

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ

ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

ИМЕНИ Н. Г. ЧЕРНЫШЕВСКОГО»

Кафедра теории функций и стохастического анализа

**АНАЛИЗ МЕТОДОВ ВЫБОРА ТАРИФНЫХ ФАКТОРОВ В
АВТОСТРАХОВАНИИ**

АВТОРЕФЕРАТ БАКАЛАВРСКОЙ РАБОТЫ

студента 4 курса 412 группы

направления 01.03.02 — Прикладная математика и информатика

механико-математического факультета

Лючева Владислава Андреевича

Научный руководитель

доцент, к. ф.-м. н.

Л. В. Борисова

Заведующий кафедрой

д. ф.-м. н., доцент

С. П. Сидоров

Саратов 2022

ВВЕДЕНИЕ

Страхование – важная отрасль коммерческой деятельности современного экономического мира ввиду опасностей, подстерегающих людей на каждом шагу. Каждый день человек подвергается всевозможным рискам, приносящим разной степени тяжести ущерб. Частота опасностей и масштабы последствий от них прогрессивно возрастают, а также появляются новые риски, поэтому страхование крайне востребовано в развитом обществе.

Автомобили стали неотъемлемой частью нашей жизни, поэтому число несчастных случаев и аварий не перестает расти с каждым годом. Почти каждый человек сегодня способен позволить себе покупку автомобиля, но далеко не все автомобилисты страхуют машины (за исключением обязательного страхования ОСАГО). Это связано, прежде всего, с экономией, нежеланием лишних затрат и не достаточным доверием к страховым компаниям.

Основа страхового дела - это правильные актуарные расчеты страховых премий, ведь именно страховые взносы формируют базу для страховых выплат, страховые резервы. Задача любого страховщика – это сделать клиенту максимально выгодное предложение, дать ему возможность максимально сэкономить, но в то же время не занизить страховые ставки из-за угрозы обанкротиться, именно в этом и заключается **актуальность** темы данной работы и ее **практическая значимость**.

Цель бакалаврской работы - анализ статистических методов отбора тарифных факторов и методов сокращения количества классов значений факторов с целью построения тарифных множителей.

Объект исследования – рынок автострахования в России.

Предмет исследования данной работы - совокупность показателей, характеризующих убыточность портфеля договоров страхования КАСКО и факторы, определяющие тарифообразование добровольного автосстрахования.

Для достижения поставленной цели необходимо решить следующие задачи **задачи**:

- проанализировать развитие и современное состояние рынка страхования КАСКО в России;

- изучить методы построения классов значений факторов риска, а также методы отбора самих тарифных факторов (наиболее важных факторов риска, на основе которых строится страховой тариф)
- построить тарифные множители для договоров КАСКО, в зависимости от региона использования транспортного средства.

Структура и содержание бакалаврской работы. Работа состоит из введения, четырех разделов, заключения, списка использованных источников, содержащего 20 наименований, и трех приложений. Общий объем составляет 54 страницы. Названия разделов:

- Экономико-статистический анализ рынка добровольного автострахования
- Методы построения классов значений факторов риска
- Анализ методов выбора тарифных факторов в автостраховании
- Вычислительный эксперимент

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обосновывается актуальность темы работы и ее практическая значимость, формулируется цель работы и ставятся основные задачи.

В **первом** разделе работы проводится экономико-статистический анализ рынка автострахования в России, рассматриваются основные тенденции и динамики рынка, такие как страховые премии, выплаты, а также количество страховых компаний на рынке. Проводится анализ влияния продаж новых автомобилей на совокупные страховые взносы компаний по сегменту КАСКО.

Рассмотрев динамику продаж новых автомобилей в России по таблице 1, можно отметить значительный рост данного показателя после упадка во время финансового кризиса 2008 года.

Таблица 1 - Динамика продаж новых автомобилей в РФ

Год	Количество продаж (млн. штук)
2007	2.746
2008	2.897
2009	1.465
2010	1.91
2011	2.653
2012	2.935
2013	2.777
2014	2.491
2015	1.601
2016	1.426
2017	1.596
2018	1.801
2019	1.76
2020	1.599
2021	1.667

Рынок КАСКО играет существенную роль в развитии страхования в России: около 11% собираемых страховых премий в стране приходится на добровольное автострахование, в таблице 2 представлена доля страхования КАСКО с 2008 года. Динамика собираемых премий по сегменту КАСКО повторяет динамику всего рынка. Необходимо отметить, что рынок КАСКО напрямую зависит от количества купленных кредитных автомобилей в силу обязательности страхования автомобиля, купленного в кредит.

