Fabian Peter Pribahsnik, BSc.

Let the data tell us their story
Machine learning with insurance policies

Master's Thesis

to achieve the university degree of

Master of Science

Financial and Actuarial Mathematics

submitted to

Vienna University of Technology

Institute of Statistics and Mathematical Methods in Economics

Supervisor

Univ.Prof. Dipl.-Math. Dr.rer.nat. Thorsten Rheinländer

Vienna, October 2017

Affidavit

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used.

Date Signature

Abstract

Insurance companies are facing a huge amount of regulations, including various guidelines addressing forecast scenario calculations for the policies in the portfolio. Taking the hundreds of thousands policies into account an average insurance company has in its portfolio on can easily see that these scenario calculations are very time consuming. Due to the rising number of policies and the very tight time schedule introduced with Solvency II insurance companies are looking for ways to reduce the computational time significantly. In the past years different approaches were developed and already used for grouping similar policies together and therefore reducing the computation time. The currently used algorithms are ranging from just grouping policies with exactly the same attributes together to basic cluster algorithms like k-means. This work highlights potential problems with the algorithms currently used and tries to implement some machine learning techniques to accomplish the task of grouping.

Contents

List of Figures

Todo list

1. Introduction

Solvency II - entered into force on 1 January 2016 - is the new European framework for a common insurance supervision. It is intended to achieve a harmonization of the European insurance sector and was implemented in accordance with the Lamfalussy architecture which works on a 4 level basis [3]. The most significant elements and aims of the new regulation framework can be studied on the homepage of the financial market authority (FMA) [8] and on the homepage of the European Insurance and Occupational Pensions Authority (EIOPA) [6]. This work is intended not to cover all aspects and aims of the new Solvency II regulation framework but focuses on the topic of data quality regarding to the actuarial function. In order to meet all the requirements imposed by Solvency II, insurance companies need to process large amounts of data within a short period. One critical aspect of these calculations is the projection horizon which however should cover the full lifetime of all obligations as stated in [1]:

3.83.

> The projection horizon used in the calculation of best estimate should cover the full lifetime of all obligations related to existing insurance and reinsurance contracts on the date of the valuation.

3.84.

> The determination of the lifetime of insurance and reinsurance obligations shall be based on up-to-date and credible information and realistic assumptions about when the existing insurance and reinsurance obligations will be discharged or cancelled or expired.

Another aspect needed to be considered is the fact that cash flow calculations need to be done for a variety of different economic scenarios which yields to an enormous computational effort. Due to the tight time schedule, insurance companies are looking for new possibilities to speed up these time consuming calculations. One approach is not to make all these calculations on per policy level, but on a grouped level where similar policies are grouped together and represented by only a few policies. This approach raises the question of how to maintain data quality as mentioned in the level 1 directive [13] while reducing the number of policies.

> *Article 82*
>
> **Data quality and application of approximations, including case-by-case approaches, for technical provisions**
>
> Member States shall ensure that insurance and reinsurance undertakings have internal processes and procedures in place to ensure the appropriateness, completeness and accuracy of the data used in the calculation of their technical provisions...

By publishing the level 2 regulations, supplementing the level 1 directive [13] the European Commission is getting more specific on data quality (Article 19 in [2]) and also formulates concrete requirements for grouped policies[2].

> *Article 35*
>
> **Homogeneous risk groups of life insurance obligations**
>
> The cash flow projections used in the calculation of best estimates for life insurance obligations shall be made separately for each policy. Where the separate calculation for each policy would be an undue burden on the insurance or reinsurance undertaking, it may carry out the projection by grouping policies, provided that the grouping complies with all of the following requirements:
>
> *a*) there are no significant differences in the nature and complexity of the risks underlying the policies that belong to the same group;
>
> *b*) the grouping of policies does not misrepresent the risk underlying the policies and does not misstate their expenses;

c) the grouping of policies is likely to give approximately the same results for the best estimate calculation as a calculation on a per policy basis, in particular in relation to financial guarantees and contractual options included in the policies.

These level 2 regulations are a reference point on what to consider when grouping policies together and they are even further specified in the level 3 guidelines issued by EIOPA[5]. Further details on the level 3 guidelines including feedback statements to the consultation paper (EIOPACP-14/036) and the guidelines can be obtained from [4].

This thesis reviews the currently used technique for grouping policies together and introduces furthermore some new approaches on how life insurance policies can be grouped together. We will therefore highlight drawbacks and advantages with a special emphasis on the regulatory requirements of every approach discussed. Theoretical considerations as well as practical implementations and tests with real world data will provide us some information on which method an insurance company should work with in order to obtain the best grouping results.

This thesis is structured as follows: First we discuss how sensitive main characteristics of a policy are with respect to the interest rate, the age or the duration. We therefore make a sensitivity analysis with real world data and a widely used projection tool . In the next chapter we introduce the currently used unsupervised learning algorithm k-means and discuss derive some theoretical findings.

Weitere Details für jedes Kapitel folgen.

