Blätter DGVFM (2008) 29: 117–126
DOI 10.1007/s11857-007-0034-y

ACTUARIAL EXAMS

Bericht zur Prüfung im Mai 2007 über Stochastische Risikomodellierung und statistische Methoden (Grundwissen)

Dietmar Pfeifer · Gerald Sussmann · Richard Herrmann

Received: 5 September 2007 / Accepted: 5 September 2007 / Published online: 1 Februar 2008

Am 5. Mai 2007 wurde zum ersten Mal eine Prüfung im Fach „Stochastische Risikomodellierung und statistische Methoden“ durchgeführt. Insgesamt konnten maximal 102 Punkte erreicht werden, zum Bestehen der Prüfung waren mindestens 50 Punkte notwendig. Von 72 Teilnehmern haben 63 die Prüfung bestanden. Zugelassen waren ein Taschenrechner und das Folienskript zum Kurs und Repetitorium „Stochastische Risikomodellierung und statistische Methoden“.

Aufgabe 1 (36 Punkte)

Bei einer Versicherungsgesellschaft wurden in den letzten 4 Jahren folgende inflations- und bestandsbereinigte Jahresschäden im Feuersegment verzeichnet (in Mio €):

Jahr	2003	2004	2005	2006
Schaden	3	1	10	6

a) Zeichnen Sie den zeitlichen Verlauf der Schadenentwicklung und kommentieren Sie die Graphik.

b) Es soll mit Hilfe von Q-Q-Plots entschieden werden, ob die Daten einer Exponentialverteilung $\mathcal{E}(\lambda)$ mit $\lambda > 0$ oder einer Pareto-Verteilung $\mathcal{P}a(\alpha)$ mit $\alpha > 0$ entstammen.
   - Geben Sie die log-Likelihood der beiden Verteilungen an!
   - Geben Sie für jede der genannten Verteilungen eine notwendige Bedingung für die Maximum-Likelihood-Schätzer $\hat{\lambda}$ und $\hat{\alpha}$ an!
   - Berechnen Sie die Maximum-Likelihood-Schätzer $\hat{\lambda}$ und $\hat{\alpha}$!

c) Zeichnen Sie auf der Basis der unter b) gefundenen Schätzer die zugehörigen Q-Q-Plots und begründen Sie, für welche Modellhypothese Sie sich entscheiden.

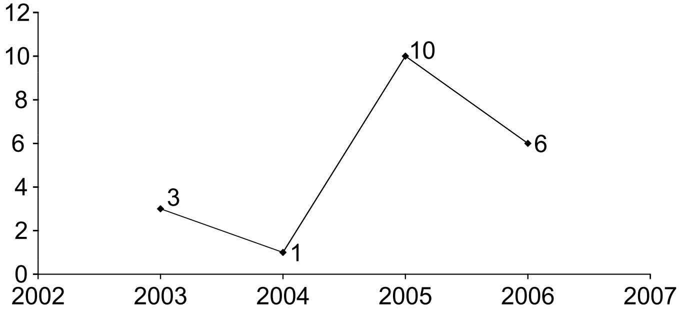
d) Berechnen Sie auf der Basis der unter c) getroffenen Entscheidung den Var und den TVaR zum Risikoniveau 0,005 (entsprechend einer Wiederkehrperiode von 200 Jahren).

Gerald Sussmann, München, Deutschland, e-mail: gerald.sussmann@vkb.de

Lösung:

Zu a)

Es ergibt sich folgende Graphik:



Beurteilung: wenig aussagefähig, zu wenige Daten, geringer positiver Trend?

Zu b)

Die Log-Likelihood-Funktionen lauten (mit $n = 4$):
Für die Exponentialverteilung

$$\ell_n(\mathbf{x};\lambda) = n\ln\lambda - \lambda\sum_{i=1}^{n} x_i.$$

Für die Pareto-Verteilung

$$\ell_n(\mathbf{x};\alpha) = n\ln\alpha - (\alpha+1)\sum_{i=1}^{n}\ln(1+x_i).$$

Mit den Ableitungen:
Für die Exponentialverteilung

$$\ell_n'(\mathbf{x};\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i, \quad \ell_n''(\mathbf{x};\lambda) = -\frac{n}{\lambda^2} < 0.$$

