



Aktuarielle Methoden der Tarifgestaltung in der Schaden-/ Unfallversicherung

Herausgeber:
DAV-Arbeitsgruppe Tarifierungsmethodik



Aktuarielle Methoden der Tarifgestaltung in der Schaden-/Unfallversicherung

Dr. Gerald Sussmann et al.

Der Vorstand der Deutschen Gesellschaft für Versicherungs- und Finanzmathematik e.V. erlaubt sich darauf hinzuweisen, dass der Verfasser jeder Abhandlung für die darin mitgeteilten Tatsachen und die ausgesprochenen Anschauungen allein verantwortlich ist.



DGVFM

DEUTSCHE GESELLSCHAFT
FÜR VERSICHERUNGS- UND
FINANZMATHEMATIK e.V.

Schriftenreihe

Versicherungs- und Finanzmathematik

Band 38

Aktuarielle Methoden der Tarifgestaltung
in der Schaden-/Unfallversicherung

Herausgeber:

DAV-Arbeitsgruppe Tarifierungsmethodik

Dr. Gerald Sussmann, Michael Buse,
Dr. Klaus Dräger, Dr. Christoph Dubowik,
Frank Ellgring, Thomas Franze, Kati Geisler,
Peter Gorontzy, Dr. Kai Haseloh, Rudolf Herter,
Dr. Olaf Kruse, Dr. Gero Nießen, Johannes Pohl-Grund,
Peter Schramm, Dr. Michael Schüte, Dr. Jörg Schult,
Axel Wolfstein



Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation
in der Deutschen Nationalbibliografie;
detaillierte bibliografische Daten sind im Internet über
<http://dnb.d-nb.de> abrufbar.

© 2011 Verlag Versicherungswirtschaft GmbH Karlsruhe

Das Werk einschließlich aller seiner Teile ist urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urhebergesetz zugelassen ist, bedarf der vorherigen Zustimmung des Verlags Versicherungswirtschaft GmbH, Karlsruhe. Jegliche unzulässige Nutzung des Werkes berechtigt den Verlag Versicherungswirtschaft GmbH zum Schadenersatz gegen den oder die jeweiligen Nutzer.

Bei jeder autorisierten Nutzung des Werkes ist die folgende Quellenangabe an branchenüblicher Stelle vorzunehmen:

© 2011 Verlag Versicherungswirtschaft GmbH Karlsruhe

Jegliche Nutzung ohne die Quellenangabe in der vorstehenden Form berechtigt den Verlag Versicherungswirtschaft GmbH zum Schadenersatz gegen den oder die jeweiligen Nutzer.

Herstellung printsystem GmbH Heimsheim

ISSN 1864-3779

ISBN 978-3-89952-607-3

Inhaltsverzeichnis

TABELLENVERZEICHNIS	IX
ABBILDUNGSVERZEICHNIS.....	XI
1 EINLEITUNG.....	1
2 DATEN	3
2.1 Einführung.....	3
2.1.1 Daten als Wirtschaftsgut	3
2.1.2 Vorgehensweise	4
2.1.3 Begriffsbestimmung.....	5
2.1.4 Aufgabenstellungen des Aktuars	6
2.2 Datenquellen	7
2.2.1 Interne Datenquellen	7
2.2.2 Externe Datenquellen.....	8
2.3 Erläuterung wichtiger Kenngrößen, Definitionen.....	10
2.3.1 Exposuremaße	10
2.3.2 Zielgrößen	14
2.3.3 Wertgrenzen und Inflation	20
2.3.4 Selbstbehalte und Deckungssummen.....	20
2.3.5 Schadenarten	21
2.4 Aufbereitung der Daten.....	22
2.4.1 Zeitliche Zuordnung von Prämien und Schäden.....	22
2.4.2 Umschlüsseln und Gruppieren	24
2.4.3 Behandlung von Großschäden und Kumulereignissen	26
2.4.4 Nullschäden.....	30
2.4.5 Abwicklungsstand.....	30
2.4.6 Probleme bei unvollständigen Daten, Imputation.....	33
2.4.6.1 Füllgrade.....	33
2.4.6.2 Fehlende oder zu kurze Datenhistorie.....	35
2.4.6.3 Fehlen einzelner Jahre in Zeitreihen	37
2.5 Prüfung der Daten	37
2.5.1 Methoden bei quantitativen Daten	37
2.5.1.1 Allgemein	37