2. Sensitivities

Grouping together single policies and representing them by just a few representative ones always comes along with a loss of information. A natural question which arises when grouping insurance contracts together is how to determine the main characteristics of the new representative policy. Some characteristics should for technical reasons be defined as the sum of the individual ones, like the sum insured, the premium or the accumulated reserve. This is needed to guarantee the equality between the un-grouped and the grouped portfolio in terms of these characteristics at the beginning of the projection horizon. For other characteristics like the age, the duration or the gender it is not intuitively clear how they should be defined for a representative policy. Possible solutions which can be implemented easily range from taking the weighted average over the value with the highest relative frequency to just taking the median of the grouped policies. Another difficulty which arises when grouping together policies from different product generations is, how the technical interest rate of the representative contract should be defined. Even if the policies are identical in terms of age, sex, sum insured, duration, costs,... and just differ on their issue date the huge possible differences with respect to the technical interest rate as shown in table 2.1 can have enormous impacts on the projected cash flows. Already a relative small difference in the technical interest of only one percent causes double digit differences in the guaranteed capital after 1 decade.

31.12. 1994	30.06. 2000	31.12. 2003	31.12. 2005	31.03. 2011	20.12. 2012	31.12. 2014	31.12. 2015	31.12. 2016
4%	3.25%	2.75%	2.25%	2%	1.75%	1.5%	1%	0.5%

Table 2.1.: Maximum technical interest rates for life insurance contracts issued after the given dates. (see [7])

It is therefore important to know how sensitive the different output variables of interest which are calculated by the projection tool react if various input parameters are changed slightly. The most basic task is to determine whether the correlation between the input and output variables is positive or negative. There are many output variables which can be important for determining if a grouping process has been successful in terms of accuracy or not, but in the subsequent we will focus only on a few of them, namely the premium, the present value of future profits at time 0, the reserve and the yearly claims. Due to the big variety of different insurance products we will just give some general guidelines based on the most important input variables and encourage the reader to do a sensitivity analysis for the specific portfolio of interest on his own. Most of the life insurance contracts can be built up by the following elementary insurance types[1]:

$$A_x = \sum_{k=0}^{\infty} v^{k+1} {}_kp_x q_{x+k} \qquad \text{(Whole life insurance)} \tag{2.1}$$

$$A^1_{x:\overline{n}|} = \sum_{k=0}^{n-1} v^{k+1} {}_kp_x q_{x+k} \qquad \text{(Term insurance)} \tag{2.2}$$

$$A_{x:\overline{n}|} = \sum_{k=0}^{n-1} v^{k+1} {}_kp_x q_{x+k} + v^n {}_np_x \qquad \text{(Endowment)} \tag{2.3}$$

$$\ddot{a}_x = \sum_{k=0}^{\infty} v^k {}_kp_x \qquad \text{(Whole life annuity)} \tag{2.4}$$

$$\ddot{a}_{x:\overline{n}|} = \sum_{k=0}^{n-1} \ddot{a}_{\overline{k+1}|} {}_kp_x q_{x+k} + \ddot{a}_{\overline{n}|} {}_np_x \qquad \text{(Temporary life annuity)} \tag{2.5}$$

These basic types already show that the main characteristics which should be taken care of, when it comes to a grouping process are the age, the duration and the technical interest rate. The following graphs and analyses are based on an endowment policy with a duration of 25 years, a sum insured of 10.000 € and an investment return of 3% p.a. over the whole projection horizon. The duration of 25 years is in all subsequent considerations equal to the duration of the premium payments which are made on a yearly basis. All other parameters especially the lapse, paid-up and surrender rates as well as first and second order assumptions can't be given here in detail.

[1] For detailed definitions and explanations see [9].

2.1. Age

The age of an insured person is one of the main factors which drives the projected outcome because it directly influences the probability of death and survival as shown in (2.1) - (2.5). When the age is changed from x to $x+1$ the death- and survival-probabilities ${}_kp_x$ and ${}_kq_x$ change as well. It is impossible to predict in general whether the probabilities will rise or fall. As known from life tables it is a bit more likely to die just after birth than a bit afterwards and the same is true for people aged around 20. The exact ages where the probability of survival increases and the probability of death decreases when a person gets a year older depends heavily on the life table and the sex of the insured person. Whether the values for (2.1) - (2.5) will rise or fall when the age x is increased by 1 year will therefore depend on n, x and the sex of the insured person. To get a better insight into the portfolio, simulation runs for various parameters should be done. In figure (2.1) the development of the yearly claims is plotted against the duration of 25 years. For a clearer chart the last cash flow which is the sum insured and therefore substantially larger than the yearly claims is omitted. We see that for younger people the claims are almost identical in the first years and only deviate slightly at the end of the duration. For an insured person with an age of 60 we see higher claims in ~~the first few years and a significantly increase of the claims during the whole projection period~~. ~~This increases the gap between a contract of a young and an old insured person as time increases~~. In figure (2.2) the development of the premiums is plotted against the age. For policyholders aged between 15 and 30 the premium stays almost constant ~~over the first few years~~ and then starts to increase exponentially. The increase of the premium is of exponential order due to the fact that the mortality rate is also increasing exponentially. Another fact that is not surprising is that the premium is the lower the higher the technical interest rate is, because the technical interest rate is used as a discount factor in (2.1) - (2.5). In figure (2.3) the present value of future profits (PVFP) at time 0 is plotted against the age. The PVFP is ~~the~~ the higher the higher the entry age of the policyholder is, which is a counter intuitive relation at first sight. ~~A profound analysis of that phenomenon can be made by looking at the yearly outputs for two persons aged 15 and 70 respectively which are given in table A.1.~~

1. When the guaranteed interest rate is roughly equal to the investment

return or even higher then it is more advantageous for the insurance company when the policyholder is older and therefore dies earlier because the difference between the guaranteed interest rate and the investment return need not to be financed over a long period.

