ISSN 0002-306X. Изв. НАН РА и ГИУА. Сер. ТН. 2007. Т. LX. № 4.

УДК 65.011:346.61:61:368

АВТОМАТИЗАЦИЯ И СИСТЕМЫ УПРАВЛЕНИЯ

Р.А. АВЕТИСЯН, А.В. ПОГОСЯН

МЕТОДЫ ВЫБОРА ТАРИФНЫХ ФАКТОРОВ В МЕДИЦИНСКОМ СТРАХОВАНИИ

Описывается один из методов выбора тарифных факторов для расчета страховых тарифов. За неимением достаточного статистического материала для отбора факторов использование методов дисперсионного или регрессионного анализов зачастую дает неверные результаты. Рассматривается адаптированный метод дисперсионного двухфакторного анализа оптимального выбора тарифных факторов для расчета страховых тарифов при наличии статистической информации за один год.

Ключевые слова: тарифный фактор, нормированный убыток, страхование, двухфакторный дисперсионный анализ.

Во многих видах рискового страхования количество факторов, влияющих на процесс убытков, довольно велико. Но между ними часто существует зависимость, и нужно следить за тем, чтобы одно и то же обстоятельство не учитывалось в тарифе многократно.

Приведем пример. Анализ выборки полисов немецкого страхования автогражданской ответственности в 1978 году показал, что около 25% автомобилей управляются (преимущественно) женщинами, и их нормированный убыток равен примерно 90% среднего нормированного убытка всех водителей. Тот же показатель (преимущественных) водителей мужского пола составил примерно 103%. Одновременно обнаружилось и различие в мощности автомобилей: у «женских» она составила в среднем 40 *кВт*, а у «мужских» — 53 *кВт*. Поскольку с ростом мощности нормированный убыток увеличивается, естественно предположить, что различие нормированных убытков мужчин и женщин во многом объясняется различием мощностей. Более подробные исследования подтвердили отсутствие значимого отличия между мужчинами и женщинами в отношении нормированного убытка при одинаковой мощности автомобиля. Значит, было бы несправедливо применять пол в качестве тарифного фактора, если мощность уже является таковой.

Однако чаще всего ситуация бывает не настолько прозрачной, как в этом примере. После того, как некоторый фактор риска выбран в качестве тарифного, влияние некоторых других факторов уменьшается, но, вообще говоря, не исчезает. Это создает трудности при выборе наиболее эффективных факторов риска, тем более, что, с практической точки зрения, выгодно иметь как можно меньше тарифных факторов [1].

Выбор факторов со значимым влиянием на целевую переменную — классическая задача многомерной статистики, решаемая в рамках как регрессионного, так и дискриминантного анализа. Но, как и методы кластер-анализа, эти методы в общем случае предполагают равнозначность объектов, характеризуемых факторами (в смысле независимых и потенциально одинаково распределенных повторений). В нашем же случае риски, как правило, различаются по «размеру» (например, имеют разные страховые суммы). Исключение составляют разве что страхование автогражданской ответственности и массовое страхование имущества, где различия размеров рисков невелики. Для этих видов страхования названные выше стандартные методы могут применяться без какой-либо корректировки при наличии данных по каждому отдельному полису. Далее рассматриваются методы, не требующие данных по отдельным полисам. При этом предполагается, что количество значений каждого фактора сокращено до обозримого числа, например, с помощью агломеративного кластер - метода.

Под классом условимся понимать объединение значений факторов следующим образом: в одни классы попадают максимально схожие значения факторов, а в другие - максимально отличающиеся значения факторов. Количественным описанием схожести служат ставки возмещений соответствующих рисков (страховых полисов).

Из всех факторов, имеющих значимое влияние на случайную величину, убыток на на один полисо – год (полисо – год: количество полисов страхового портфеля на один год), логично в первую выбрать фактор с наиболее отчетливыми различиями реализаций этой случайной величины по классам значений. Количественное различие классов можно определить степенью неприятия гипотезы равенства математических ожиданий нормированных годовых убытков. Тем самым задача отбора факторов сводится к проверке гипотезы равенства средних. Самый известный способ проверки гипотезы равенства математических ожиданий – дисперсионный анализ [1]. Он подразумевает нормально распределенные величины с одинаковой дисперсией.

Привести распределение случайной величины к нормальному можно посредством логарифмирования значений случайной величины, получая в итоге так называемое логнормальное распределение. Будем считать, что у всех факторов риска нормированный убыток в каждом классе распределен логнормально с параметрами ϑ и σ^2/ν, где ν – объем класса. Параметр ϑ варьирует по классам и по факторам риска, параметр σ^2 в дисперсионном анализе предполагается постоянным, это допускается с учетом эвристического характера отбора факторов.

Введем следующие обозначения : Z_{ik} - значение i - го нормированного совокупного убытка в k -ом классе ; v_{ik} - объем i-го нормированного совокупного убытка в k -ом классе.

Под нормированным совокупным убытком понимаем отношение суммы страховых возмещений (полисов) к количеству полисов в страховом портфеле (классе).

Обозначим $W_{ik} = \ln(Z_{ik})$. При наличии годичной статистики (как, например, в Республике Армения, где медицинское страхование начало развиваться с 2006 года) нецелесообразно применять вышеперечисленные методы, так как модели, составленные с их помощью, оказываются чрезмерно параметризированными, и поэтому приходится учитывать нормированные убытки и объемы классов в течение нескольких лет, т.е. каждой фазе итерации соответствует $I \cdot K + 1$ параметров (ϑ_{ik} и σ^2).

