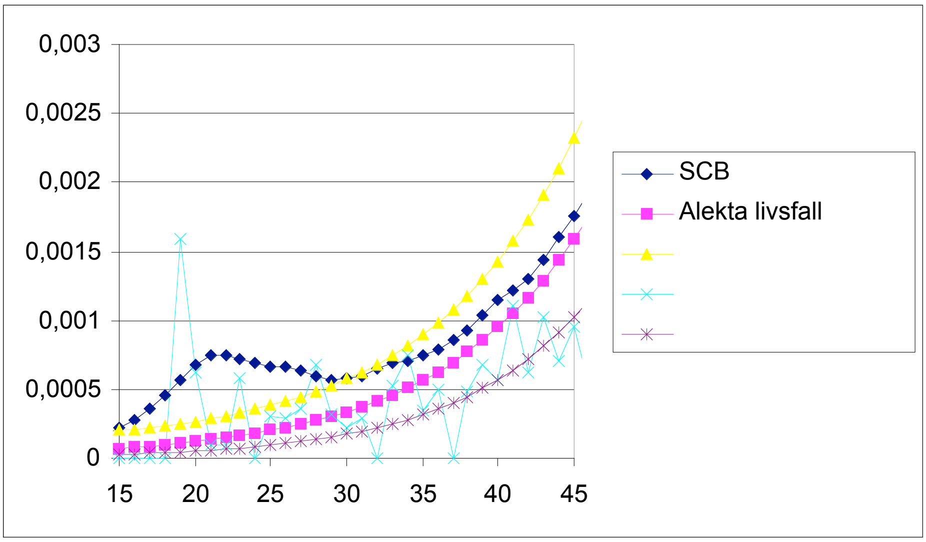
Kvinnor, alla

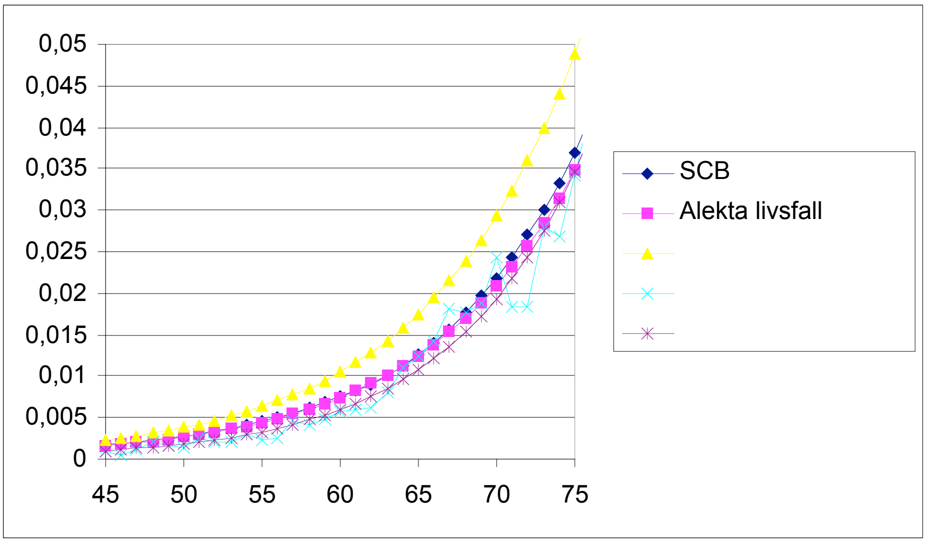


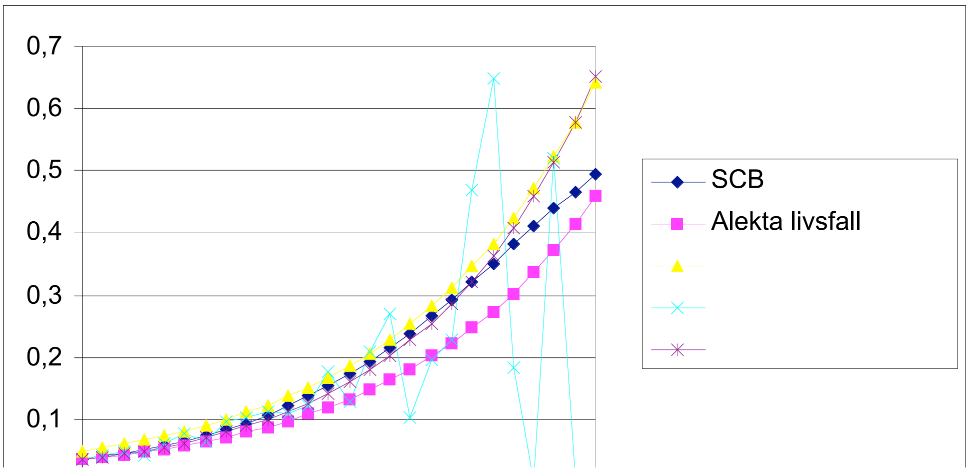




Män, alla

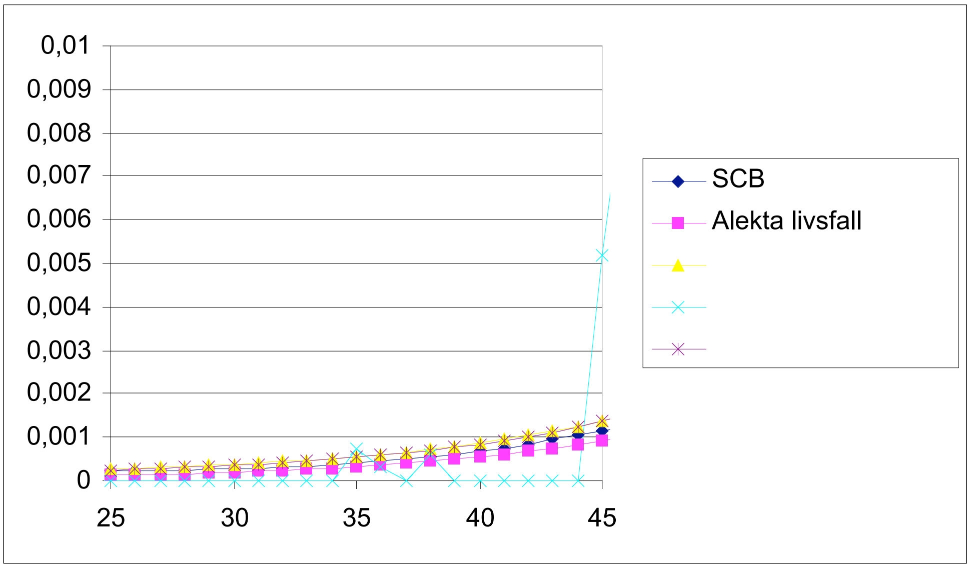

0,003
0,0025
0,002
0,0015
0,001
0,0005
0
15
20
25
30
35
40
45
SCB
Alekta livsfall