2. The differences of the first and second order assumptions of the mortality rates are in absolute values the bigger the higher the age is. This directly leads to a higher risk margin for elder persons in absolute values. The resulting higher premium therefore increases the surplus in absolute values and leads to a higher present value of future profits.

In figure (2.4) the value of the reserve is plotted against the time. In all three cases the reserve is zero at the beginning and then starts to increase. This is due to the fact, ~~that the projection made, started with the first quarter of 2017.~~ We see that for the ages of 15 and 30 the difference in the reserve is negligible for all different values of the technical interest rate. The reserve for a 45 year old person is almost identical to the one ~~for the~~ younger policyholders at the first 10 years of the endowment but then increases at a slower rate. For a person aged 60 the reserve increases after ~~the first~~ 10 years at a much lower rate and even starts to decrease ~~after 18 years.~~ ~~For the first years the premium is relatively high and compensates the high mortality claims so that the reserve develops similar to the ones for the younger policyholders. After some years the initial population of 1 person has dramatically decreased which yields to a lower premium income but also to lower claims which need to be payed in the future.~~

2.2. Technical interest rate

The technical interest rate is one of the key assumptions in the life insurance business. It determines the factor by which the reserve and the savings premium increases during the contract period. After the contract has been concluded the technical interest rate is fixed and can't be changed by the insurance company. The maximum technical interest rate has been reduced dramatically by the Financial Market Authority in recent years from 4% to 0.5% as shown in table 2.1. The regulation of this upper bound is also relevant as it is often related to the minimum interest rate guaranteed to the policy

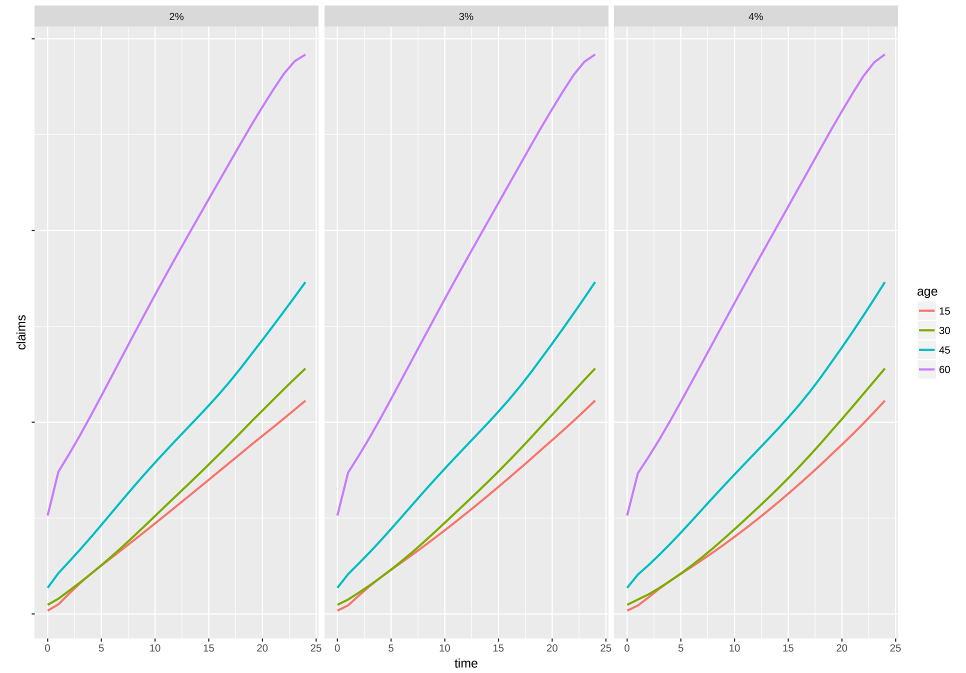

Figure 2.1.: Yearly cash flow for claims depending on the age and the technical interest rate.