2.5.1.2	Histogramme	38
2.5.1.3	Box-Plots	40
2.5.2	Q-Q-Plots	41
2.5.2.1	Prioritätenstatistiken	43
2.5.2.2	Mahalanobis-Distanz	44
2.5.3	Methoden bei qualitativen Daten	45
2.5.3.1	Allgemeines	45
2.5.3.2	Prüfung auf formale Korrektheit	46
2.5.3.3	Inhaltliche Prüfung	46
2.5.3.4	Zusammenfassung	50
2.6	Daten vor Anwendung statistischer Modelle	51
3	MODELLIERUNG, UMSETZUNG, TECHNIK	53
3.1	Univariate Analysen	54
3.2	Gewichtungsverfahren	56
3.3	Ausgleichsverfahren	59
3.4	Verallgemeinerte Lineare Modelle (GLM)	61
3.4.1	Modellformulierung	61
3.4.2	Devianz und verallgemeinerte χ^2 -Statistik	63
3.4.3	Auswahl der Verteilungsannahme	65
3.4.4	Overdispersion	67
3.4.5	Power-Varianz- und Power-Link-Funktionen	67
3.4.6	Quasi-Likelihood	69
3.4.7	Parametrisierung und Merkmalsschachtelung	70
3.4.8	Zielgröße in der Tarifikalkulation	72
3.4.9	Inferenzanalyse und Signifikanztests	73
3.4.10	Merkmalsauswahl und Modellanpassung	77
3.4.11	Modelldiagnose	82
3.4.12	Berücksichtigung von a-priori-Informationen	85
3.4.13	Systematische Nullschadenbedarfe	90
3.5	Credibility-Verfahren	90
3.6	Geografische Glättungsverfahren	92
3.7	Clusterverfahren	94

4	PRAKTISCHE HINWEISE ZUR RISIKOMODELLERSTELLUNG	97
4.1	Generelles.....	97
4.2	Erster Überblick über die Daten	97
4.3	Kriterien zur Auswahl von Risikomerkmale	102
4.4	Aussagekraft statistischer Tests	106
4.5	Abhängigkeiten im Bestand (er)kennen	108
4.6	Gruppieren von Ausprägungen.....	111
4.7	Untersuchung von Zeit- und Zufallskonsistenz.....	116
4.8	Interaktionen und Korrelationen.....	119
4.9	Informationen aus Verbänden und Datenpools.....	127
4.9.1	Übernahme einer Tarifstruktur	127
4.9.2	Übernahme eines Tarifniveaus	130
4.10	Klassifikationsverfahren	130
4.11	Struktur der Modelle in den Tarifierungsphasen	133
5	GLOBALES NIVEAU	137
5.1	Ziel.....	137
5.2	Niveaubestimmung im Risikomodell	139
5.2.1	Globales Niveau als Teil von Regressionsansätzen.....	139
5.2.2	Behebung nicht gegebener Repräsentativität.....	141
5.2.2.1	Sachliche Repräsentativität	141
5.2.2.2	Zeitliche Repräsentativität.....	142
5.2.3	Abwicklung und Diskontierung.....	144
5.2.3.1	Abwicklung	144
5.2.3.2	Diskontierung	146
5.2.3.3	HGB-Reservierung und Barwert des Schadens	147
5.2.4	Großschäden und Rückversicherung	149
5.2.5	NatCat	151
5.2.5.1	Einsatz von Naturgefahrenmodellen in der Tarifierung	157