Таблица 2 - Доля сегмента КАСКО на страховом рынке

Год	КАСКО	Весь рынок	% КАСКО
2008	138.2	551.6	25 %
2009	112.5	513.1	21.9 %
2010	139.1	557.2	25 %
2011	164.6	664.4	24.8 %
2012	195.3	809	24.1 %
2013	212.3	904.9	23.5 %
2014	218.5	987.8	22.1 %
2015	187.2	1023.8	18.2 %
2016	170.7	1180.6	14.4 %
2017	118	1277.5	9.2 %
2018	168.7	1478.5	11.4 %
2019	170.5	1482	11.5 %
2020	175.4	1536.3	11.4 %
2021	148	1353.9	11 %

Анализ основных показателей рынка показывает, что количество страховых компаний на рынке с каждым годом уменьшается, что связано, в основном, с банкротством и отзывом лицензий. По таблице 3 можно сделать вывод, что за 19 лет с 2002 года с рынка ушло более 84% страховых компаний.

Таблица 3 - Количество страховых компаний в РФ

Год	Количество страховых компаний
2004	1416
2005	1280
2006	920
2007	931
2008	929
2009	850
2010	803
2011	759
2012	662
2013	594
2014	569
2015	505
2016	376
2017	313
2018	275
2019	256
2020	233
2021	226

Убыточность по рынку добровольного автострахования является очень высокой. В 2020 году, например, совокупные выплаты по договорам страхования КАСКО составили 98 млрд. рублей. В таблице 4 приведена динамика выплат страховых компаний по договорам добровольного автострахования.

Таблица 4 - Страховые выплаты по КАСКО

Год	Выплаты
2002	6.8
2003	12.5
2004	25.6
2005	42.5
2006	56.6
2007	89.6
2008	102.3
2009	95.6
2010	103.1
2011	109.6
2012	125.7
2013	155.8
2014	171.8
2015	144
2016	97.5
2017	83.3
2018	83.2
2019	97.2
2020	98
2021	81

На представленном графике наблюдается похожие тенденции, что и у совокупных страховых взносов.

Возвращаясь к таблице 1, можно сделать вывод, что количество продаж новых автомобилей, как и предполагалось, имеет ту же динамику, что и страховые взносы и выплаты по КАСКО.

Во **втором** разделе представлен обзор методов построения классов значений факторов риска.

Количество учитываемых факторов риска и их возможных значений должно, с одной стороны, обеспечивать достаточную однородность объектов в каждой группе, а с другой - не сильно затруднять практическое применение системы. Их увеличение ведет к сокращению числа объектов в каждом классе и может вызвать ухудшение точности статистических оценок. Поэтому страховщику при тарификации страхового продукта приходится искать компромисс между степенью однородности объектов и количеством групп.

Если множество значений фактора велико, то уже из соображений простоты тарифной сетки желательно сократить его, объединив отдельные значения в классы. С точки зрения тарификации объединять следует значения со схожим влиянием на процесс убытков. Эта задача относительно проста в случае метрически или ординально шкалируемых факторов (например, объем двигателя или страховая сумма) - у них соседние значения, как правило, обладают и схожим потенциалом убытков.

Когда же фактор с большим числом значений шкалирован номинально (географические регионы, или области почтовых индексов, или различные виды промышленных предприятий), построение классов становится непростой задачей.

Но, в целом, речь идет о задаче, схожей с рассматриваемой в кластер-анализе. Это позволяет взять за основу методы кластер-анализа. В общем случае кластер-анализ предназначен для объединения некоторых объектов в классы (кластеры) таким образом, чтобы в один класс попадали максимально схожие, а объекты различных классов максимально отличались друг от друга. Количественный показатель сходства рассчитывается заданным способом на основании данных, характеризующих объекты.

Методы кластерного анализа применяются всякий раз, когда необходи-

мо преобразовать «горы» информации в наглядные структуры, т.е. сократить число значений номинально шкалированного фактора (марка автомобиля, род предприятия, почтовый индекс, географический регион и т.д.).

Агломеративные методы (англ. *agglomerative*) создают новые кластеры путем объединения более мелких кластеров. Эта группа методов характеризуется последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров. В агломеративных кластер-методах задается мера сходства или расстройния, и каждый объект (значение фактора) первоначально является отдельным классом, затем последовательно объединяются классы наиболее близко находящиеся друг к другу относительно введенной меры расстояния. Результат зависит в большинстве своем от выбора меры расстояния. В разделе рассматривается агломеративный кластер метод на основе критерия равенства математических ожиданий и агломеративный кластер-метод с максимизацией функции правдоподобия, данные методы необходимо применять для уменьшения количества классов значений с целью облегчения дальнейшего анализа этих факторов для выявления тарифных факторов, построения множителей для выбранных факторов и т.д.