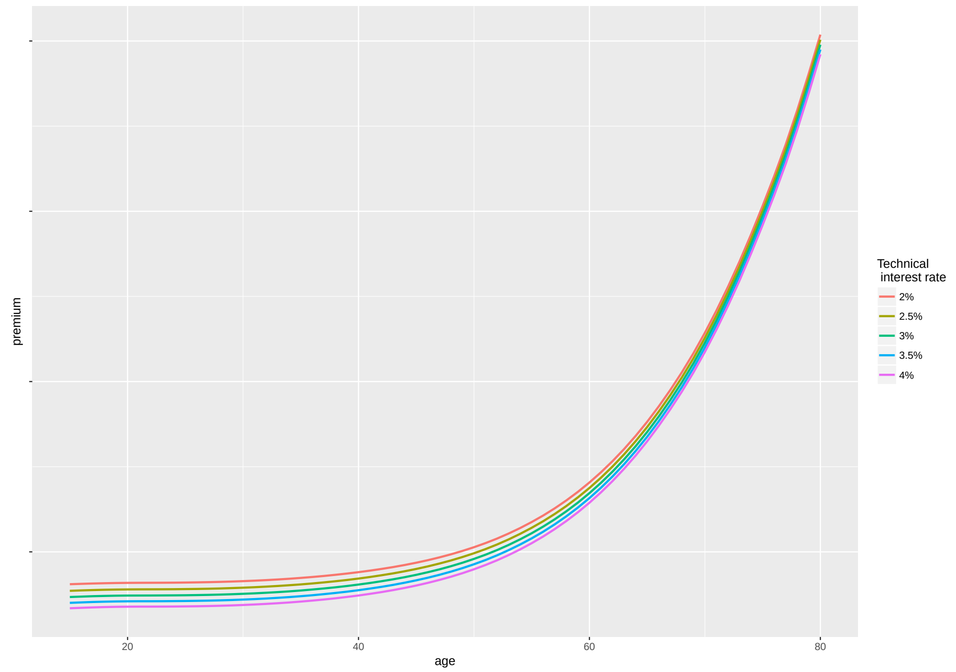

Figure 2.2.: Premiums depending on the age and the technical interest rate.

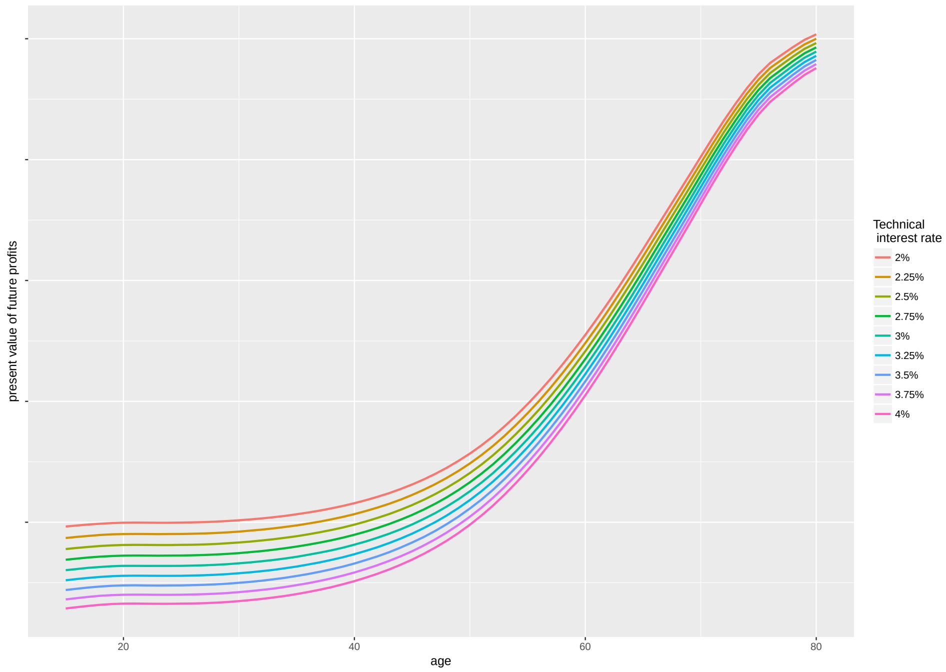

Figure 2.3.: Present value of future profits at time 0 depending on the age and the technical interest rate.

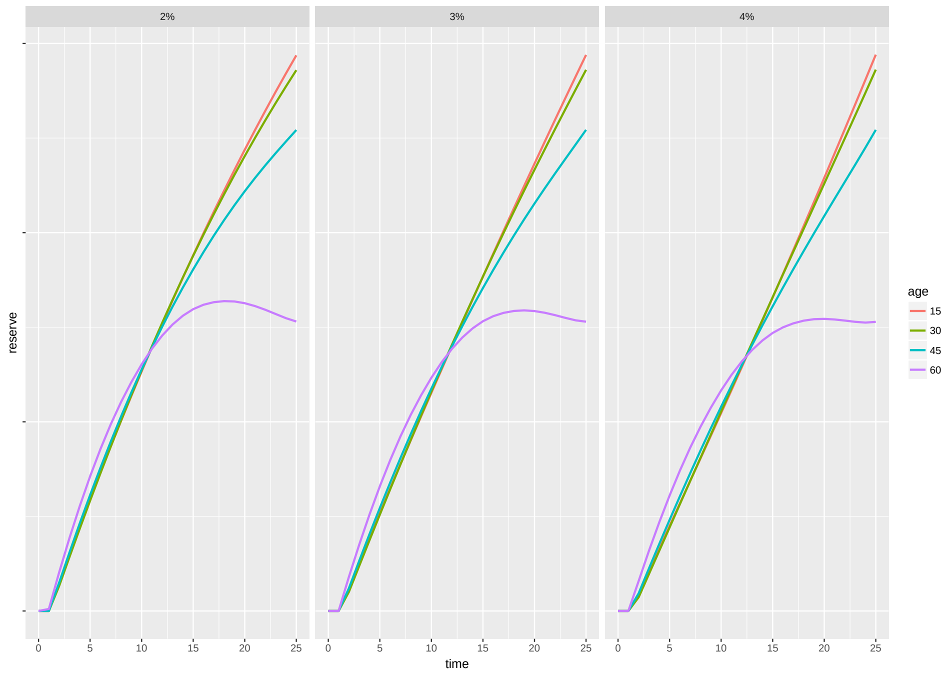

Figure 2.4.: Yearly reserve depending on the age and the technical interest rate.

holder. ~~A high technical interest rate also guarantees a higher benefit for the policyholder.~~ When ~~grouping policies from~~ different product generations with same payout characteristics but different technical interest rates ~~together~~ ~~several sensitivities must be taken into account.~~ When the technical interest rate increases, the present value of future profits as well as the premium will decrease as shown in figure (2.3) and (2.2) respectively. The reserve and the claims are not that much affected by an increase of the technical interest rate as shown in figure (2.4) and (2.1) respectively.