5.2.5.2	Kombination von Erfahrungs- und Exposuretarifizierung	161
5.2.6	Trends.....	163
5.2.7	Abstimmung zwischen Tarifizierung und Reservierung	166
5.3	Übergang vom Risikomodell zum Tarifmodell.....	167
5.3.1	Kostenzuschläge.....	167
5.3.2	Kapitalkosten, Sicherheits- und Schwankungszuschlag.....	169
5.4	Übergang vom Tarifmodell zum Tarifbuch.....	171
5.5	Berücksichtigung von Rückversicherung.....	174
5.6	Beispiel	177
5.6.1	Grundlagen.....	177
5.6.2	Von der Kalkulationsstatistik zum Risikomodell	178
5.6.3	Vom Risikomodell zum Tarifmodell	181
5.6.4	Vom Tarifmodell zum Tarifbuch.....	182
6	EINBINDUNG IN DIE UNTERNEHMENSPROZESSE	183
6.1	Control Cycle.....	183
6.2	Dokumentation.....	185
6.2.1	Vorgehensdokumentation	185
6.2.2	Ergebnisdokumentation	187
6.3	Einschränkungen bezüglich Zeit/Ressourcen/Daten.....	190
6.4	Abgrenzung zwischen Erst- und Rückversicherung.....	191
6.4.1	Beitragskalkulation in der Rückversicherung.....	192
6.4.2	Beitragskalkulation in der Erstversicherung.....	194
	LITERATURANGABEN UND REFERENZEN	195
	STICHWORTVERZEICHNIS	199

Tabellenverzeichnis

Tabelle 1:	Zusammenhang zwischen Exposuremaß und Zielgröße	19
Tabelle 2:	Schadenkennzahlen für verschiedene Versicherungssparten.....	19
Tabelle 3:	Schadensatz in der Sturmversicherung.....	36
Tabelle 4:	Beispiel für Prioritätenstatistik.....	43
Tabelle 5:	Datenstruktur in KFZ-Bestand, Ausschnitt.....	51
Tabelle 6:	Datenstruktur in Haftpflicht-Bestand, Ausschnitt.....	52
Tabelle 7:	Datenbeispiel für univariate Datenanalyse: Jahreseinheiten, Schadenbedarf	55
Tabelle 8:	Datenbeispiel für Gewichtungsverfahren	57
Tabelle 9:	Varianzfunktion und kanonische Linkfunktion wichtiger GLM.....	63
Tabelle 10:	Devianz-Funktionen	64
Tabelle 11:	GLM mit Power-Link- bzw. Power-Varianz-Funktionen.....	68
Tabelle 12:	Power-Varianz-Funktionen: Anpassungsfaktoren $\xi^{(p)}$	88
Tabelle 13:	Verteilung und Schadeninformation nach den Merkmalen Alter und Geschlecht.....	119
Tabelle 14:	Verteilung und Schadeninformation nach den Merkmalen Alter und Geschlecht.....	121
Tabelle 15:	Typisches Eingabedatenformat für ein aggregiertes Modell....	152
Tabelle 16:	Typisches Eingabedatenformat für ein detailliertes Modell (Policeninformationen).....	153
Tabelle 17:	Typisches Eingabedatenformat für ein detailliertes Modell (Standortinformationen).....	153
Tabelle 18:	Event Loss Table	154
Tabelle 19:	Ausgabe von μ und σ aus dem Naturkatastrophenmodell	159
Tabelle 20:	Faktoren für die Statistikjahre in den Teilmodellen für Personen- und Sachschäden.....	179
Tabelle 21:	Kostenzuschläge im Tarifmodell.....	181