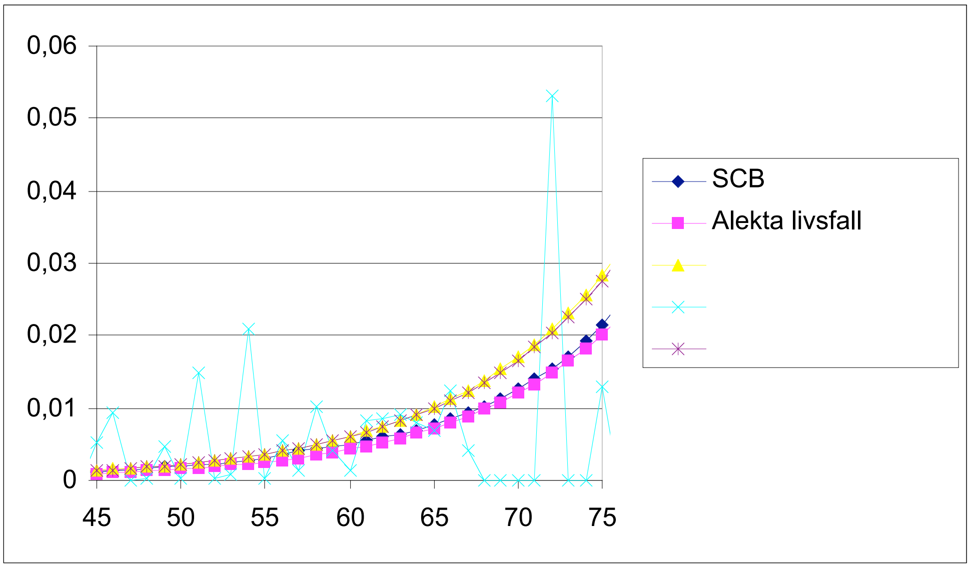

0,05
0,045
0,04
0,035
0,03
0,025
0,02
0,015
0,01
0,005
0
45
50
55
60
65
70
75
SCB
Alekta livsfall

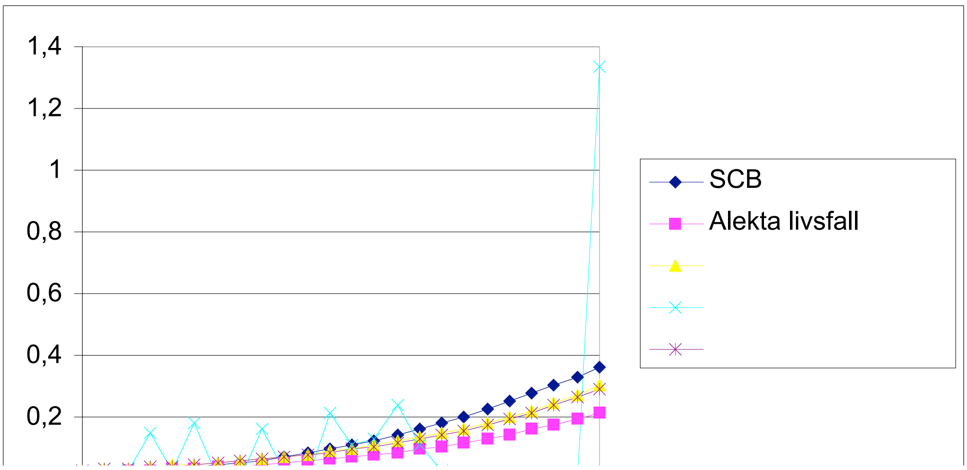

0,7
0,6
0,5
0,4
0,3
0,2
0,1
SCB
Alekta livsfall

3.2.2. Dödlighetsintensitet, ekonomisk.

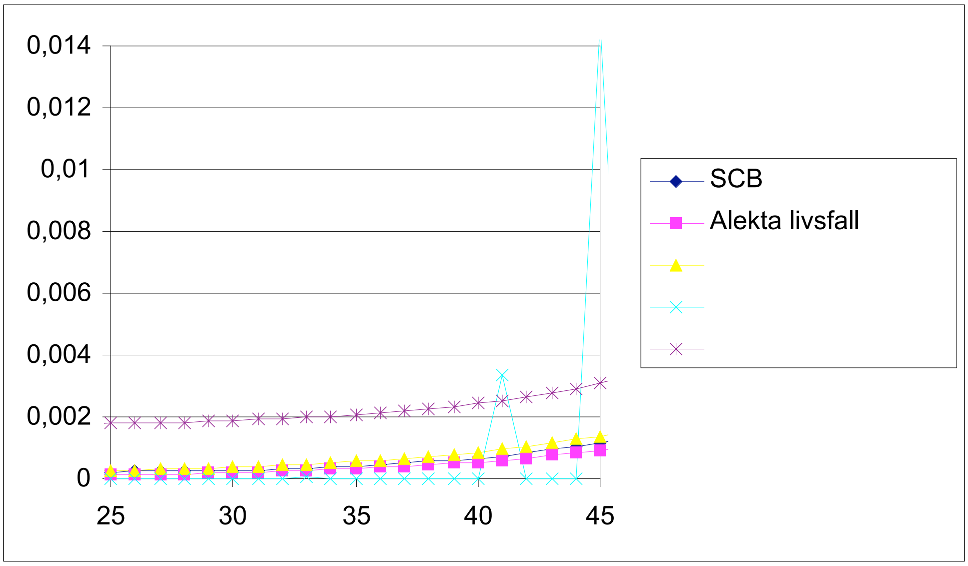
Kvinnor, livsfallsförsäkring, hög ekonomisk nivå.

