
Статистический анализ анкет клиентов гостиниц

*Ветохин А.Н.,
к. ф-м.н., доцент ИТиГ
(филиал РГУТиС),
Силаева И.В.,
доцент ИТиГ (филиал РГУТиС)*

В статье рассматриваются непараметрический статистический критерий анализа категориальных данных, а также возможности применения этого критерия для статистического анализа анкет клиентов гостиниц. Предлагаемые авторами задачи в полной мере учитывают специфику туристской отрасли и наглядно иллюстрируют указанный критерий.

Ключевые слова: статистика туризма, статистический анализ, непараметрический статистический критерий.

The authors look into the distribution-free test of the categorical data analysis and, besides, into an opportunity of applying this criterion to the statistic analysis of hotel guests questionnaires. The tools offered by the authors abide by the tourism industry specifics and illustrate the given criterion.

Key words: Tourism Statistics, Statistic Analysis, Distribution-Free Test.

Усиление прикладного характера изучаемых дисциплин и их тесная связь с индустрией гостеприимства являются одними из важнейших задач совершенствования туристского образования в современных условиях. В связи с этим представляется актуальным внесение изменений и дополнений в программы дисциплин «Статистика» и «Статистика туризма» для студентов, обучающихся в профильных ВУЗах. В этой работе мы подробно остановились на различных аспектах применения критерия Пирсона для статистического анализа анкет клиентов гостиниц. Считаем, что этот критерий может быть предложен для изучения в курсе «Статистика» (специальности 080502 – Экономика и управление на предприятии и 080507 – Менеджмент организации).

Исходя из личного опыта преподавания в высшей профессиональной школе, считаем, что курс статистики является одним из наиболее сложных для студентов, многих из которых пугают математические вычисления и довольно громоздкие термины и формулы. Для преодоления некоторого предубеждения студентов преподаватель должен показывать преимущества, которые дает знание и пра-

вильное использование статистических методов в бизнесе, причем на конкретных примерах.

Статистическое образование студентов экономистов и менеджеров в туристском ВУЗе предполагает подробную иллюстрацию статистического анализа данных на примерах из сферы туризма и гостеприимства. Между тем, при наличии значительного количества литературы по статистике учебники, которые рассматривают особенности и возможности использования статистических методов в туризме и гостиничном бизнесе, практически отсутствуют. Авторами подготовлена серия статей в этом направлении, которые, надеемся, будут полезны как студентам, изучающим статистику, так и преподавателям профильных ВУЗов.

В условиях постоянно растущей конкуренции специалистам, работающим в гостиничном бизнесе, приходится ежедневно находить ответы на непростые вопросы. Как оценить качество обслуживания клиентов гостиницы и степень удовлетворенности их запросов и

ки качественных данных, позволяет выявить существующие проблемы уровня обслуживания и сравнить его с тем, что предлагают конкуренты.

Изучив анкеты нескольких гостиниц, отметим, что все анкеты содержат вопрос «Остановитесь ли Вы в нашей гостинице в следующий раз?» с вариантами ответа «Да» и «Нет». Утвердительный ответ на этот вопрос свидетельствует об общей удовлетворенности запросов гостей.

Рассмотрим процедуру, которая на основе анализа анкет гостей трех гостиниц позволяет выявить разницу в степени их удовлетворенности (качестве обслуживания) и, после изучения жалоб клиентов, дает возможность найти причины недовольства.

Пример 1.

Результаты ответов гостей гостиниц 1, 2 и 3 на вопрос: «Остановитесь ли Вы в нашей гостинице в следующий раз?» – оформляются в виде таблицы, которая называется таблицей сопряженности. Отметим, что основная часть

Таблица 1

Таблица сопряженности признаков для оценки качества обслуживания клиентов гостиниц				
Остановитесь ли Вы в нашей гостинице в следующий раз?	Гостиница 1	Гостиница 2	Гостиница 3	Итого
	f_1	f_2	f_3	
Да	165	193	140	498
Нет	90	76	94	260
Итого	255	269	234	758

ожиданий? Как ранжировать требования гостя к отелю, выявить причины недовольства отдельных клиентов и оценить вероятность того, что гость снова приедет в отель?

Оценка эффективности работы гостиниц включает в себя изучение как количественных показателей, так и качественных характеристик. В настоящее время практически в любой гостинице проводится анкетирование гостей с целью осуществления контроля качества обслуживания. Анализ анкет, который опирается на статистические методы обработ-

таблицы содержит две строки и три столбца.

Приступая к изучению статистических методов, необходимо понимать, что в основе статистики вывода лежит вероятность. Методы статистического исследования используют выборочные данные, поэтому статистические оценки являются только предположениями, они показывают величину неопределенности или ошибки с некоторой вероятностью.

В данной работе будет использоваться критерий χ^2 (хи-квадрат), который в 1900-ом

году был предложен английским математиком Карлом Пирсоном.

Прежде чем приступить к дальнейшему решению задачи, нам необходимо выдвинуть две взаимоисключающие гипотезы. Гипотеза формулируется как определенное утверждение, которое может быть верным или неверным. Гипотезы проверяются с помощью статистических критериев. Статистическая гипотеза, подлежащая проверке, называется нулевой и обозначается H_0 .