Abbildungsverzeichnis

Abbildung 1: Beispiel eines äquidistanten Bandings für das Merkmal "Bevölkerungsdichte"	25
Abbildung 2: Entwicklung des Anfalljahres 2007.....	32
Abbildung 3: Histogramm: Verteilung von PH-Sachschäden bis 1.000 € Schadenhöhe auf einzelnen Schadenklassen	39
Abbildung 4: Box-Plot.....	40
Abbildung 5: Q-Q-Plot für die Exponentialverteilung.....	42
Abbildung 6: Q-Q-Plot für die Gammaverteilung.....	42
Abbildung 7: Verteilung der Jahreseinheiten über alle SF-Stufen in KH.....	47
Abbildung 8: SF-Indizes für Schadenbedarfe in Kraftfahrt.....	48
Abbildung 9: Schadenbedarfsabstand Fahrleistung zur Fahrleistungsklasse „km-Leistung bis 6000 km“.....	49
Abbildung 10: Devianz-Profile für zwei Link-Funktionen und dreidimensionales Devianz-Profil.....	69
Abbildung 11: Residuenanalyse: Verteilung der Devianzresiduen.....	83
Abbildung 12: Residuenanalyse: Streudiagramm der Devianzresiduen.....	84
Abbildung 13: Residuenanalyse: Streudiagramm der Devianzresiduen.....	84
Abbildung 14: Vertragsdauer als kategoriales Merkmal vs. Glättung durch Polynom.....	89
Abbildung 15: Beispiel auf Postleitzahlebene: Rohdaten (links) mit $a=20$ ausgeglichen (rechts).....	93
Abbildung 16: Schadenhöhenverteilung von Glasschäden in der Kasko-Versicherung	98
Abbildung 17: Schadenhöhenverteilung von Glasschäden in der Kasko-Versicherung für zwei unterschiedliche Selbstbehaltsvarianten.....	99
Abbildung 18: Jeweils auf 100% normierte Exposureverteilung des Merkmals "Fahrzeughalter", aufgesplittet nach den einzelnen Statistikjahren	100
Abbildung 19: Beobachtete Frequenz für Fahrraddiebstahl je Fahrradzone	101

Abbildung 20: Beobachtete Frequenz für Fahrraddiebstahl, unterschieden nach Wohneigentum-Verhältnis	101
Abbildung 21: Modell 1 (ohne das Merkmal „Kaskoanbindung“).....	102
Abbildung 22: Modell 2 (inklusive Merkmal "Kaskoanbindung")	103
Abbildung 23: Prüfung des Merkmals "Kaskoanbindung" auf Zeitkonsistenz	104
Abbildung 24: Prüfung des Merkmals "Kaskoanbindung" auf Zufallskonsistenz	105
Abbildung 25: Beispiel für ein nicht zeitkonsistentes Merkmal	107
Abbildung 26: Beispiel für ein nicht zufallskonsistentes Merkmal	107
Abbildung 27: Beobachtete Frequenz für Fahrraddiebstahl nach Alter des VN	108
Abbildung 28: Beobachtete Frequenz für Fahrraddiebstahl nach Zahlungsperiode.....	109
Abbildung 29: Beobachtete Frequenz sowie Modellvorhersage für Fahrraddiebstahl nach Zahlungsperiode	109
Abbildung 30: Abhängigkeiten im Bestand, hier: Verteilung der Zahlungsperiode je Alter des VN.....	110
Abbildung 31: Beobachtete Frequenz für Privathaftpflichtschäden nach Alter des VN	112
Abbildung 32: Test auf Zeitkonsistenz der Schätzer für Alter des VN.....	113
Abbildung 33: Test auf Zufallskonsistenz der Schätzer für Alter des VN.....	113
Abbildung 34: Verwendung eines Polynoms anstelle der einzelnen Schätzer je Ausprägung.....	114
Abbildung 35: Das Polynom bleibt innerhalb von zwei Standardfehlern um den "Einzelschätzer" herum	116
Abbildung 36: Frequenz von Glasschäden in der Kasko-Versicherung nach Alter des Kfz	117
Abbildung 37: Frequenz von Glasschäden in der Kasko-Versicherung nach Alter des Kfz inklusive einer möglichen Glättung	117
Abbildung 38: Überprüfung auf Zeitkonsistenz	118
Abbildung 39: Überprüfung auf Zufallskonsistenz	118
Abbildung 40: Visualisierung der Verteilung.....	120
Abbildung 41: Visualisierung der Schadeninformation.....	120

