The Individual Life Experience Committee’s analytics crowdsourcing competition for 2021 has drawn to a close. The goals of the competition are two-fold: to encourage actuaries and non-actuaries to engage with life insurance data in meaningful and innovative ways, and to expose innovative data analysis methods and insights to the committee and to the profession. The competition asked competitors to predict 2017 mortality using the provided 2009 – 2016 ILEC dataset.

About the Winning Entry

This year’s winner applied a novel approach to estimate mortality curves: multiple output Gaussian processes (MOGP). A major feature of MOGP models is their ability to jointly model multiple, interrelated outputs. The theoretical basis of MOGP models rests in part on an unexpected source from the life actuarial perspective, temporo-spatial analytics methods from geostatistics. The interrelated outputs in this instance are the mortality rates from the population mortality datasets from the Human Mortality Database, and the same from the ILEC dataset.

The authors of the winning entry then used these attained age curves as an offset for a ridge regression step, where they employed a Gaussian elastic net GLM on the difference between the MOGP prior mortality rates and the ILEC rates.

In general, their model did the best overall in capturing expected interrelationships in the ILEC data, such as risk class, face amount, and product differentials.

No entry is ever perfect, and we feel that the authors could have improved their entry in at least two respects as it relates to the ridge regression step.

1. Using a Poisson GLM instead of a Gaussian GLM, with the log of expected mortality from the MOGP model as the offset
2. Modeling smooth covariates, such as duration, with spline bases as warranted

About the Other Entries

We had two other entries. The authors of these entries took approaches which relied heavily on a decision tree in the one case and boosted decision trees in the other. Neither of their approaches adequately captured all the variation present in the data.

We invite our interested colleagues to look at an example where boosted decision trees were successful in modeling mortality: Interpretable Machine Learning for Insurance by Larry Baeder, Peggy Brinkman, FCAS, and Eric Xu, FCAS. Baeder, et al, used boosted decision trees on the 2009 – 2015 ILEC term data. Their models relied on hundreds of boosting iterations with shallow individual trees. An entry which mimicked this approach would have done well.