Für die Pareto-Verteilung

$$\ell_n'(\mathbf{x};\alpha) = \frac{n}{\alpha} - \sum_{i=1}^{n}\ln(1+x_i)\,, \quad \ell_n''(\mathbf{x};\alpha) = -\frac{n}{\alpha^2} < 0.$$

Nach Randwertbetrachtung ergeben sich die eindeutigen ML-Schätzer zu:

$$\hat{\lambda} = \frac{n}{\sum\limits_{i=1}^{n} x_i} = \frac{4}{3+10+6+1} = \frac{4}{20} = 0{,}2$$

für die Exponentialverteilung,

$$\hat{\alpha} = \frac{n}{\sum\limits_{i=1}^{n}\ln(1+x_i)} = \frac{n}{\ln\prod\limits_{i=1}^{n}(1+x_i)} = \frac{4}{\ln 616} = 0{,}6227$$

für die Pareto-Verteilung.

Die zugehörigen (geschätzten) Quantilfunktionen lauten:

$$Q_{Ex}(u) = -\frac{1}{\hat{\lambda}} \ln(1-u), \quad Q_{Par}(u) = \frac{1}{(1-u)^{1/\hat{\alpha}}} - 1 \qquad \text{für } 0 < u < 1.$$

Die zugehörigen Q-Q-Plots ergeben sich aus folgenden Tabellen:
Für die Exponentialverteilung:

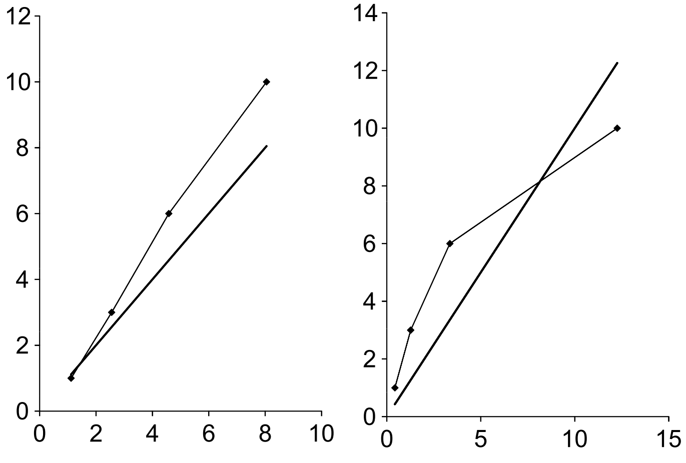
k	1	2	3	4
$\frac{k}{5}$	0,2	0,4	0,6	0,8
$Q_{Ex}\left(\frac{k}{5}\right)$	1,1157	2,5541	4,5815	8,0472
$x_{(k)}$	1	3	6	10

Für die Pareto-Verteilung:

k	1	2	3	4
$\frac{k}{5}$	0,2	0,4	0,6	0,8
$Q_{Par}\left(\frac{k}{5}\right)$	0,4310	1,2713	3,3557	12,2581
$x_{(k)}$	1	3	6	10

Zu c)

Daraus resultieren folgende Q-Q-Plots:



Beurteilung: linker Plot passt besser (lineare Form), rechter Plot hat deutlichen Knick.

Der linke Plot weist darauf hin, dass der Parameter $\hat{\hat{\lambda}}$ nach Q-Q-Plot etwas kleiner ausfällt (Regressionsgerade durch den Ursprung; Achtung: hier liegt eine Skalenfamilie in $\sigma = \frac{1}{\lambda}$ vor!); er ist hier gegeben durch

$$\hat{\hat{\lambda}} = \frac{\sum_{i=1}^{n} \ln^2\left(\frac{k}{n+1}\right)}{\sum_{i=1}^{n} \left(-\ln\left(\frac{k}{n+1}\right)\right) \cdot x_{(k)}} = 0{,}1602\,.$$

Es ist $\widehat{\mathrm{VaR}}_{0,005} = Q_{Ex}(0{,}995) = -\frac{1}{\hat{\lambda}} \ln(0{,}005) = 26{,}49$ und (wegen der Gedächtnislosigkeit der Exponentialverteilung) $\widehat{\mathrm{TVaR}}_{0,005} = \widehat{\mathrm{VaR}}_{0,005} + \frac{1}{\hat{\lambda}} = 31{,}69$. (Die bedingte Verteilung des Risikos $X - \mathrm{VaR}_\beta$ unter der Bedingung $X > \mathrm{VaR}_\beta$ ist dieselbe Exponentialverteilung!)