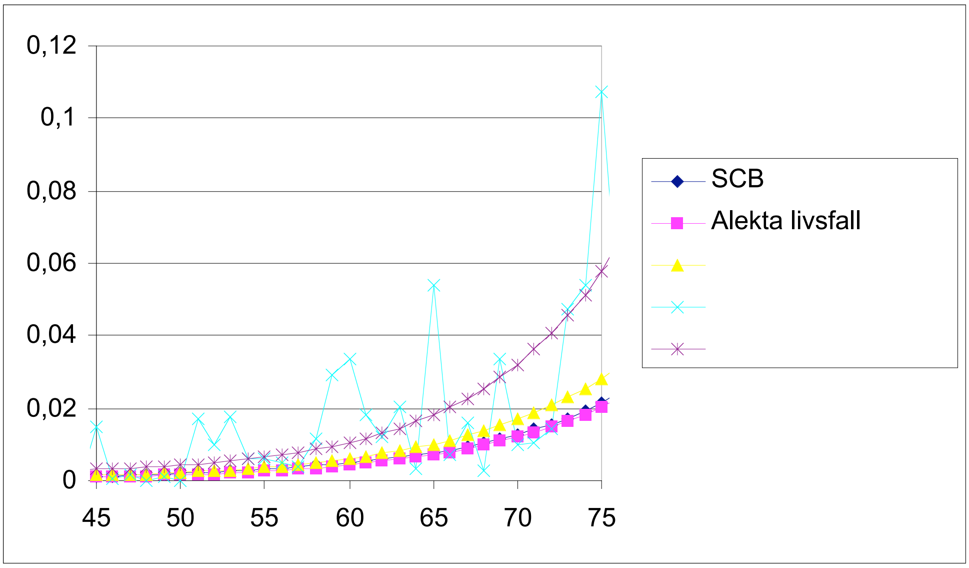


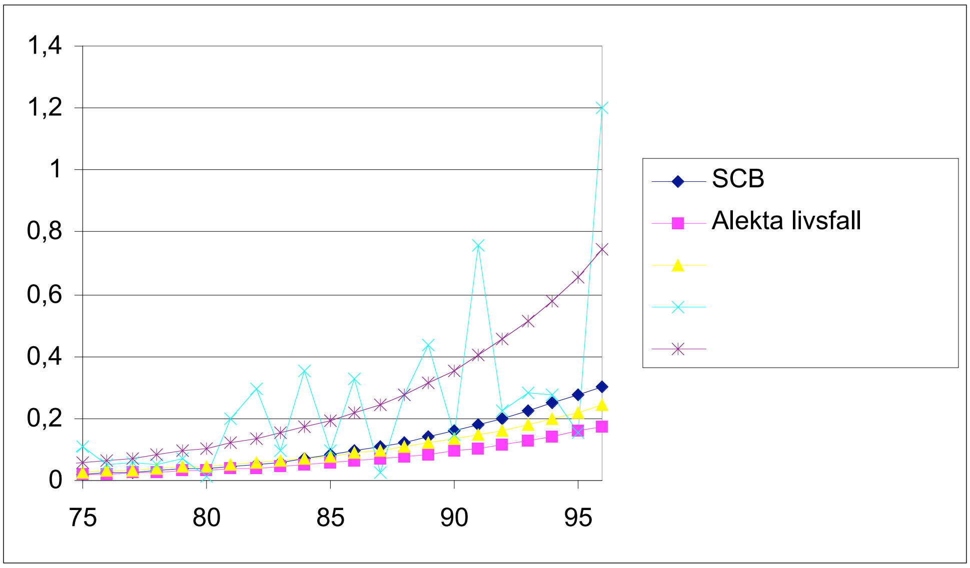




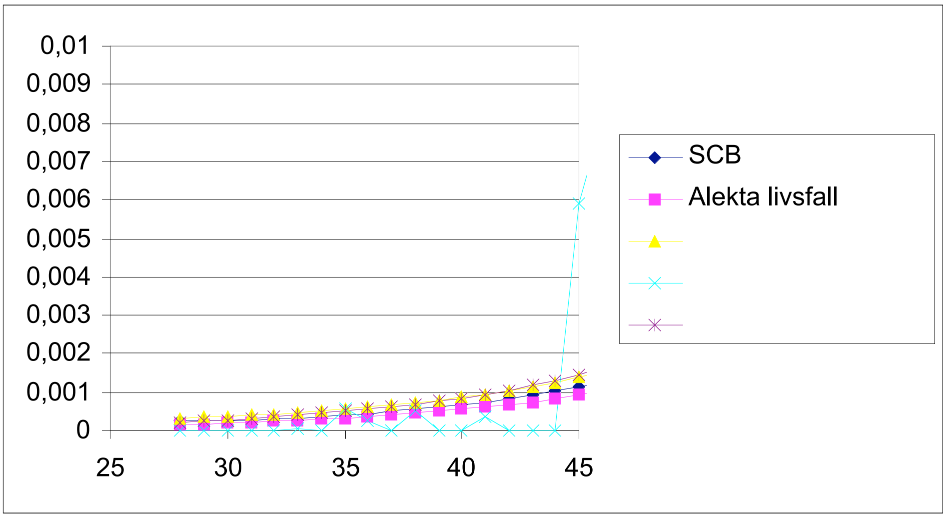
Kvinnor, livsfallsförsäkring, låg ekonomisk nivå.

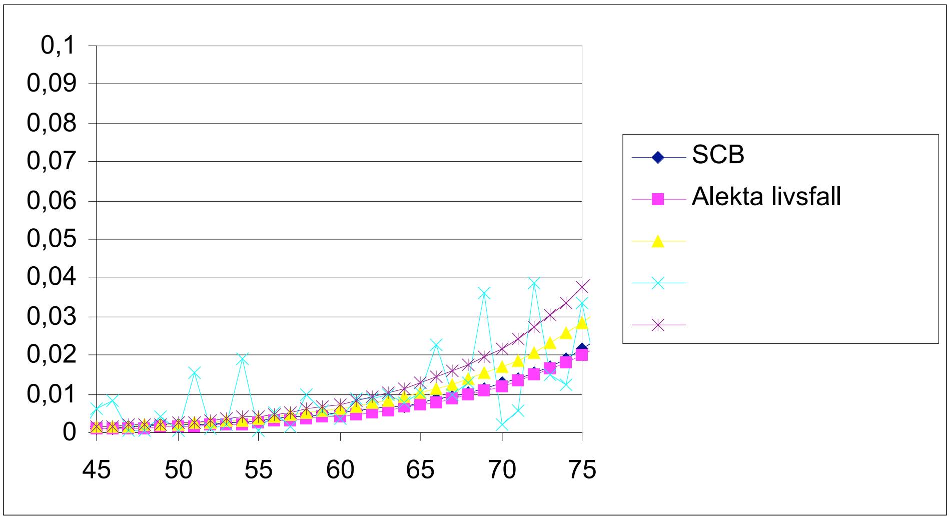


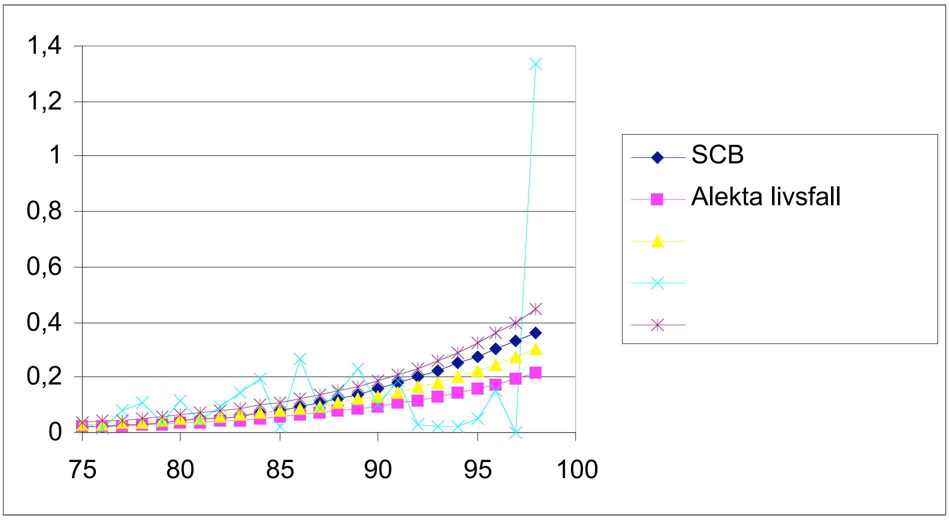




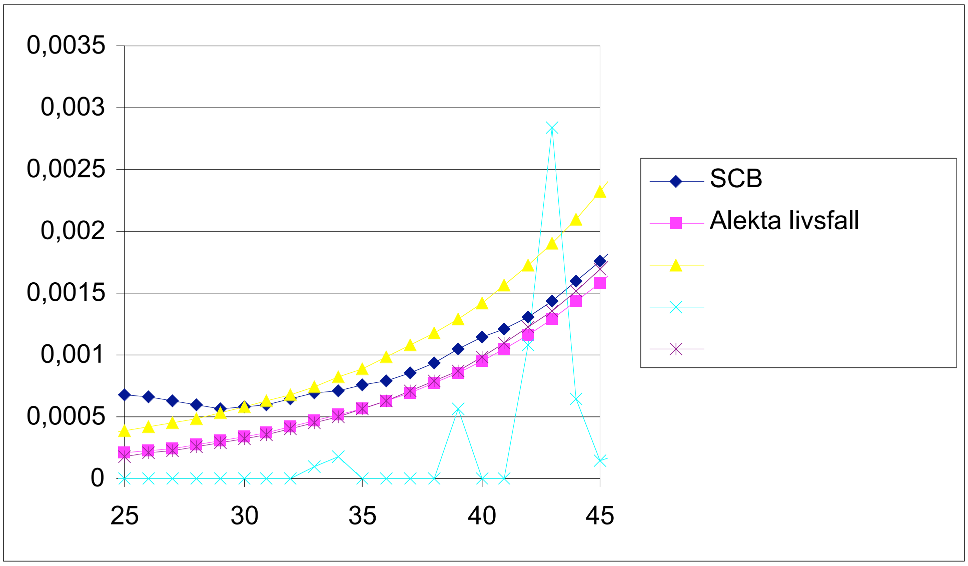
Kvinnor, livsfallsförsäkring, alla.

