

预测分析框架—可复用的经验分析工作 流程



摘要

预测建模是一种基于统计技术进行结果预测的有效工具，广泛应用于医疗、金融等行业，为决策者提供基于历史数据的专业洞察，帮助其做出明智决策。本报告阐述了精算经验研究中开发预测模型的框架。为了将该框架与实际精算问题相结合，报告将重点分析和建模不同产品（终身寿险、定期寿险等）的死亡率差异。通过建模方法，我们得出以下关键发现：

- 优选死亡率的相对差异因产品而异：
在两类优选体系中，定期寿险的标准死亡率显著高于优选死亡率；
在四类优选体系中，万能寿险/变额寿险及带保证型（UL/VL/ULSG/VLSG）产品的死亡率差异范围明显大于其他产品。
- 保额因子差异：灵活型产品（xL）、终身产品（Perm）和定期产品（Term）的保额因子之间存在分化，其中 xL 相对于 Term 差异呈现收窄趋势。
- 核保年龄率差异：
65 岁以下年龄段：核保年龄斜率显示 Term 比 Perm 和 xL 更陡；
65 岁以上年龄段：Perm 的核保年龄斜率相对于 xL 呈现更陡的趋势。
- 不同产品在保单生效后的第 1 和第 2 个年度的死亡率存在差异。
- 自 1990-1999 核保年度以来，xL 相对于 Term 的相对死亡率呈现小幅但持续的增长趋势，目前 xL 已接近 Term 水平。

1. 机器学习在死亡率研究中的精算应用

Caveat and Disclaimer

The opinions expressed and conclusions reached by the authors are their own and do not represent any official position or opinion of the Society of Actuaries Research Institute, the Society of Actuaries or its members. The Society of Actuaries Research Institute makes no representation or warranty to the accuracy of the information.

经验研究是精算师用来量化和理解历史经验的主要工具。将统计技术应用于经验研究以发现新的相关见解，是精算建模的自然演进。对于精算师来说，这种方法具有多重优势：

- 避免繁琐且可能具有误导性的单变量分析；
- 适当考虑可信度并充分利用数据中固有的所有可信度；
- 更容易发现并合理调整变量间的交互作用；
- 统计控制死亡率研究中控制多个协变量给定单元格内不同来源的影响。

2. 问题陈述

精算师关注的核心问题之一在于：不同产品为何会表现出差异化的死亡率结果。尽管很难单独识别和量化，但众所周知，核保规则、目标市场、保单持有人行为和社会经济因素等都会直接影响死亡率结果。通过统计或机器学习模型，我们有可能解释这些变量的影响。对于本项目，我们试图回答的关键问题是：在个险经验委员会（ILEC）数据集中，死亡率如何因产品而异。在本文使用的简化数据集中，产品类别包括定期寿险（Term）、长期寿险（Perm）、万能/变额寿险（UL/VL）和其他产品（Other）。为了理解不同产品的死亡率差异，我们将构建机器学习模型来预测死亡率结果，并分析结果以获得相关见解。

3. 方法论

该框架规范了预测建模的实施路径，涵盖代码设置、模型创建、数据预处理和验证的过程，并针对精算建模中的常见挑战提出解决方案，例如：纳入非线性关系、确定交互作用、处理欠拟合和过拟合（偏差-方差权衡）以及模型可解释性。

本项目的目标是提供实用的技术、代码和思路，以帮助精算师未来对死亡率研究的分析。常见建模过程都有几个关键步骤：数据预处理、数据探索、模型选择、模型验证和模型解释。关于这些主题可以讨论的内容远多于我们在此探讨的范围，我们旨在解决与经验研究相关的关键考虑因素。

4. 建模方法

在将统计和机器学习应用于经验研究时，可以采取多种不同的建模方法。我们将重点关注最常用的方法：广义线性模型（GLM）、带惩罚项的广义线性模型（也称为弹性网络 GLM）和梯度提升机（GBM 或 GBDT）。除此之外，还有许多其他方法或这些方法的变体也是合理的。

4.1 广义线性模型（GLM）

广义线性模型是精算领域历史最悠久的研究方法，从某种意义上说也是最简单的。GLM 的一个优点是它们允许进行统计假设检验。例如，可以对各个模型系数进行统计检验，并执行多种统计检验以验证结果和比较模型。并且，GLM 的结果也相对容易解释。

然而，GLM 也存在一些局限性：由于其相对简单，它们的预测能力低于其他方法。为了获得 GLM 的最佳性能，需要额外的努力来建模非线性关系和交互作用。因此，相对其他方法来说，其模型优化过程更加耗时。

GLM 可以扩展为正则化 GLM，例如 LASSO 或 Ridge，其本质是对拟合模型的目标函数进行修正。这种正则化项具有若干优点、缺点以及对建模过程的改变。首先，添加惩罚项使得置信区间和假设检验失效。因此不再使用参数假设检验和似然比检验来评估模型的相对拟合优度，而是应用机器学习范式通过交叉验证来优化模型。

值得强调的是，正则化 GLM 仍然保持了线性模型的可解释性优势，并且可以提高模型的整体预测准确性。此外 LASSO 惩罚可实现模型自动执行变量选择。

4.2 梯度提升决策树 (GBDT)

梯度提升决策树是一种分阶段构建的决策树集成模型。每个决策树都是在前一个树的残差上递归训练的。第一棵树是基于目标结果的决策树，第二棵树是基于前一棵树的残差的决策树，以此类推。通过这种方式，模型不断重新聚焦于其预测最弱的区域。

梯度提升决策树的流行框架包括 LightGBM、CatBoost 和 XGBoost。该模型是表格数据进行分类和回归最有效的方法之一。

梯度提升机（此处使用 LightGBM 实现）由于其极高的准确性、易用性和自动捕捉变量的重要交互作用的能力，已成为许多表格机器学习任务的主流方法。然而其模型解释较为困难，为此，我们将使用 SHAP 值这一主流的解释方法来辅助分析。

5. 总结

在本报告中，我们探讨了预测建模框架在精算经验研究中的应用，重点关注 ILEC 数据集中不同产品类型死亡率差异。通过分析，我们揭示了若干关键发现和死亡率差异，这些结果有助于理解死亡率的影响因素。

本框架及其研究结果展示了预测建模的应用价值及所需考量，同时也凸显了其在支持精算决策方面的重要意义。

About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its [strategic research programs](#): aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of [topical research available](#), including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
8770 W Bryn Mawr Ave, Suite 1000
Chicago, IL 60631
www.SOA.org