Abbildung 42: Visualisierung der Verteilung.....	121
Abbildung 43: Visualisierung der Schadeninformation.....	122
Abbildung 44: Interaktion Zahlungsperiode x Zahlungsart.....	123
Abbildung 45: Interaktion Zahlungsperiode x Zahlungsart relativ zu einem Basisniveau (hier: Lastschrift)	124
Abbildung 46: Interaktion Zahlungsperiode x Zahlungsart relativ zu einem Basisniveau (hier: Lastschrift) für das Jahr 2006.....	125
Abbildung 47: Interaktion Zahlungsperiode x Zahlungsart relativ zu einem Basisniveau (hier: Lastschrift) für das Jahr 2007.....	126
Abbildung 48: Interaktion Zahlungsperiode x Zahlungsart relativ zu einem Basisniveau (hier: Lastschrift) für das Jahr 2008.....	126
Abbildung 49: Zinserträge und Barwert in Entwicklungsjahren.....	148

1 Einleitung

Die folgende Ausarbeitung befasst sich mit der Tarifgestaltung von Erstversicherungsunternehmen der Schaden-/Unfallversicherung speziell im Breitengeschäft. Sie richtet sich an die dort praktisch tätigen Aktuar und setzt die Kenntnis der wesentlichen Inhalte des aktuariellen Grundwissens, speziell hinsichtlich Schadenversicherungsmathematik und statistischer Methoden, voraus. Es wird weder der Anspruch auf Verbindlichkeit oder Vollständigkeit erhoben, noch sollte dieser Text als abgeschlossenes Werk angesehen werden.

Grundlage des folgenden Textes ist der DAV-Hinweis „Berufspflichten des Aktuars in der Tarifgestaltung“, verabschiedet am 27.6.2007. Darin wird abstrakt auf die Verantwortung des Aktuars in Abgrenzung und Zusammenarbeit mit anderen an der Tarifgestaltung tätigen Funktionen abgestellt. Konkrete Ansätze und Methoden werden dort nicht diskutiert. Im vorliegenden Text werden mathematische und statistische Verfahren zusammengestellt, die im weitesten Sinne zur Erstellung von statistischen Auswertungen in der Tarifikalkulation relevant sind bzw. relevant werden könnten. Es werden übliche Datenkonstellationen und Ansätze zur Konzeption von Analysen beschrieben. Somit ist diese Ausarbeitung in Konkretisierung des oben erwähnten Hinweises als ein Methoden-Handbuch für den Aktuar der Schaden-/Unfallversicherung zu betrachten. Die Aspekte der Schaden-/Unfallversicherung, die nach Grundsätzen der Personenversicherung kalkuliert werden, werden hier nicht betrachtet.

In der praktischen Tätigkeit stellt der Aktuar häufig fest, dass es gerade im Bereich der deskriptiven und explorativen Statistik mehrere zum Ziel führende Wege gibt. Die Wahl der verwendeten Methode ist stets abzustellen auf die konkreten Rahmenbedingungen. Dennoch wurde im vorliegenden Text versucht, konkrete Vorgehensweisen für die typischen aktuariellen Problemstellungen herauszuarbeiten, um den Lesern eine unverbindliche Entscheidungshilfe an die Hand zu geben. In diesem Sinne wird dem Aktuar mit diesem Text ein Werkzeugkasten an die Hand gegeben, um auftretende Probleme adäquat angehen zu können. In keinem Fall darf ein Nutzer die hier vorgestellten Ansätze verwenden, ohne sich über deren Anwendbarkeit klar geworden zu sein.

Obwohl das Handbuch elementar gehalten wurde, erleichtert eine gewisse Vertrautheit mit grundlegenden Erkenntnissen der Statistik das Verständnis.