Введем необходимые далее обозначения:

Рассмотрим сначала однородную группу рисков, где все риски $i = 1, \dots, I$ независимы и имеют одинаковые распределения суммарного (годового) убытка R_i . Введя обозначения

$$m = E(R_i), s^2 = Var(R_i), 1 \leq i \leq I$$

(где E - мат ожидание, Var - дисперсия). Для совокупного убытка группы рисков

$$S = \sum_{i=1}^I R_i$$

получим

$$E(s) = I \cdot M, Var(s) = I \cdot s^2$$

Вместо зависящей от объема величины «совокупный убыток» удобнее использовать величину нормированную на объем - убыток на один полисо-год(или по другому "ставка убытка")

$$Z = S/I,$$

для которой

$$E(Z) = m, Var(Z) = s^2/I$$

Агломеративный кластер-метод на основе критерия равенства математических ожиданий. Обозначим через Z_{ij} нормированный на объем совокупный убыток i -го значения фактора, $1 \leq i \leq I$, в j -м году наблюдения, $1 \leq j \leq J$. Предположим, что СВ Z_{ij} независимы и имеют логнормальные распределения с параметрами θ_i и σ_i^2/ν_{ij} , где объем ν_{ij} (число полисо-лет) известен. Тогда СВ

$$W_{ij} = \ln(Z_{ij})$$

распределена нормально и

$$E(W_{ij}) = \theta_i, 1 \leq j \leq J,$$

$$Var(W_{ij}) = \frac{\sigma_i^2}{\nu_{ij}}.$$

Модифицированная с учетом этих условий статистика t -критерия для проверки равенства средних θ_i и θ_k соответственно i -го и k -го значений фактора имеет вид:

$$T^2(i, k) = \frac{(\underline{W}_{i+} - \underline{W}_{k+})^2}{1/\nu_{i+} + 1/\nu_{k+}} \cdot \frac{2(J-1)}{SS_i + SS_k},$$

где

$$\underline{W}_{i+} = \sum_{j \geq 1} \nu_{ij} W_{ij} / \nu_{i+}$$

$$\nu_{i+} = \sum_{j \geq 1} \nu_{ij},$$

$$SS_i = \sum_{j \geq 1} \nu_{ij} (W_{ij} - \underline{W}_{i+})^2.$$

\underline{W}_{k+} , ν_{k+} и SS_k определяются аналогично. При справедливости основной гипотезы: $\theta_i = \theta_k$ и $\sigma_i = \sigma_k = \sigma$ разность $\underline{W}_{i+} - \underline{W}_{k+}$ как и в случае обычного t-критерия, распределена нормально с математическим ожиданием 0 и дисперсией $\sigma^2(1/\nu_{i+} + 1/\nu_{k+})$. Величина

$$(SS_i + SS_k)/\sigma^2$$

не зависит от $\underline{W}_{i+} - \underline{W}_{k+}$ и при условии $\sigma_i = \sigma_k = \sigma$ распределена как хи-квадрат с $2(J - 1)$ степенями свободы. Тогда статистика

$$T(i, k) = \text{sign}(\underline{W}_{i+} - \underline{W}_{k+}) \cdot \sqrt{T^2(i, k)}, \quad (1)$$

где sign - знак выражения, при справедливости основной гипотезы имеет t -распределение с $2(J - 1)$ степенями свободы (соответственно, $T^2(i, k)$ имеет F -распределение с 1 и $2(J - 1)$ степенями свободы). Применение этого критерия разрешается только при наличии по каждому фактору не менее $J = 5$ наблюдений совокупного убытка. Предположение равенства $\sigma_i = \sigma_k$ должно быть проверено с помощью F-критерия. Поскольку величины SS_i/σ_i^2 и SS_k/σ_k^2 независимы и распределены как хи-квадрат с $J - 1$ степенями свободы, статистика SS_i/SS_k в случае $\sigma_i = \sigma_k = \sigma$ имеет F-распределение с $J - 1$ и $J - 1$ степенями свободы. Слишком высокое или слишком низкое значение

$$SS_i/SS_k \quad (2)$$

заставляет отвергнуть гипотезу $\sigma_i = \sigma_k$ и отказаться от применения t-критерия к паре (i,k). В этом случае неправильно объединять i-е и k-е значения фактора в один класс.