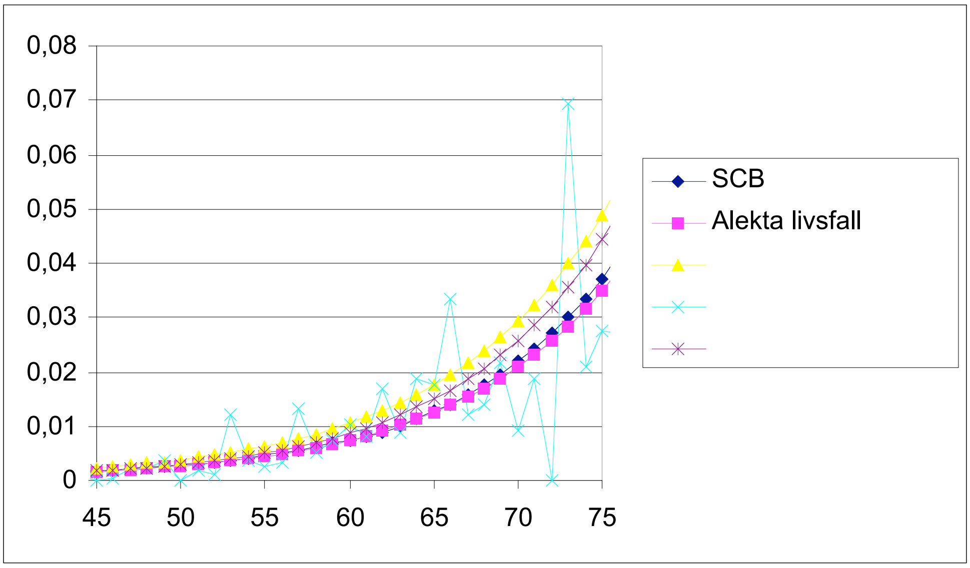


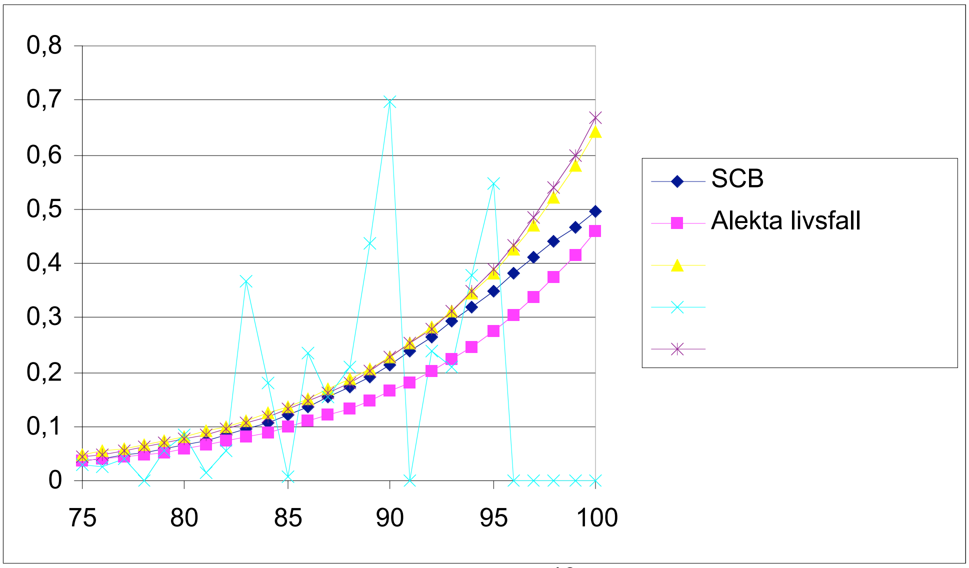




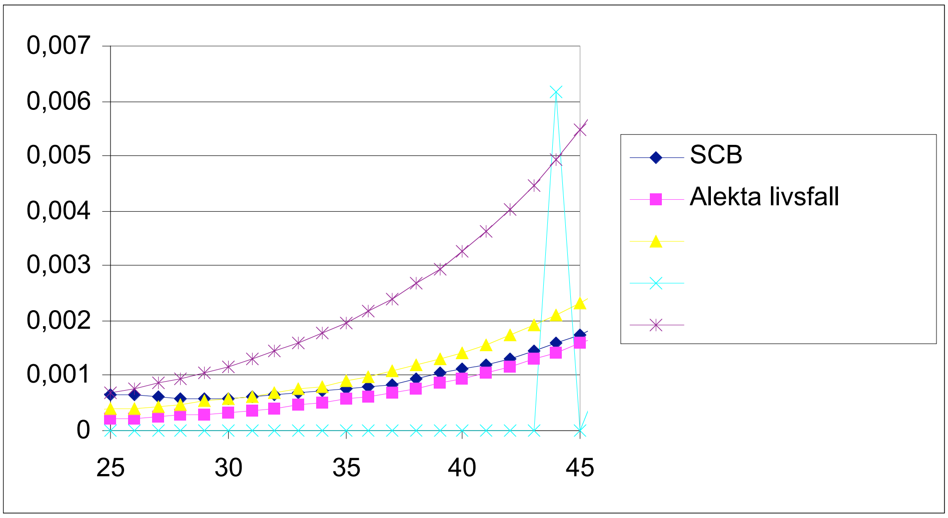
Män, livsfallsförsäkring, hög ekonomisk nivå.