2.3. Duration

The duration n of an insurance contract is next to the age and the technical interest rate another main characteristic which need to be taken care of when a grouping process is carried out. In figure (2.5) the yearly claims except the last claim which is the maturity claim, are shown for different ages. One obvious conclusion that can be derived is that ~~the claims~~ ~~are~~ getting the higher the higher the age is. Another observation that can be made is that for any given time t the claims are the higher the shorter the duration gets when we keep the age fixed. This is not surprising at all, because a shorter duration goes along with a higher premium (see figure (2.6)) and ~~therefore surrender claims are higher in absolute values even if they are assumed to be a fixed percentage of the population.~~ In figure (2.6) we see that the premium is decreasing with an exponential order when the duration is increased. The difference between the premiums for policyholders with different ages is indistinguishable small for short term contracts and is getting bigger as duration increases. In figure (2.7) the present value of future profits at time 0 is plotted against the duration of the contract. We see the same effect as in figure (2.3) where contracts with higher ages lead to a higher PVFP. The effect of the absolute difference between the first and second order mortality assumptions is getting the bigger the longer the duration is and therefore the PVFP is getting the higher the longer the duration is. In the last sensitivity chart (2.8) the reserve is plotted against the time for different values of x and n. We see that for every time t the reserve is the higher the shorter the duration is, because a shorter duration leads, ceteris paribus, to a higher premium which then results in a higher reserve. For short term contracts up to 10 years the reserve is approximately the same across

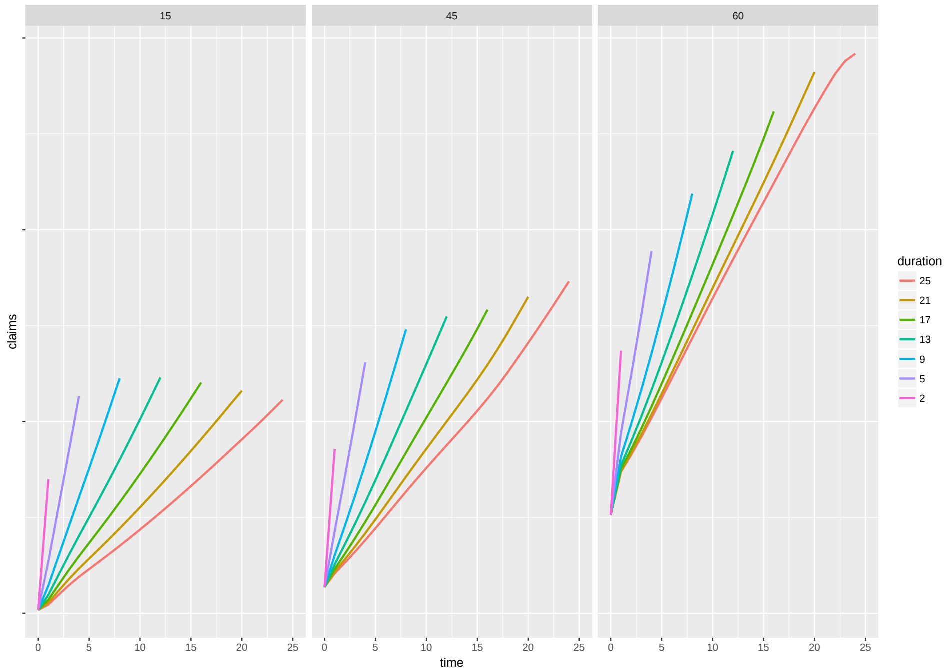

Figure 2.5.: Yearly cash flow for claims depending on the duration and the age.

different ages and is then getting the lower the higher the age and the longer the duration is.