Алгоритм кластеризации:

- (1) Задаем вероятность попадания в допустимую область при справедливости основной гипотезы, например 95 %.
- (2) Фиксируем произвольный класс i . ($1 \leq i \leq I$)
- (3) Считаем значение статистики $T(i,k)$ (2.1), для каждого класса k ($1 \leq$

$k \leq I, k \neq i$), при условии, что F-критерий(2.2) не отвергает гипотезу $\sigma_i = \sigma_k$.

- (4) Находим среди всех полученных $T(i,k)$ значение наиболее близкое к 0. Если не нашлось ни одного класса k , допускаемого F-критерием (2.2), то класс i нельзя объединить ни с одним другим классом, следовательно, мы возвращаемся к шагу (2), предварительно записав класс i в таблицу результатов.
- (5) Объединяем класс i с классом k (складываем наблюдения совокупных убытков по каждому году классов i и k , и их объемы в каждом году). Получаем годовые величины убытков нового класса:

$$Z_{i \cup k, j} = (\nu_{ij} Z_{ij} + \nu_{kj} Z_{kj}) / (\nu_{ij} + \nu_{kj})$$

- (6) Проверяем устойчивость нового класса с помощью F-критерия(2.2) и T- критерия(2.1), если в нем более 2 значений фактора.

(6.1) Класс устойчив - Возвращаемся к пункту (2), в качестве фиксированного класса берем новообразованный.

(6.2) Класс неустойчив - возвращаемся к пункту (4), исключив $T(i,k)$ из списка полученных значений T-критерия (под списком полученных значений понимаются все результаты $T(i,k)$ в пункте(3)).

Пункты (2)-(6) повторяются до тех пор, пока все классы не будут исключены из итерации. В результате применения метода получим новое разбиение по классам значений.

Третий раздел посвящен методам отбора тарифных факторов. Задача деления на тарифные группы является достаточно сложной и не всегда легко поддается формализации. При ее решении используются различные методы статистического анализа данных, цель которых - выявить факторы, влияющие на степень риска. Для этого оценивается связь (корреляции) между различными физическими (наблюдаемыми) характеристиками объектов и параметрами риска. В результате проведенного анализа страховщик получает перечень факторов, которые с высокой степенью достоверности определяют вероятность наступления страхового случая и (или) ожидаемую величину убытка.

По результатам анализа выделяют некоторый набор независимых наи-

более значимых факторов, которые в дальнейшем будут учитываться при расчете тарифов. Их называют «тарифными факторами». Объекты, имеющие одинаковые значения данных показателей, будут образовывать однородную с точки зрения риска группу или класс.

Количество факторов риска может оказаться довольно большим, но между ними часто существует взаимосвязь, и приходится следить, чтобы одно и то же обстоятельство не учитывалось в тарифе несколько раз. В разделе рассмотрены: **схема пошагового отбора**, на основе которой строятся большинство методов выбора тарифных факторов; **метод отбора тарифных факторов с помощью дисперсионного анализа**, в рамках которого проверяется гипотеза равенства математических ожиданий для каждого фактора, и на основе значения статистики принимается решение о выделении какого-либо фактора в качестве значимого(тарифного), также метод учитывает влияние каждого следующего фактора на уже имеющуюся систему; **метод отношения правдоподобий**, который отличается от метода выбора с помощью дисперсионного анализа тем, что может использоваться на основе любого распределения, а не только логнормального распределения.

Четвертый раздел посвящен построению тарифных множителей для страхования КАСКО в зависимости от региона использования транспортного средства. Построение множителя произведено на основе данных о страховых выплатах по добровольному страхованию в 74 субъектах РФ. Для анализа данных было написано консольное приложение на языке программирования C++, реализующее агломеративный кластер-метод, основанный на критерии равенства математических ожиданий. В результате анализа статистики выплат по КАСКО удалось существенно сократить количество классов значений такого номинально шквалированного как «регион использования транспортного средства» с 74 до 33 классов значений, построить систему тарифных множителей.

В **заключении** приведены результаты бакалаврской работы.

Основные результаты

1. Проанализирован рынок автострахования Российской Федерации с I квартала 2002 года по III квартал 2021 года, а по большей части рынок

добровольного страхования КАСКО.

2. Изучены методы построения классов значений факторов риска.

3. Дано определение тарифных факторов, а также рассмотрены методы их отбора.

4. Написано консольное приложение на языке C++, фрагмент которого представлен в приложении А. Разработанная программа позволяет строить новые классы значений для конкретного фактора риска. С помощью этого приложения произведен анализ статистики выплат по добровольному автострахованию в 74 субъектах РФ. На основе результатов работы программы были подсчитаны тарифные множители при использовании схемы с маргинальными множителями.