Methodisch gliedert sich das Handbuch in fünf Teile: In Kapitel 2 werden übliche Datenkonstellationen sowie die zur Tarifierung notwendigen typischen datentechnischen Vorarbeiten beschrieben. In Kapitel 3 wird auf statistische Methoden eingegangen, die im Rahmen der Tarifikalkulation eingesetzt werden. Besonderes Gewicht wird dabei auf die Verallgemeinerten Linearen Modelle als modernes statistisches Verfahren zur Tarifikalkulation gelegt. Die praktischen Aspekte bei der Erstellung eines Risikomodells werden in Kapitel 4 dargestellt. Darüber hinaus fokussiert dieser Abschnitt auch auf Überlegungen zur Konzeption der Struktur der Auswertungen. In Kapitel 5 werden unterschiedliche Aspekte zur Bestimmung eines angemessenen Tarifniveaus diskutiert. Die Einbindung der aktuariellen Tarifierungsarbeit in die Abläufe des Unternehmens wird abschließend in Kapitel 6 beleuchtet. Neben der Beschreibung des aktuariellen Control Cycles gibt dieser Teil praktische Hinweise für die Dokumentation und für den Umgang mit den in der Praxis regelmäßig auftretenden Restriktionen. Den Abschluss bildet eine Abgrenzung der Tarifierung in der Erstversicherung von der Preisgestaltung in der Rückversicherung.

2 Daten

2.1 Einführung

2.1.1 Daten als Wirtschaftsgut

Ein wesentlicher Bestandteil der aktuariellen Tarifentwicklung ist es, aus Daten Informationen zu gewinnen. Bei der Beschreibung dieses Prozesses wird – auch in der Fachliteratur – üblicherweise der Fokus mehr auf technische bzw. statistische Fragestellungen als auf Fragen der Datenqualität gelegt. Dabei ist es unbestritten, dass in der Praxis der wesentliche Aufwand bei der Datensammlung bzw. -aufbereitung und nicht unbedingt bei der statistischen Analyse anfällt. Der Einsatz immer komplexerer und fortgeschrittener statistischer Verfahren führt zu stetig steigenden Anforderungen an Umfang und Qualität von Daten.

Mehr als jemals zuvor haben Versicherer die Möglichkeit, auf detaillierte Kundeninformation nicht nur aus ihren Beständen zuzugreifen. Der Umgang mit bzw. Zugriff auf diese Informationen hat sich in den letzten Jahren aufgrund der rasanten Entwicklung der DV-Technologie stark gewandelt. Zusätzlich hat sich aufgrund des Wettbewerbes die Notwendigkeit der einzelnen Unternehmen erhöht, ihre eigenen internen Informationen bei der Tarifikalkulation optimal auszuwerten.

Auch Gesetzesnovellen – sowohl in Europa (KonTraG/Solvency II) als auch in den USA (Sarbanes-Oxley) –, die sich u.a. mit der Transparenz und Dokumentation innerbetrieblicher Abläufe befassen, sollten dazu führen, dem Thema der Datenqualität größere Aufmerksamkeit zu widmen. Der Ansatz, Daten als Wirtschaftsgut zu betrachten, liegt also auf der Hand. Dass nicht erkannte Datenprobleme ernsthafte wirtschaftliche Folgen haben können, wird an den folgenden beiden Beispielen deutlich:

Die NASA verlor 1999 eine 100 Millionen € teure Mars-Sonde, als sie zu tief in das Schwerkraftfeld des Planeten eintauchte. Ein beteiligtes Projektteam hatte Entfernungsbestimmungen nicht im metrischen sondern im angelsächsischen Maßsystem durchgeführt.

Als bislang größter Konkurs eines britischen Versicherers musste 2001 „The Independent“ seinen Geschäftsbetrieb einstellen. Der Zusammenbruch ereignete sich nach einer missglückten Kapitalerhöhung aufgrund von Enthüllungen, dass die Firma nicht quantifizierbare Verluste aus Kundenforderungen, die nicht oder fehlerhaft bilanziell verbucht worden waren, zu erwarten hatte.