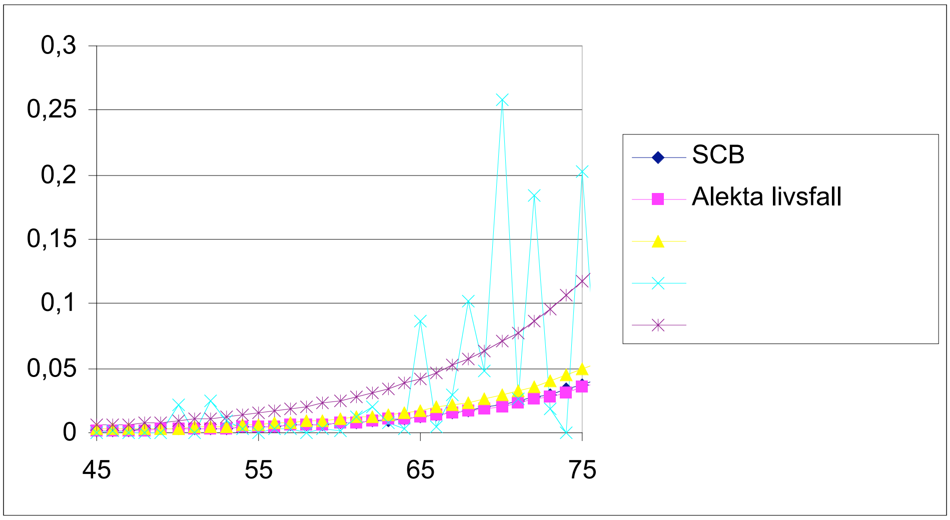


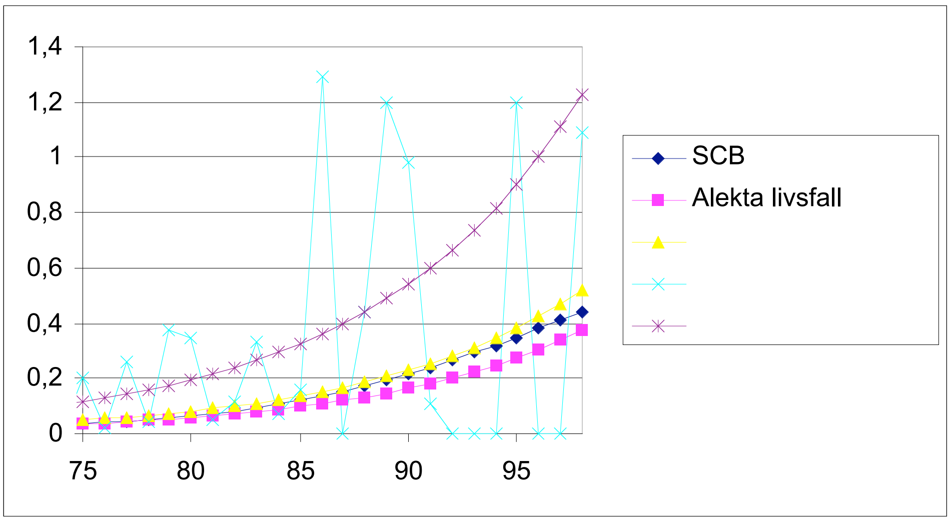




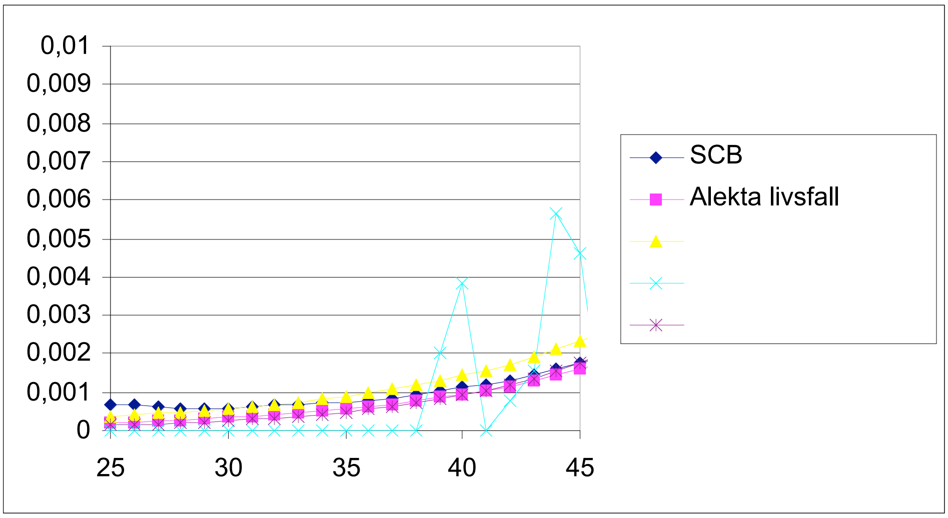
Män, livsfallsförsäkring, ekonomisk medelnivå.

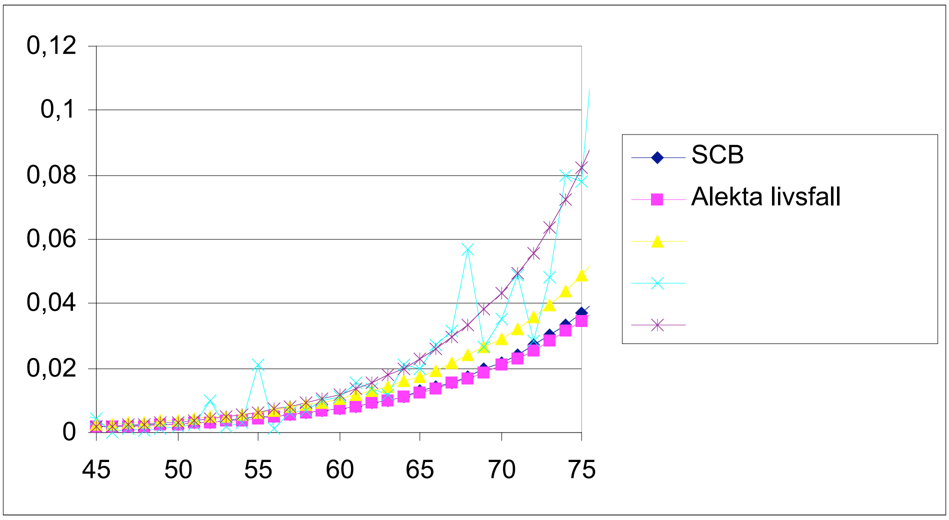


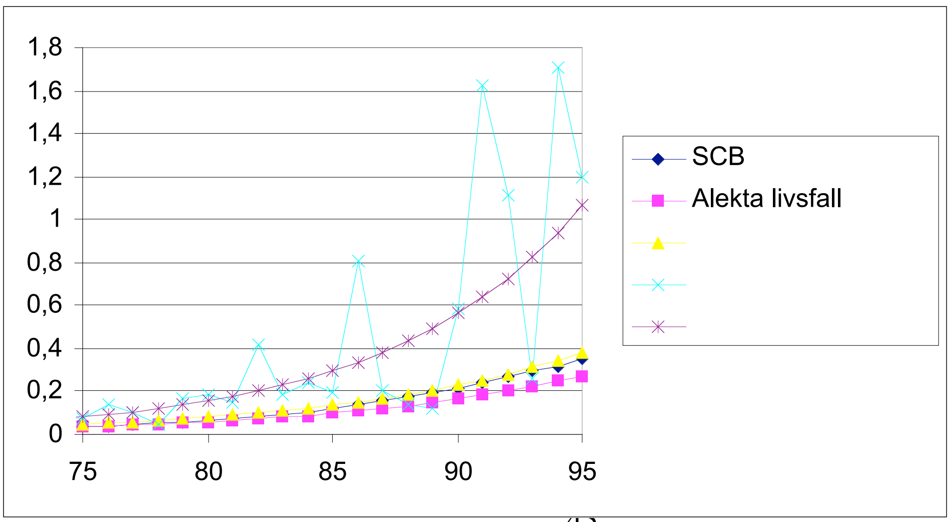




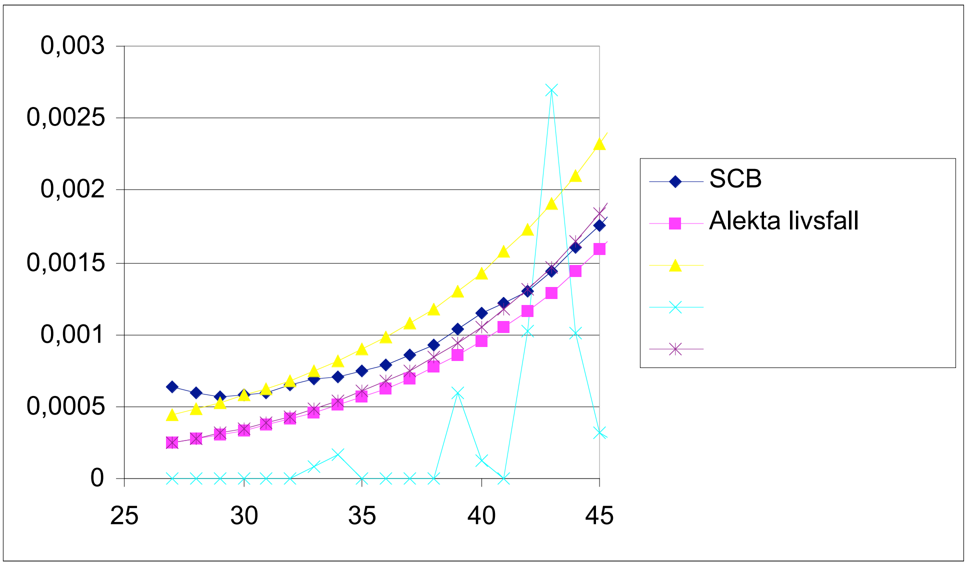
Män, livsfallsförsäkring, låg ekonomisk nivå.