Aufgabe 2 (16 Punkte)

In einer Fabrik, die Metallwaren produziert, sollen Betriebsunterbrechungszeiten durch einen homogenen Markov-Prozess mit 2 Zuständen modelliert werden. Aus der Vergangenheit ist bekannt, dass der Betrieb durchschnittlich 3 Tage in Folge auf Grund von Störungen still liegt und die Dauer des unterbrechungsfreien Betriebs im Mittel 100 Tage beträgt.

a) Stellen Sie die Intensitätsmatrix des Prozesses auf.
b) Wie groß ist die Wahrscheinlichkeit, dass der Betrieb am 20. Januar einen Störfall hat, wenn er am Jahresanfang (zum 1.1.) störungsfrei begonnen hat?
c) Geben Sie die asymptotische Verteilung des Prozesses an. Ist diese stationär?

Lösung:

Zu a)

Zustandsraum $\mathcal{S} = \{0,1\}$ mit $0 =$ gestört, $1 =$ in Betrieb. Bei einem homogenen Markov-Prozess sind die Verweildauern jeweils exponentialverteilt (etwa $\mathcal{E}(\lambda)$ in Zustand 0, $\mathcal{E}(\nu)$ in Zustand 1). Aus den gemachten Angaben folgt: $\lambda = \frac{1}{3}$, $\nu = \frac{1}{100}$; die Intensitätsmatrix ist also gegeben durch

$$Q = \begin{bmatrix} -\lambda & \lambda \\ \nu & -\nu \end{bmatrix}$$

mit diesen Werten.

Zu b)

Hierzu wird die Übergangsmatrix

$$\Pi^t = e^{t\mathbf{Q}} = \frac{1}{\lambda+\nu} \begin{bmatrix} \nu + \lambda \cdot e^{-(\lambda+\nu)t} & \lambda \cdot \left(1 - e^{-(\lambda+\nu)t}\right) \\ \nu \cdot \left(1 - e^{-(\lambda+\nu)t}\right) & \lambda + \nu \cdot e^{-(\lambda+\nu)t} \end{bmatrix}$$

benötigt (Skript Teil I, Kapitel 3.4 Seite 98) mit $t = 20$. Die gesuchte Wahrscheinlichkeit beträgt dann

$$\frac{\nu}{\lambda+\nu}(1 - e^{-20(\lambda+\nu)}) = 0{,}029096\,,$$

also knapp 3%.
Zu c)

Es ist $\mathbf{p}^* = \left(\frac{\nu}{\lambda+\nu} \quad \frac{\lambda}{\lambda+\nu}\right) = (0{,}02913 \quad 0{,}97087)$. Dies ist zugleich asymptotische und stationäre Verteilung.

Aufgabe 3 (30 Punkte)

Gegeben: Schaden-Statistik nach 3 Tarif-Segmenten und 2 Vertriebswegen. Gegeben sind jeweils Vertragsbestände und Schadenanzahlen im Laufe des Jahres. Aus Erfahrung sei bekannt, dass pro Vertrag mehr als 1 Schaden auftreten kann, dass pro Vertrag die Schadenwahrscheinlichkeit proportional zur versicherten Zeit unter Risiko ist, und dass pro Vertrag gleichzeitig nicht mehr als 1 Schaden auftreten kann.

	Vertragsbestände		Schadenanzahlen	
	Vertriebsweg		Vertriebsweg	
Segment	1	2	1	2
1	100.000	200.000	3100	5800
2	100.000	200.000	2300	3800
3	10.000	400.000	680	23.200

a) Man berechne die relativen Schaden-Häufigkeiten und zeichne die Mittelwert-Plots.

b)
- Welche diskrete Verteilung ist zur Modellierung von Schadenhäufigkeiten naheliegend? Warum? (*Hinweis*: Es kommen nur Verteilungen aus der Panjer-Klasse in Frage!)
- Welche Struktur der Erwartungswerte liegt auf Basis der Erkenntnisse aus den Mittelwert-Plots nahe? Warum, qualitative Begründung?

c)
- Man gebe die (Zähl-)Dichte der in b) genannten Verteilung an.
- Man gebe die kanonische Link-Funktion der in b) genannten Verteilung, sowie den Erwartungswert-Parameter abhängig vom linearen Prädiktor an.
- Man gebe für die Parametrisierung mit Haupteffekten die Design-Matrix und eine übliche Darstellung für den Parameter Θ_{ij} der Verteilung in der Zelle ij an.
- Man gebe die Log-Likelihood des in b) genannten Modells abhängig von den Haupteffekten bei kanonischer Link-Funktion an.