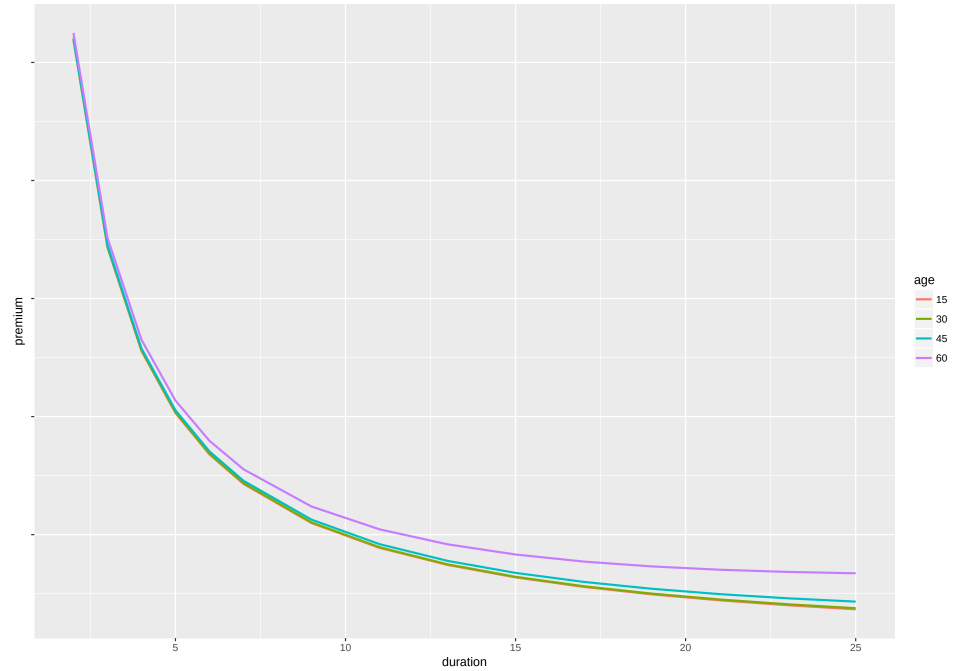

Figure 2.6.: Premiums depending on the duration and the age

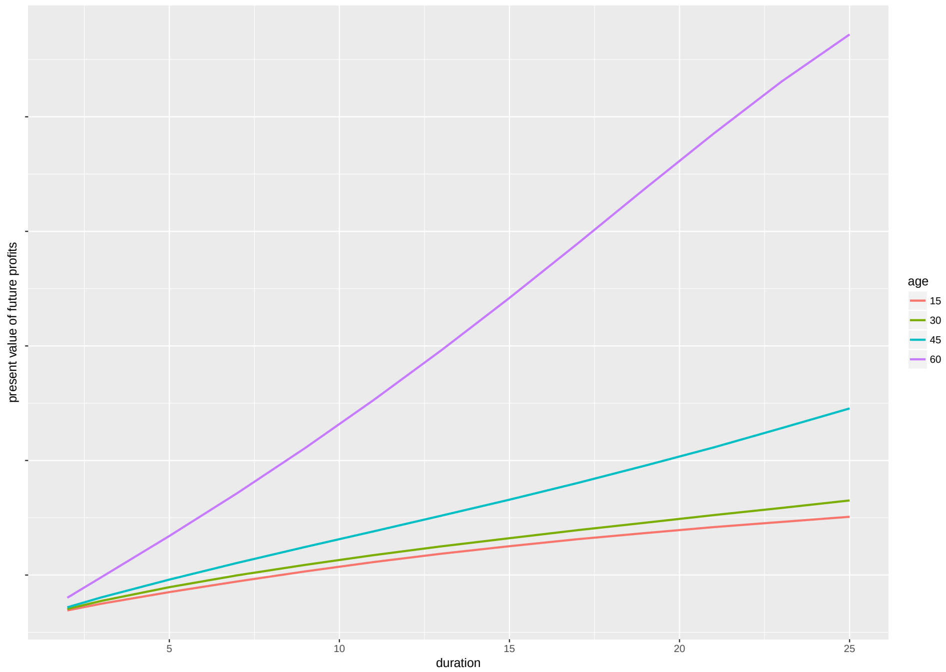

Figure 2.7.: Present value of future profits at time 0 depending on the duration and the age.

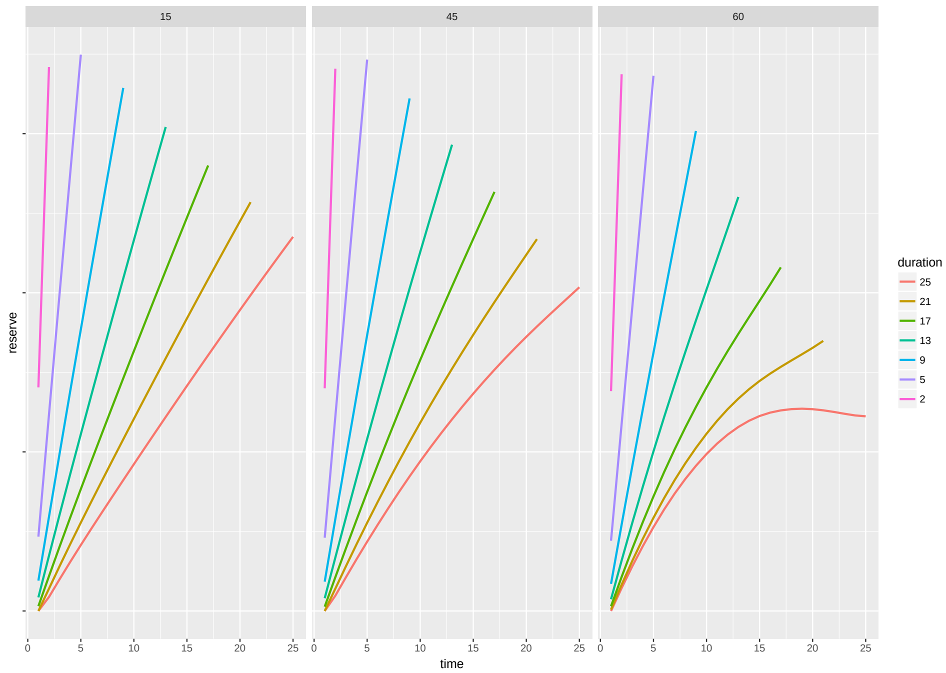

Figure 2.8.: Yearly reserve depending on the duration and the age.