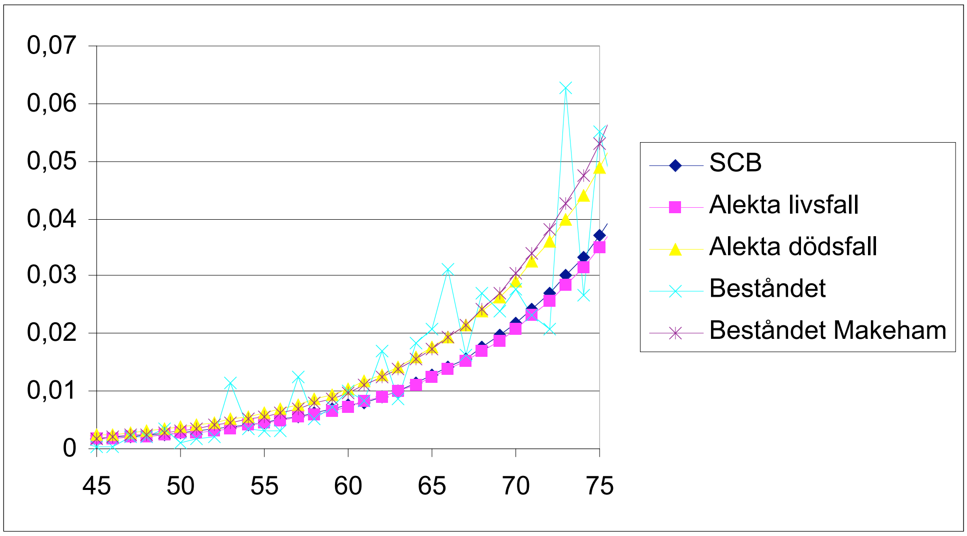


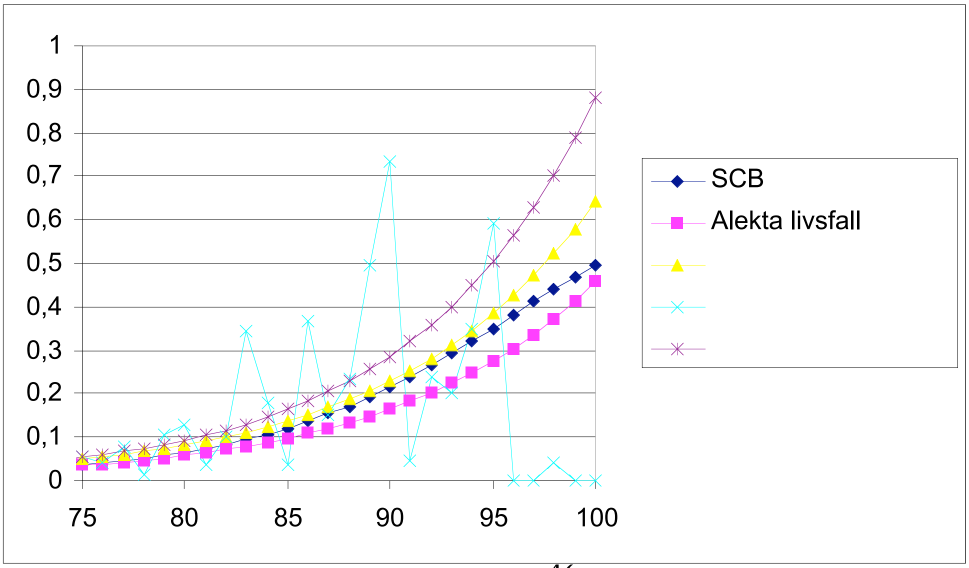




Män, livsfallsförsäkring, alla.