Die Robustheit der aktuariellen Einschätzung steht und fällt mit der zugrunde liegenden Datenqualität. Da der Aktuar in der Regel auf nicht von ihm selbst erhobene Daten zurückgreifen wird, ergibt sich für ihn die Aufgabe, eine Qualitätssicherung auf Vollständigkeit, Anwendbarkeit und Richtigkeit vorzunehmen und die sich daraus ergebende Unsicherheit in seiner Arbeit zu berücksichtigen. Innerhalb (und auch außerhalb, insbesondere bei Rückversicherern, welche vielfach auf externen Daten operieren müssen) seines Unternehmens ist er gefordert, auf die Auswirkungen schlechter Daten hinzuweisen und Anforderungen an die Dateninfrastruktur, -qualität und -integrität zu formulieren und an zuständige Stellen weiterzuleiten.

2.1.2 Vorgehensweise

Die Ausarbeitung orientiert sich in Teilen an [10], [11] und [9]. „Datenqualität“ lässt sich auf folgende Themenbereiche herunter brechen:

- Die Bedeutung der Datenerhebung und -verarbeitung
- Prüfung von Daten in Bezug auf Verwendbarkeit, Genauigkeit und Vollständigkeit
- Bewertung der Qualität der Daten
- Abschätzung der Auswirkung von fehlerhaften oder unvollständigen Daten
- Zuständigkeiten für diesen Prozess

Da Aktuare häufig gefordert sind, mit unvollständigen und vielfach fehlerbehafteten Daten zu arbeiten, werden in den folgenden Kapiteln verschiedene Ansätze und Methoden beschrieben, um die angesprochenen Aspekte des Prozesses der Bewertung und Sicherstellung der Datenqualität

zu unterstützen. Aufgrund der Breite und Komplexität des Themas können nur Hinweise und Anregungen gegeben werden. Entsprechend wird auch kein Anspruch auf Vollständigkeit erhoben. Handlungsschemata müssen in Abhängigkeit von der jeweiligen Problemstellung erarbeitet werden. Hier soll vor allem ein stärkeres Bewusstsein für diese Thematik geweckt werden.

2.1.3 Begriffsbestimmung

Eine effektive Bewertung der Datenqualität bedeutet die Untersuchung der Daten im Hinblick auf

- Gültigkeit
- Genauigkeit (absolute ~/effektive ~/relative ~)
- Repräsentativität und
- Vollständigkeit

Unter „Gültigkeit“ wird die Einhaltung eines definierten Wertebereichs verstanden. Der Wertebereich deutscher fünfstelliger Postleitzahlen ist z.B. auf ca. 8.400 Ausprägungen begrenzt. So sind „00000“ und „99999“ keine gültigen Postleitzahlen.

Die Begriffe „Genauigkeit“ und „Vollständigkeit“ entsprechen weitgehend ihrer umgangssprachlichen Bedeutung. „Vollständigkeit“ adressiert aber nicht nur das Thema fehlender Daten, sondern auch die Frage, ob ein Datensatz genau einmal und nicht z.B. mehrfach verarbeitet wird.

Die „absolut genaue“ Postleitzahl der DAV-Geschäftsstelle in Köln ist „50674“ und nicht „50931“, auch wenn „50931“ in unmittelbarer Nähe von „50674“ liegt und z.B. für die Frage der Zuordnung zu einem Regierungsbezirk „effektiv genau“ wäre. „Relativ genau“ ist eine Information, die konsistent falsch erhoben wurde, so dass die aus dieser Information abgeleiteten Aussagen nur unwesentlich beeinträchtigt werden. Werden von den Postleitzahlen systematisch nur die ersten drei Stellen erfasst, ergeben sich nur geringe Fehler bei der Zuordnung auf Regierungsbezirke.