d) Mit den Schätzern für μ_{ij} gemäß Tabelle:

	Vertriebsweg	
	1	2
Segment 1	6,1%	2,9%
Segment 2	4,2%	2,0%
Segment 3	13,7%	6,6%

sowie der Verteilungsannahme gemäß b) berechne man die Pearson-Residuen:

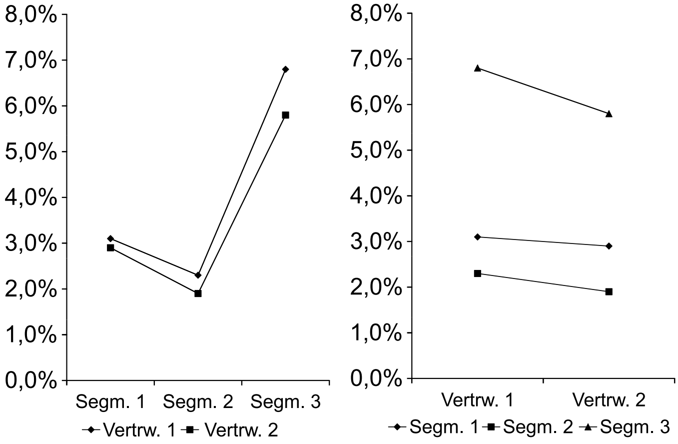
- Varianzfunktion
- Formel der Pearson-Residuen
- Ergebnis (Vertriebsweg 1 und Segment 1 genügt)

Lösung:

Zu a) Relative Häufigkeiten y_{ij}

	Vertriebsweg	
	1	2
Segment 1	3,1%	2,9%
Segment 2	2,3%	1,9%
Segment 3	6,8%	5,8%

MW-Plots



Zu b)

- Poissonverteilung. Diese erfüllt genau die angegebenen Eigenschaften.
- Modellierung mit Haupteffekten und mit einer Link-Funktion, die den Erwartungswert μ als Argument des ln enthält.
  Grund: In jeder der (beiden) Dimensionen verlaufen die MW-Plot weitgehend parallel (eigentlich: gehen durch Streckung in einander über; oder: die logarithmierten Werte verlaufen parallel).

Zu c)

- $\tilde{y}_{ij}$: Anzahl Schäden in Zelle ij, n_{ij}: Anzahl Verträge in Zelle ij, y_{ij}: relative Schadenhäufigkeit in Zelle ij mit (unbekanntem) Erwartungswert μ_{ij}.
  Poissonverteilung: $P(Y_{ij} = y_{ij}) = e^{-n_{ij}\cdot\mu_{ij}} \frac{(n_{ij}\cdot\mu_{ij})^{\tilde{y}_{ij}}}{\tilde{y}_{ij}!}$
- Kanonische Link-Funktion: ln: $\Theta_{ij} = \ln(\mu_{ij})$ (aus Tabelle 8.2.) Damit: $\mu_{ij} = e^{\Theta_{ij}}$
- Design-Matrix bei Parametrisierung von Haupteffekten:

$$X = \begin{pmatrix} 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta_1 \\ \beta_2 \end{pmatrix}$$

(in der 2. Darstellung spielen die β eine etwas andere Rolle!)
Jede der Komponenten von $X\beta$ lässt sich darstellen als $\Theta_{ij} = \alpha_i + \beta_j$ mit $i = 1, 2, 3$ und $j = 1,2$.

- Log-Likelihood:

$$
\begin{aligned}
l(y^T) &= \sum_{i=1}^{3}\sum_{j=1}^{2} \ln\left(e^{-n_{ij}\cdot e^{\alpha_i+\beta_j}}\,\frac{\left(n_{ij}\cdot e^{\alpha_i+\beta_j}\right)^{\tilde{y}_{ij}}}{\tilde{y}_{ij}!}\right)\\
&= \sum_{i=1}^{3}\sum_{j=1}^{2}\left(-n_{ij}\cdot e^{\alpha_i+\beta_j} + n_{ij}\cdot y_{ij}\cdot \ln\left(n_{ij}\cdot e^{\alpha_i+\beta_j}\right) - \ln(\tilde{y}_{ij}!)\right)\\
&= \text{Const} + \sum_{i=1}^{3}\sum_{j=1}^{2}\left(-n_{ij}\cdot a_i\cdot b_j + n_{ij}\cdot y_{ij}\cdot \ln(n_{ij}\cdot a_i\cdot b_j)\right)
\end{aligned}
$$

(mit: $a_i := e^{\alpha_i}$, $b_j := e^{\beta_j}$; beachte: $n_{ij}\cdot y_{ij} = \tilde{y}_{ij}$ Anzahl Schäden!).