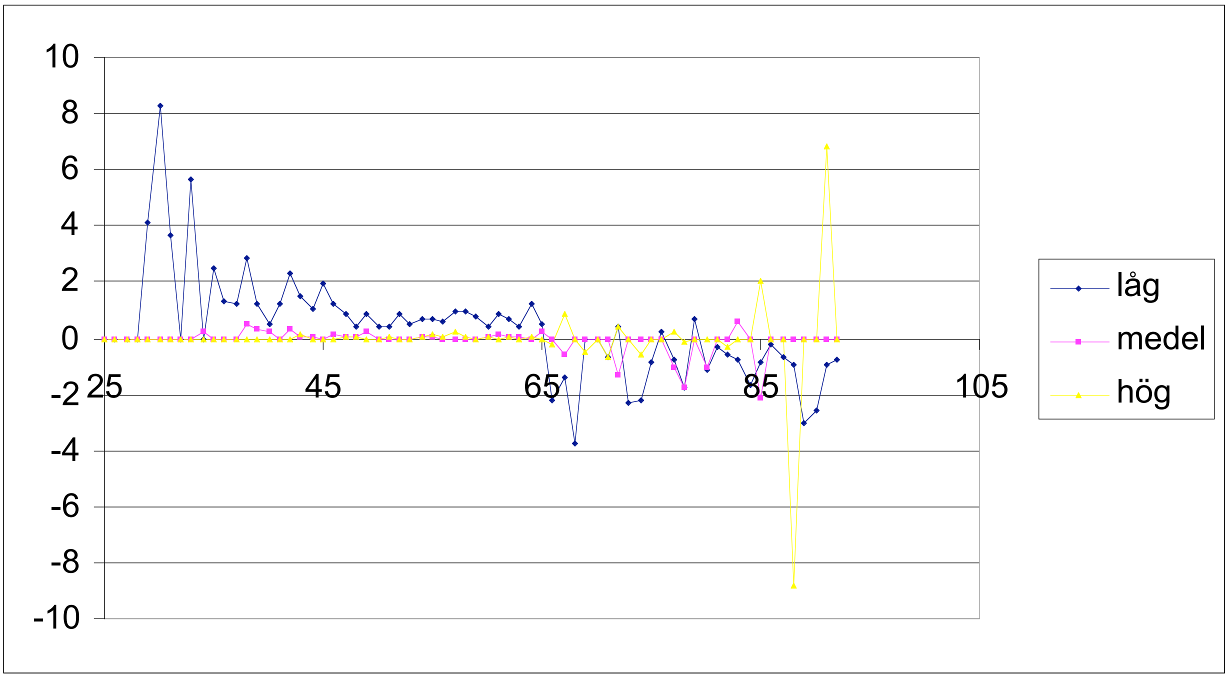




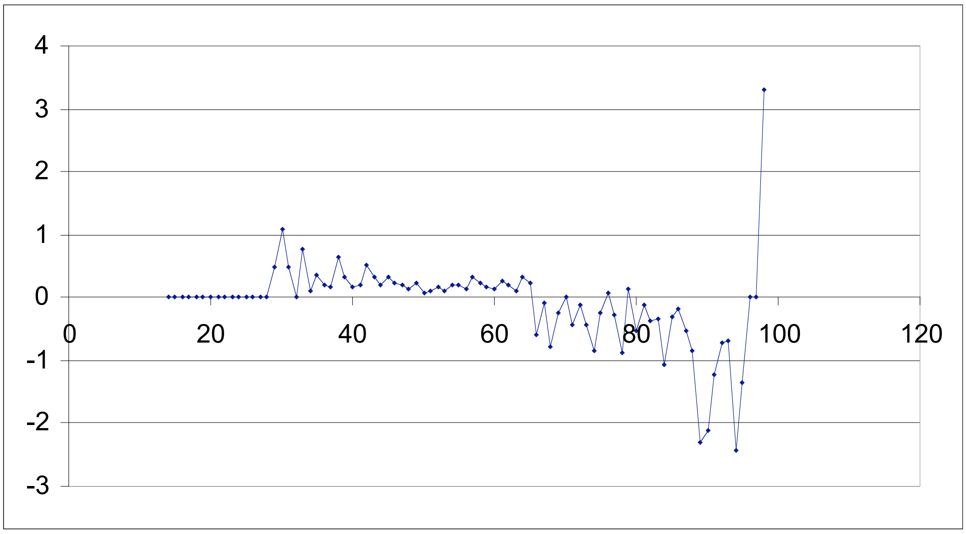


3.3. Relativ ekonomisk dödlighet, dödsfallsförsäkring.

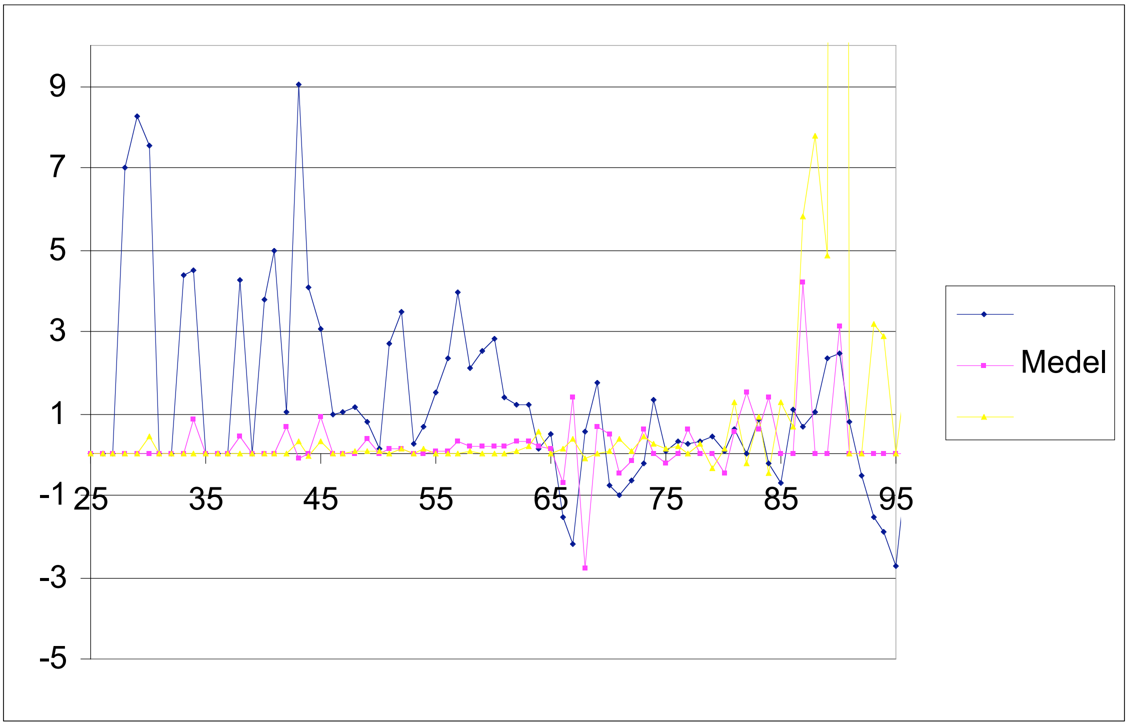
Kvinnor, uppdelat på ekonomisk nivå.



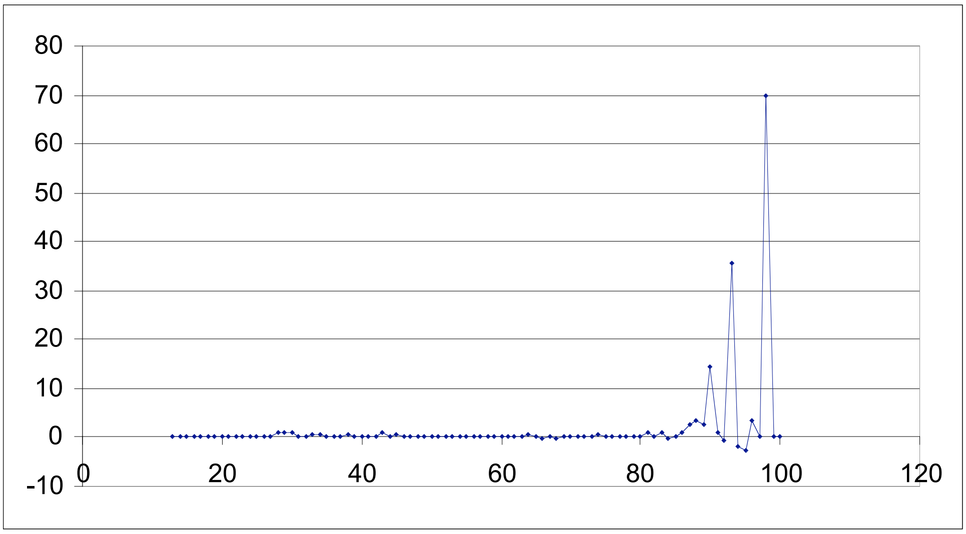
Kvinnor, alla.



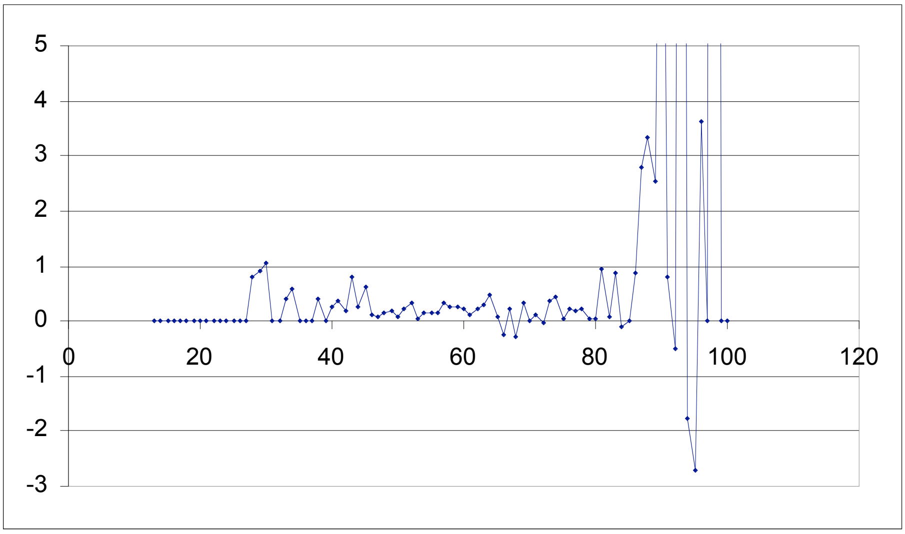
Män, uppdelat på ekonomisk nivå.



Män, alla.



Män, alla, beskuret diagram.



3.4. Utfall hela beståndet.

	livsfallsförsäkring			dödsfallsförsäkring		
	riskpremie	frigjord risksumma	kvot	riskpremie	frigjord risksumma	kvot
1994	-162857	-654712	4,02	624141	-74188	-0,12
1995	-915927	-277429	0,30	2254433	-211608	-0,09
1996	-1641108	-1162570	0,71	5057872	-167469	-0,03
1997	-2402267	-2483886	1,03	8098168	-302918	-0,04
1998	-4955531	-4135552	0,83	91348088	-205841	-0,02
1999	-12736155	-13067747	1,03	12070144	2909486	0,24
2000	-20110456	-19875819	0,99	16009095	3023796	0,19
2001	-23473617	-19565237	0,83	50531166	10379040	0,21
2002	-28766761	-20110041	0,70	82994840	13571729	0,16
2003	-32306205	-26503096	0,82	103729475	19649231	0,19
2004	-37618291	-26340795	0,70	114044936	38409494	0,34

Standardavvikelse livsfallsförsäkring

1994	134316
1995	315166
1996	502800
1997	585639
1998	851189
1999	1121239
2000	1719162
2001	2045544
2002	2186620
2003	2288569
2004	2462162

3.5. Åldersförskjutning.