3. K-means

To be able to process, summarize and understand huge amounts of data one is interested in methods which can find patterns in the data and represent them by just a few representatives. Given a data set of observations the challenge is, based on a measure of similarity, to find groups of observations which are quite similar within each group but quite different to all the other groups. If this task has to done with unlabeled data it is referred to as unsupervised clustering [10]. One of the most widely used unsupervised clustering approaches is the K-means clustering. The K-means algorithm is a simple approach in cluster analysis which splits a data set of n p-dimensional observations into K distinct clusters. Each observation belongs uniquely to exactly one of the K clusters, where K is a user defined number of clusters. Let $C_1, C_2 \ldots, C_K$ donate the sets containing the indices of the observations related to the clusters, then we get:

$$C = C_1 \cup C_2 \cdots \cup C_K = \{1, \ldots, n\}$$

$$C_j \cap C_k = \emptyset \quad \forall j \neq k$$

$$\sum_{k=1}^{K} |C_k| = n$$

The basic idea of the k-means clustering is to minimize the variation within the clusters.Typically the Euclidean metric is used for computing the distance between the different points. As a result the algorithm will find clusters which are spherical shaped.We therefore define the variation in p dimension within a single cluster C_k by:

$$D(C_k) = \frac{1}{|C_k|} \sum_{j=1}^{p} \sum_{i,i' \in C_k} (x_{ij} - x_{i'j})^2$$

Using $\bar{x}_{kj} = \frac{1}{|C_k|}\sum_{i\in C_k} x_{ij}$ and $\sum_{i\in C_k}(x_{ij}-\bar{x}_{kj})^2 = \left(\sum_{i\in C_k} x_{ij}^2\right) - |C_k|\bar{x}_{kj}^2$ we get:

$$
\begin{aligned}
D(C_k) &= \frac{1}{|C_k|}\sum_{j=1}^{p}\sum_{i,i'\in C_k}(x_{ij}-x_{i'j})^2 \\
&\overset{(1)}{=} \sum_{j=1}^{p}\left(\sum_{i\in C_k} x_{ij}^2 - 2\bar{x}_{kj}\sum_{i\in C_k} x_{ij} + \sum_{i\in C_k} x_{ij}^2\right) \\
&\overset{(1)}{=} 2\sum_{j=1}^{p}\left(\sum_{i\in C_k} x_{ij}^2 - |C_k|\bar{x}_{kj}^2\right) \\
&\overset{(2)}{=} 2\sum_{i\in C_k}\sum_{j=1}^{p}(x_{ij}-\bar{x}_{kj})^2
\end{aligned}
\tag{3.1}
$$

Using (3.1) we are able to formulate the optimization problem that defines the K-means clustering:

$$
C^* = \min_{C_1,\ldots,C_K}\sum_{k=1}^{K} D(C_k) = \min_{C_1,\ldots,C_K} 2\sum_{k=1}^{K}\sum_{i\in C_k}\sum_{j=1}^{p}(x_{ij}-\bar{x}_{kj})^2 \tag{3.2}
$$

Solving this problem is NP hard as shown in [12], but it is possible to provide algorithms which converges to a local optimum. K-means starts either with an initial assignment of observations to clusters or with K initially selected means and begins then to reduce the squared error. The squared error always decreases with an increase in the number of clusters with the extreme case of (3.2) is zero for K = n. One possible formulation for an algorithm that converges to a local optimum which is given below is provided in [11]. Note that algorithm 1 has a fixed number of clusters K, but never increases the objective function (3.2) because we have a finite set of possible partitions and

1. For any set of observations S it hods that:

   $$\bar{x}_S = \operatorname*{argmin}_x \sum_{i\in S}(x_i - x)^2$$

   Hence, step $2a$) minimizes the sum of squared deviation.

2. Step 2b) which reassigns the observations to the new nearest centroid can only reduce the objective function.

Algorithm 1 K-means clustering

---

1. Assign all observations initially to one cluster.(e.g. Randomly assign a number, from 1 to K, to each of the observations. These serve as initial cluster assignments for the observations.)
2. Iterate still the cluster assignments stop changing:
   a) For each of the K cluster, compute the cluster centroid. The kth cluster centroid is the vector of the p parameter means for the observations in the kth cluster.
   b) Assign each observation to the cluster whose centroid is closest.

Because the K-means algorithm finds a local rather than a global optimum, different initial cluster assignments in step 1 of algorithm 1 will may lead to different results. It is therefore important to run the algorithm multiple times with different initial assignments and select than the (best) solution based on the value of the minimized objective function (3.2).

$$\sum_{i=0}^{\infty} \tag{3.3}$$

$$\sum_{i=0}^{\infty} \tag{3.3*}$$

Wir können verweise auf Gleichung 3.3 und auch auf Gleichung 3.3*

Appendix A.