Zu d)

- Varianzfunktion:

$$V(\mu_{ij}) = \mu_{ij} = a_i \cdot b_j\,.$$

- Formel:

$$P_{ij} = \frac{y_{ij}-\mu_{ij}}{\sqrt{\mu_{ij}/n_{ij}}} \quad \text{mit} \quad \mu_{ij} = a_i\cdot b_j\,.$$

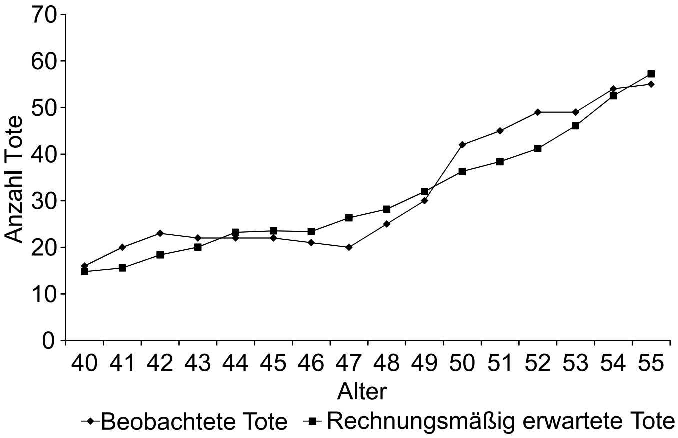
- Ergebnis: Tabelle der P_{ij}

$$P_{11} = \frac{3{,}1\% - 6{,}1\%}{\sqrt{6{,}1\%/100.000}} = -38{,}41\,.$$

Aufgabe 4 (20 Punkte)

Gegeben seien die folgenden Werte beobachteter und rechnungsmäßig erwarteter Toter:

Alter x	Bestand l_x	unterstellte Sterbewahrscheinlichkeiten q_x	Beobachtete Tote Z_x	Erwartete Tote $E_x = l_x\,q_x$
40	13.253	0,0011151	16	14,78
41	12.588	0,0012369	20	15,57
42	13.402	0,0013706	23	18,37
43	13.233	0,0015148	22	20,05
44	13.896	0,0016709	22	23,22
45	12.785	0,0018403	22	23,53
46	11.568	0,0020216	21	23,39
47	11.862	0,0022190	20	26,32
48	11.586	0,0024325	25	28,18
49	12.003	0,0026628	30	31,96
50	12.455	0,0029127	42	36,28
51	12.052	0,0031843	45	38,38
52	11.837	0,0034790	49	41,18
53	12.121	0,0038024	49	46,09
54	12.635	0,0041566	54	52,52
55	12.577	0,0045493	55	57,22



a) Überprüfen Sie mithilfe des

- Vorzeichentests,
- des Iterationstests,
- des χ^2-Tests,

ob die unterstellten Sterbewahrscheinlichkeiten bei einem Signifikanzniveau von $\alpha = 5\%$ angemessen erscheinen oder nicht.
Formulieren Sie zuerst die Null- bzw. Alternativhypothese. Geben Sie zu jedem Test die entsprechende Teststatistik und deren Wert an. Die kritischen Werte können Sie den nachfolgenden Tabellen entnehmen.
(*Hinweis*: Beim χ^2-Test ergibt sich als Wert der Teststatistik 8,97.)

Bin(16;0,5)

k	$F(k)$
0	0,00002
1	0,00026
2	0,00209
3	0,01064
4	0,03841
5	0,10506
6	0,22725
7	0,40181
8	0,59819
9	0,77275
10	0,89494
11	0,96159
12	0,98936
13	0,99791
14	0,99974
15	0,99998
16	1,00000

Bin(15;0,5)

k	$F(k)$
0	0,00003
1	0,00049
2	0,00369
3	0,01758
4	0,05923
5	0,15088
6	0,30362
7	0,50000
8	0,69638
9	0,84912
10	0,94077
11	0,98242
12	0,99631
13	0,99951
14	0,99997
15	1,00000

Quantile der χ^2-Verteilung

Anzahl Freiheitsgrade	15	16	17
95%-Quantil	24,996	26,296	27,587

b) Erläutern Sie die den Tests zugrunde liegenden Ideen und interpretieren Sie Ihre Testergebnisse aus a) entsprechend.