Мы предлагаем действовать по аналогии с двухфакторным дисперсионным анализом и предложить модель (E - оператор математического ожидания)

$$E(W_{ik}) = \vartheta + a_i + b_k, \quad 1 \le i \le I, 1 \le k \le K,$$
$$\sum a_i = 0 \quad \text{и} \quad \sum b_k = 0.$$

содержащую всего $I + K - 1$ свободных параметров. Из статистики критерия двухфакторного дисперсионного анализа посредством взвешивания по объему получается статистика для нашей логнормальной модели:

$$\frac{SS/(K-1)}{RSS/((I-1)(K-1))} = (I-1)SS/RSS,$$

где

$$SS = \sum_{k \ge 1} v_{+k}(\underline{W}_{+k} - \underline{W}_{++})^2,$$

$$RSS = \sum_{i,k} v_{ik}(W_{ik} - \underline{W}_{i+} - \underline{W}_{+k} + \underline{W}_{++})^2,$$

$$\underline{W}_{+k} = \sum_{i \ge 1} v_{ik} W_{ik} / \sum_{i \ge 1} v_{ik}, \quad \underline{W}_{i+} = \sum_{k \ge 1} v_{ik} W_{ik} / \sum_{k \ge 1} v_{ik},$$

$$\underline{W}_{++} = \sum_{i \ge 1} \sum_{k \ge 1} v_{ik} W_{ik} / \sum_{i \ge 1} \sum_{k \ge 1} v_{ik}.$$

При справедливости нулевой гипотезы $b_1 = b_2 = \ldots = b_k = 0$ эта статистика имеет F-распределение (распределение Фишера) с $K-1$ и $(I-1)(K-1)$ степенями свободы. Нулевая гипотеза означает, что различие в математических ожиданиях $E(W_{ik})$ связано только с различием параметров $a_1, \ldots, a_I$ и дополнительный фактор с K классами значений не влияет на случайную величину убытка.

Условие $I = 1$ не допускает проведения двухфакторного дисперсионного анализа на начальном шаге отбора — при выборе первого тарифного фактора. Критерием для отбора первого тарифного фактора может служить значение статистики

$$T = \sum_{k \ge 1} v_k (W_k - \underline{W}_+)^2 / (K-1).$$

При нулевой гипотезе равенства математических ожиданий величин $W_k = \ln(Z_k)$ всех классов $k = 1, \ldots, K$ значений рассматриваемого фактора эта статистика имеет математическое ожидание σ^2, Первым выбирается фактор с наибольшим значением статистики T, т. е. наиболее отчетливыми различиями средних по классам (в любом случае предполагается значимость фактора) [1]. Обратим внимание на то, что сравниваемые факторы должны содержать примерно одинаковое число K классов, поскольку для факторов с небольшим числом классов более вероятны высокие значения T. Это следует из формулы $Var(T)=2\sigma^4/(K-1)$, где Var - оператор дисперсии.

Рассмотренный метод двухфакторного дисперсионного анализа отбора тарифных факторов сводится к проверке гипотезы равенства нескольких математических ожиданий. Не гарантируя движения оптимума, последовательный отбор, тем не менее, позволяет легко выявить взаимную зависимость факторов риска. Пошаговая оптимизация является стандартным методом многомерной статистики [2]. Для выбора первого тарифного фактора еще подойдет критерий Крюскала - Уэллиса (Kruskal – Wallis). Но в фазе итерации обойтись без модели распределения не удастся — к двухфакторной модели не применим ни один критерий, не завися от распределения.

СПИСОК ЛИТЕРАТУРЫ

1. **Мак Томас.** Математика рискового страхования. – М.: Изд. "ОЛИМП-БИЗНЕС", 2005. - 432 с.
2. **Аветисян Р., Погосян А.** Страхование как метод минимизации финансовых ущербов // Сборник материалов годичной научной конференции ГИУА.Т2.- 2007.- С. 687 – 690.

ГИУА. Материал поступил в редакцию 27.07.2007.

Ռ.Ա. ԱՎԵՏԻՍՅԱՆ, Ա.Վ. ՊՈՂՈՍՅԱՆ

ԱՊԱՀՈՎԱԳՐԱՎՃԱՐԱՅԻՆ ԳՈՐԾՈՆՆԵՐԻ ԸՆՏՐՈՒԹՅԱՆ ՄԵԹՈԴՆԵՐԸ ԲԺՇԿԱԿԱՆ ԱՊԱՀՈՎԱԳՐՈՒԹՅՈՒՆՈՒՄ

Նկարագրված է ապահովագրավճարների հաշվարկման ժամանակ ապահովագրավճարային գործոնների ընտրության մեթոդներից մեկը: Վիճակագրական տվյալների քիչ լինելու դեպքում դիսպերսիոն կամ ռեգրեսիոն վերլուծության մեթոդների կիրառումը հաճախակի բերում է սխալ արդյունքի: Բերված է երկգործոն դիսպերսիոն վերլուծության համաձայնեցված մեթոդը ապահովագրավճարային գործոնների ընտրության համար ապահովագրավճարների հաշվարկման ժամանակ մեկ տարվա վիճակագրական տվյալների առկայության դեպքում:

Առանցքային բառեր՝ ապահովագրավճարային գործոն, նորմավորված ծախս, ապահովագրություն, երկգործոն դիսպերսիոն վերլուծություն:

R.A. AVETISYAN, A.V. POGHOSYAN

METHODS OF RATE FACTOR SELECTION IN MEDICAL INSURANCE

One of the methods of rate factor selection for calculation of the rates of insurance is introduced. In the absence of sufficient statistic material for the selection of factors, application of deviation or regression analyses provide incorrect (false) results. An adapted method of two-factor deviation analysis of the rate factors optimal choice for calculation of the rates of insurance in the presence of statistical information for one year is considered.

Keywords: rate factor, standardized loss, insurance, two-factor deviation analysis.