Tables

test

month	pop_15	pop_70	prem_15	prem_70	prem_diff	claims_15	claims_70	claims_diff
0	1	1						
9	0.999834	0.987259	385.60	1128.40	742.80	1.70	130.61	128.91
21	0.977140	0.947577	379.73	1092.63	712.90	4.73	185.70	180.97
33	0.952445	0.905554	370.17	1045.64	675.47	10.08	201.42	191.34
45	0.928273	0.863652	360.80	998.80	638.00	15.22	216.81	201.60
57	0.904661	0.821842	351.64	952.09	600.45	19.87	232.12	212.25
69	0.881638	0.780093	342.69	905.48	562.79	24.11	247.29	223.18
93	0.837332	0.696652	325.47	812.40	486.93	32.45	277.16	244.71
105	0.816021	0.654913	317.19	765.86	448.68	36.70	291.89	255.19
117	0.795253	0.613146	309.11	719.31	410.19	41.02	306.32	265.29
129	0.775013	0.571358	301.25	672.72	371.47	45.41	320.21	274.81
141	0.755288	0.529577	293.58	626.11	332.54	49.85	333.38	283.53
165	0.717332	0.446430	278.82	533.03	254.20	58.90	354.70	295.80
177	0.699076	0.405693	271.73	486.94	215.21	63.50	359.06	295.56
189	0.681284	0.365984	264.81	441.73	176.92	68.15	359.15	291.01
201	0.663944	0.327583	258.07	397.76	139.68	72.83	355.49	282.66
213	0.647046	0.290722	251.51	355.31	103.81	77.54	348.61	271.07
237	0.614523	0.222339	238.87	275.96	37.09	87.12	326.91	239.79
249	0.598856	0.191115	232.78	239.40	6.62	91.79	312.50	220.71
261	0.583562	0.162081	226.85	205.15	-21.70	96.52	295.62	199.10
273	0.568621	0.135426	221.05	173.38	-47.67	101.35	276.01	174.66
285	0.554015	0.111340	215.39	144.32	-71.07	106.29	253.63	147.35
297	0.539728	0.089980	209.85	118.16	-91.69	111.28	228.95	117.68
309	0.000000	0.000000	0.00	0.00	0.00	5983.32	991.18	-4992.14

Table A.1.: Yearly outputs for PVFP-sensitivity analysis.

The todo text is the text that will be shown in the todonote and in the list of todos. The optional argument options, allows the user to customize the appearance of the inserted todonotes. For a description of all the options see sadf The todo text is the text that will be shown in the todonote and in the list of todos. The optional argument options, allows the user to customize the appearance of the inserted todonotes. For a description of all the options see The todo text is the text that will be shown in the todonote and in the list of todos. The optional argument options, allows the user to customize the appearance of the inserted todonotes. For a description of all the options see The todo text is the text that will be shown in the todonote and in the list of todos. The optional argument options, allows the user to customize the appearance of the inserted todonotes. For a description of all the options see fasd fadsf asdf

Make a cake . . .

Make a sketch of the structure of a trebuchet.

Bibliography

[1] CEIOPS. *CEIOPS' Advice for Level 2 Implementing Measures on Solvency II: Technical provisions*. Oct. 2009. URL: https://eiopa.europa.eu/CEIOPS-Archive/Documents/Advices/CEIOPS-L2-Final-Advice-on-TP-Best-Estimate.pdf (cit. on p. 1).

[2] European Commission. *Official Journal of the European Union*. Vol. 58. L. (Cit. on p. 2).

[3] European Commission. *Regulatory process in financial services*. URL: https://ec.europa.eu/info/business-economy-euro/banking-and-finance/financial-reforms-and-their-progress/regulatory-process-financial-services/regulatory-process-financial-services_en (cit. on p. 1).

[4] EIOPA. *Final Report on Public Consultation No. 14/036 on Guidelines on valuation of technical provisions*. 2.02.2015. URL: https://eiopa.europa.eu/Publications/Guidelines/Final_Report_Val_tech_prov_GLs.pdf (cit. on p. 3).

[5] EIOPA. *Guidelines on valuation of technical provisions*. 2.02.2015. URL: https://eiopa.europa.eu/Publications/Guidelines/TP_Final_document_EN.pdf (cit. on p. 3).

[6] EIOPA. *Solvency II*. URL: https://eiopa.europa.eu/regulation-supervision/insurance/solvency-ii (cit. on p. 1).

[7] FMA. *Fakten, Trends & Strategien 2016*. 2016. URL: https://www.fma.gv.at/publikationen/fakten-trends-strategien/ (cit. on p. 5).

[8] FMA. *Solvency II*. URL: https://www.fma.gv.at/en/insurance/solvency-ii/ (cit. on p. 1).

[9] Hans U. Gerber. *Life Insurance Mathematics* -. Berlin Heidelberg: Springer Science & Business Media, 2013. ISBN: 978-3-662-03460-6 (cit. on p. 6).

[10] Anil K Jain. “Data clustering: 50 years beyond K-means.” In: *Pattern recognition letters* 31.8 (2010), pp. 651–666 (cit. on p. 15).

[11] Gareth James et al. *An Introduction to Statistical Learning - with Applications in R*. 1st ed. 2013, Corr. 6th printing 2016. Berlin Heidelberg: Springer Science & Business Media, 2013. ISBN: 978-1-461-47138-7 (cit. on p. 16).

[12] Meena Mahajan, Prajakta Nimbhorkar, and Kasturi Varadarajan. “The Planar k-Means Problem is NP-Hard.” In: *WALCOM: Algorithms and Computation: Third International Workshop, WALCOM 2009, Kolkata, India, February 18-20, 2009. Proceedings*. Ed. by Sandip Das and Ryuhei Uehara. Springer Berlin Heidelberg, 2009, pp. 274–285. ISBN: 978-3-642-00202-1. URL: `http://dx.doi.org/10.1007/978-3-642-00202-1_24` (cit. on p. 16).

[13] EUROPEAN PARLIAMENT and THE COUNCIL. *Directive 2009/138/EC on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II)*. 25 Nov. 2009. URL: `http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32009L0138&from=EN` (cit. on p. 2).