Unter Repräsentativität wird die Frage nach der Übereinstimmung von Datenstrukturen mit erwarteten oder bekannten/historischen Mustern zusammengefasst. Ist ein Sturmereignis wie Kyrill in 2007 repräsentativ für die Kasko-Versicherung bzw. Wohngebäudeversicherung oder stellt es eher ein Ausnahmeereignis dar? Ist die Ansammlung von Großschäden in einem Zeitsegment zufällig auf ein externes Ereignis zurückzuführen oder evtl. ein Anzeichen von Codierungs-Problemen? Im Gegensatz zu den erst genannten Punkten werden hier nicht Einzelsätze sondern aggregierte Datenbestände analysiert.

2.1.4 Aufgabenstellungen des Aktuars

Fast alle in aktuariellen Studien verwendeten Daten werden schon während der Erhebung und Weiterverarbeitung routinemäßig verschiedensten Prüfungen unterzogen. Dem Aktuar obliegt nun die Entscheidung, ob die Daten gut genug im Sinne der definierten Kriterien oder ob weiterführende Prüfungen notwendig sind. Kurz gesagt wird die Frage beantwortet „Sind die Daten adäquat?“.

Eine gute Dokumentation beginnt mit einem erfahrungsbasierten Urteil und ergänzt es mit objektiven Fakten. Zum Beispiel wird ein Vergleich mit historischen Daten – z.B. aus einer gleichgelagerten Analyse – immer einen sinnvollen Ausgangspunkt darstellen.

Die Struktur und der Umfang einer Dokumentation basiert auf den spezifischen Umständen: Dem Verwendungszweck der Daten, dem vorhandenen Datenvolumen, dem Vorliegen bereits bekannter Datenbeschränkungen, dem vorhandenen Zeitrahmen und anderen Faktoren. Sie sollte

- Umfang der Prüfung und ggf. notwendige Dateneingriffe beschreiben,
- fragwürdige Datenstrukturen oder Inkonsistenzen aufzeigen und eine
- Abschätzung der Auswirkung von fehlerhaften/unvollständigen Daten treffen.

Vor allem bei dem letzten Punkt sind die Expertise und die Erfahrung des Aktuars gefragt. Eine Dokumentation könnte nach folgenden Punkten strukturiert werden:

- Verwendete interne Datenquellen
- Verwendete externe Datenquellen
- Datenverzerrung aufgrund unvollständiger Daten
- Durchgeführte Dateneingriffe bzw. -anpassungen
- Mögliche Einschränkungen in Bezug auf die Analyse
- Ungelöste Probleme/Bedenken in Bezug auf die Datenqualität

Um die Datenqualität nachhaltig zu sichern liegt es im Interesse des Aktuars, auf die Implementierung eines (revisionssicheren) Prozesses zur Erstellung, Aufbereitung und Sicherung der Daten hinzuwirken. Damit wird seine Aufgabe, die Qualität und Vergleichbarkeit der Daten zu beurteilen, wesentlich erleichtert. Gleichzeitig liegt in einem solchen Prozess eine Beschreibung der Daten vor, die der Aktuar für die Dokumentation seiner Arbeit ohnehin benötigt.

Für weitere Informationen und auch Modelle im Sinne eines formalisierteren Datenqualitätsberichts sei auf [26] verwiesen. Hier finden sich auch viele andere interessante Informationsquellen und Querverweise.

2.2 Datenquellen

Grundsätzlich sollte ein besonderes Augenmerk auf die Art und Weise des Zustandekommens der Daten – auch mit Blick auf ihren Verwendungszweck – geworfen werden.

2.2.1 Interne Datenquellen

Unter internen Daten (im primären Sinn der jeweiligen Sparten, aber im sekundären Sinn auch aus anderen Sparten sowie ggf. im Konzern vorhandene) seien solche verstanden, die durch und während des Versicherungsvertrages erhoben wurden und folglich dem VU im direkten