Män livsfall hög – kvinnor livsfall hög 13,6 år

Män livsfall medel – kvinnor livsfall medel 6,3 år

Män livsfall låg – kvinnor livsfall låg 4,1 år

Män dödsfall hög – kvinnor dödsfall hög 3,4 år

Män dödsfall medel – kvinnor dödsfall medel 10,8 år

Män dödsfall låg – kvinnor dödsfall låg 8,4 år

Män dödsfall hög – man livsfall hög 3,4 år

Man dödsfall medel – man livsfall medel -4,6 år

Man dödsfall låg – man livsfall låg 1,6 år

Kvinnor dödsfall hög – kvinnor livsfall hög -0,4 år

Kvinnor dödsfall medel – kvinnor livsfall medel –1,2 år

Kvinnor dödsfall låg – kvinnor livsfall låg -7,2 år

3.6. Resultat logistisk regression.

Modellen logistisk regression innebär att dödlighetsintensiteten kan skrivas som

$\mu(x_i) = \dfrac{e^{\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_r x_{ri}}}{1 + e^{\beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_r x_{ri}}}$, där β_j-parametrarna beskriver hur parametern x_j påverkar μ. 95% konfidensintervall för parametervärdena liksom diskrepanser, givet de olika modellerna, ges av följande tabeller: (Observera att åldersparametrarna inte redovisas.)

Åldersintervaller med bredd ett år

	Kön	Försäkrings-typ	Ekonomisk klass		Trend	Diskrepans
			Hög	Låg		
H1	0,58 – 0,70					2247
H2	0,58 – 0,69	0,03 – 0,14				2237
H3	0,70 – 0,81		-0,35 – -0,10	1,19 – 1,40		45,1
H4	0,57 – 0,68				-0,28 – -0,12	2225
H5	0,56 – 0,68	0,03 – 0,14			-0,28 – -0,12	2215
H6	0,70 – 0,82	-0,23 – -0,12	-0,32 – -0,07	1,24 – 1,45		8,7
H7	0,69 – 0,80		-0,35 – -0,10	1,19 – 1,40	-0,21 – -0,04	36,9
H8	0,69 – 0,81	-0,23 – -0,11	-0,32 – -0,07	1,24 – 1,45	-0,21 – -0,05	

Åldersintervaller med bredd fem år

	Kön	Försäkrings-typ	Ekonomisk klass		Trend	Diskrepans
			Hög	Låg		
H1	0,59 - 0,71					2253
H2	0,59 – 0,70	0,04 – 0,15				2241
H3	0,71 – 0,83		-0,35 – -0,10	1,19 –1,40		42,4
H4	0,57 – 0,69				-0,28 – -0,12	2230
H5	0,57 – 0,68	0,04 – 0,15			-0,28 – -0,12	2218
H6	0,71 – 0,83	-0,22 – -0,11	-0,32 – -0,08	1,24 – 1,45		8,7
H7	0,70 – 0,82		-0,35 – -0,10	1,19 – 1,40	-0,20 – -0,04	34,0
H8	0,70 – 0,82	-0,22 – -0,11	-0,22 – -0,11	1,24 – 1,45	-0,20 – -0,04	

Åldersintervaller med bredd tio år

	Kön	Försäkrings-typ	Ekonomisk klass		Trend	Diskrepans
			Hög	Låg		
H1	0,64 – 0,75					2298
H2	0,62 – 0,74	0,06 – 0,17				2281
H3	0,76 – 0,87		-0,38 – -0,14	1,18 – 1,39		44,9
H4	0,62 – 0,74				-0,27 – -0,12	2275
H5	0,61 – 0,73	0,05 – 0,17			-0,28 – -0,12	2259
H6	0,77 – 0,88	-0,23 – -0,12	-0,36 – -0,11	1,24 – 1,45		7,5
H7	0,75 – 0,86		-0,38 – -0,14	1,17 – 1,38	-0,19 – -0,03	37,8
H8	0,75 – 0,87	-0,23 – -0,12	-0,35 – -0,11	1,24 – 1,45	-0,19 – -0,03	

Här vill jag påminna om att värdena i tabellen är skattad förändring i logoddskvot då man går från kvinna till man, från dödsfallsförsäkring till livsfallsförsäkring, från ekonomisk klass Medel till Hög, från ekonomisk klass Medel till Låg, respektive från år 1994-1999 till 2000-2004.
För att diskrepanserna skall vara approximativt χ^2 – fördelade krävs att antalet döda i åtminstone 80% av antalet celler skall vara minst fem st. Så är långt från fallet. Vi kan alltså inte använda någon färdig metod som ger statistiskt signifikanta resultat, utan måste använda mer skönsmässiga metoder för hypotesprövning. Även konfidensintervallen är osäkra.
Vi ser dock att vissa effekter tydligt fångas upp. I alla 18 modeller är parametervärdet för kön ca 0,6 – 0,9.
Konfidensintervallen för försäkringstyp är i alla modeller nära 0 och dessutom på olika sidor. Alltså kan man dra slutsatsen att försäkringstypens inverkan är i det närmaste ingen alls. Vidare är diskrepansskillnaderna mellan olika närliggande modeller med respektive utan försäkringstyp liten, vilket styrker påståendet. Detta är inte så underligt eftersom de i materialet ingående personerna ofta är både livsfalls- och dödsfallsförsäkrade. I ett bestånd med individuellt försäkrade skulle man kunna förvänta sig att dödsfallsförsäkrade dör mer än livsfallsförsäkrade, p.g.a. moturval.
Ekonomisk klass påverkar dödligheten tydligt. Går man från en modell med ekonomisk klass till en annan modell med samma ingående parametrar utom just ekonomisk klass ökar diskrepansen väsentligt. Vidare är parametervärdet för ekonomisk klass Låg högre än för kön. Modellerna ger alltså att kvinnor med låg ekonomisk klass har högre dödlighetsintensitet än lika gamla män med medelhög ekonomisk klass. Personer med hög ekonomisk klass har något lägre dödlighetsintensitet än personer med medelhög ekonomisk klass.
Värdena för Trend har konfidensintervallsgränser nära noll. Diskrepansskillnaderna mellan modeller med och utan Trend är måttliga. Vi törs alltså inte dra någon säker slutsats om eventuell trend. Trenden skulle möjligen vara svagt nedåtgående.
Jag har inte använt något färdigt programpaket som t.ex. SAS, utan programmerat "från början" i Mathcad.

Referenslista

Andersson, Gunnar; Livförsäkringsmatematik, Svenska Försäkringsföreningen, 2005

Bowers, N.L. ; Actuarial Mathematics, The Society of Actuaries, Schaumburg, Illinois. (s.38 – 41) . 1997

Blom, Richard, Livförsäkringsmatematik, Ifu förlag, 2002

Gut, Allan; An intermediate course in Probability Theory, Springer Verlag, 1995

KPA; PFA Pension för dig född 1938 eller senare. Broschyr PFA.

Ohlsson, Esbjörn; Log-linjära modeller och Logistisk regression, Stockholms universitet, 2000.

www.scb.se; Ettåriga dödsrisker