Lösung:

H$_0$: Die tatsächlichen und die rechnungsmäßig unterstellten Sterbewahrscheinlichkeiten stimmen überein.
H$_1$: Die tatsächlichen und die rechnungsmäßig unterstellten Sterbewahrscheinlichkeiten sind verschieden.

Zu a)

Vorzeichentest:
Die Teststatistik lautet hier $T = \sum_{x=40}^{55} 1_{\{Z_x > E_x\}}$.
Unter der Nullhypothese gilt $T \sim Bin(16, \frac{1}{2})$. Als kritische Werte ergeben sich daher mithilfe der obigen Tabelle $k = 4$ und $k = 12$ und damit der Verwerfungsbereich $B = \{0, 1, 2, 3\} \cup \{13, 14, 15, 16\}$.
Wie man dem Diagramm bzw. der Tabelle entnehmen kann, ergibt sich hier 9 als Wert der Teststatistik.
Da $9 \notin B$ ist, kann die Nullhypothese bei einem Signifikanzniveau von $\alpha = 5\%$ also nicht abgelehnt werden!

Iterationstest:
Die Teststatistik lautet hier $T = \sum_{x=41}^{55} 1_{\{Sign(Z_x - E_x) \neq Sign(Z_{x-1} - E_{x-1})\}}$, unter Gültigkeit der Nullhypothese gilt also $T \sim Bin(15, \frac{1}{2})$.
Als kritischer Wert ergibt sich nach obiger Tabelle $k = 4$ und als Verwerfungsbereich damit $B = \{0, 1, 2, 3\}$.
Wir zählen 3 Vorzeichenwechsel. Als Wert der Teststatistik ergibt sich daher $T = 3$.
Da $T \in B$ ist, muss die Nullhypothese bei einem Signifikanzniveau von $\alpha = 5$ % also abgelehnt werden!

χ^2-Test:
Die Teststatistik $T = \sum_{x=40}^{55} \frac{(Z_x - E_x)^2}{E_x}$ ist näherungsweise χ^2-verteilt mit 16 Freiheitsgraden.
Als Wert der Teststatistik erhält man $8{,}97 < 26{,}3 = \chi^2_{16;0,95}$.
Daher kann die Nullhypothese beim χ^2-Test bei einem Signifikanzniveau von $\alpha = 5$ % nicht abgelehnt werden!

Zu b)

Vorzeichentest:
Geht man davon aus, dass die tatsächlichen und die unterstellten Sterbewahrscheinlichkeiten übereinstimmen, dann kann man annehmen, dass bei den Differenzen zwischen beobachteten und rechnungsmäßig erwarteten Toten gleich viele positive wie negative Vorzeichen auftreten. Die Anzahl der beobachteten Toten liegt in unserem Fall 9mal oberhalb und 7mal unterhalb der rechnungsmäßig erwarteten Toten, d.h. es gibt fast so viele positive wie negative Vorzeichen. Daher kann die Nullhypothese hier nicht abgelehnt werden.

Iterationstest:
Wenn man davon ausgeht, dass die tatsächlichen und die unterstellten Sterbewahrscheinlichkeiten übereinstimmen, dann kann man folgern, dass die beobachteten Wahrscheinlichkeiten „mal größer und mal kleiner" sind, d.h. dass das Auftreten der Vorzeichen „plus" und „minus" gleich wahrscheinlich ist.
In unserem Fall haben wir aber nur 3 von möglichen 15 Vorzeichenwechseln gezählt. Auf dem 5%-Niveau ist das zu wenig. Daher muss die Nullhypothese abgelehnt werden.

χ^2-Test:
Die beobachteten Daten können nur dann eine Realisation der erwarteten Verteilung sein, wenn die Abweichungen nicht zu groß werden. Daher fließt bei der Berechnung des Werts der χ^2-Teststatistik die Größe der Abweichungen mit ein.
In unserem Fall liegt der Wert der Teststatistik unterhalb des 95%-Quantils der χ^2-Verteilung, d.h. die gemessenen Abweichungen zwischen beobachteten und erwarteten Toten widersprechen nicht der Nullhypothese.