В нулевой гипотезе выдвигаются предпо-

$$d_{nem} = 1 - d_{oa} = 1 - 0,657 = 0,343 \text{ (34,3 \%)}$$

Умножая эти две доли на количество опрошенных постояльцев гостиниц, мы получим значения теоретических частот для ответивших «да»:

$$f_1^T = d_{oa} * \Sigma f_1 \quad f_2^T = d_{oa} * \Sigma f_2$$

$$f_3^T = d_{oa} * \Sigma f_3$$

и для ответивших «нет»:

$$f_1^T = d_{nem} * \Sigma f_1 \quad f_2^T = d_{nem} * \Sigma f_2$$

$$f_3^T = d_{nem} * \Sigma f_3$$

На основе полученных таблиц с эмпирическими f_1, f_2, f_3 и теоретическими f_1^T, f_2^T, f_3^T

Таблица 2				
Таблица теоретических частот для оценки качества обслуживания клиентов гостиниц				
Теоретические частоты	f_1^T	f_2^T	f_3^T	Итого
Да	167,53	176,73	153,74	498
Нет	87,47	92,27	80,26	260
Итого	255	269	234	758

ложения относительно одного или нескольких параметров исходной совокупности. Альтернативная гипотеза противоположна нулевой гипотезе и обозначается H_1 . Если в нулевой гипотезе утверждается, что между долями совокупностей, из которых были извлечены выборки, различия не существует, то альтернативная гипотеза утверждает, что такое различие есть.

Сформулируем нулевую гипотезу: доли постояльцев гостиниц 1, 2 и 3, которые планируют остановиться в гостинице в следующий раз, одинаковы. Найдем теоретические частоты f_1^T, f_2^T и f_3^T , которые рассчитываются, исходя из сформулированной нулевой гипотезы об отсутствии связи между признаками группировки.

Если нулевая гипотеза является истинной, то доля гостей, которые собираются остановиться в гостинице в следующий раз, одинакова для каждой гостиницы и равна 65,7%:

$$d_{oa} = \frac{498}{758} = 0,657$$

Доля гостей, оставшихся недовольными обслуживанием, вычисляется следующим образом:

значениями вычислим тестовую характеристику χ^2 , которая показывает различия между двумя таблицами.

$$\chi^2 = \sum \frac{(f_i - f_i^T)^2}{f_i^m}$$

Вычисления χ^2 -статистики оформим в виде таблицы:

Таблица 3				
f	f^T	$f - f^T$	$(f - f^T)^2$	$(f - f^T)^2 / f^T$
165	167,53	-2,53	6,4009	0,038
193	176,73	16,27	264,7129	1,498
140	153,74	-13,74	188,7876	1,228
90	87,47	2,53	6,4009	0,073
76	92,27	-16,27	264,7129	2,869
94	80,26	13,74	188,7876	2,352
Итого	-	-	-	$\chi^2 = 8,058$

Значение χ^2 является суммой значений всех ячеек последнего столбца таблицы. Это значение сравниваем с критическим значением $\chi^2_{крит}$, которое определяется исходя из

χ^2 -распределения с числом степеней свободы $\nu = (n - 1) * (m - 1)$, где n – число строк и m – число столбцов в таблице сопряженности. Значение $\chi^2_{\text{крит}}$ можно найти по статистическим таблицам или с помощью компьютера, создав рабочий лист EXCEL, а затем применив встроенную статистическую функцию ХИ2ОБР (уровень значимости α ; степени свободы ν). Если расчетное значение χ^2 удовлетворяет неравенству $\chi^2 < \chi^2_{\text{крит}}$, то нулевую гипотезу отклонить нельзя.

При проверке статистических гипотез существует риск прийти к неверным выводам, допустить ошибку. Ошибка возникает, когда отклоняется истинная нулевая гипотеза. Вероятность такой ошибки называется уровнем значимости критерия и обозначается через α . В статистических исследованиях чаще всего пользуются тремя уровнями значимости: $\alpha = 0,10$; $\alpha = 0,05$; $\alpha = 0,01$. Мы применяем уровень значимости $\alpha = 0,05$, т. е. в нашем случае вероятность сделать ошибочный вывод составляет 5%.

Сделаем вывод: при уровне значимости $\alpha = 0,05$ и числе степеней свободы $(2-1)*(3-1) = 2$ критическое значение статистики $\chi^2_{\text{крит}} = 5,991$, расчетное значение $\chi^2 = 8,058$, расчетное значение превосходит критическое, поэтому нулевая гипотеза, утверждающая, что

цами есть статистически значимая разница в обслуживании гостей, причем в гостинице 2 клиенты обслуживаются лучше, чем в гостиницах 1 и 3. Такой вывод можно сделать на основе сравнения соответствующих долей клиентов в каждой из гостиниц: среди клиентов 1-ой гостиницы 165 человек из 255 (или 64,71 %) собираются остановиться в гостинице в следующий раз, во 2-ой гостинице таких клиентов 193 человека из 269 (или 71,75%), а в третьей – 140 из 234 (или 59,83 %).

Отметим, что существуют два методических аспекта применения критерия χ^2 . В описанной выше ситуации рассматривался один фактор (удовлетворенность обслуживанием) и две категориальные переменные: да и нет. Возможно использование критерия χ^2 при рассмотрении двух факторов, например, для выяснения причин недовольства клиентов гостиниц. Первым фактором здесь является причина недовольства, а вторым – отель, причем каждый из факторов имеет несколько уровней.

Пример 2.

Предположим, что в ходе опроса клиентов, которые заявили о нежелании останавливаться в гостинице в следующий раз, просили указать причину недовольства. Результаты ответов представлены в таблице.

Таблица 4

Таблица сопряженности признаков для выяснения причин недовольства клиентов гостиниц				
Причина недовольства	Гостиница 1	Гостиница 2	Гостиница 3	Итого
	f_1	f_2	f_3	
Расположение гостиницы	14	19	38	71
Удобства номера	15	15	8	38
Качество обслуживания	25	32	30	87
Качество питания	36	10	18	64
Итого	90	76	94	260

между числовыми значениями долей клиентов, удовлетворенных обслуживанием в обеих гостиницах, нет существенной разницы, должна быть отклонена. Существуют основания утверждать, что между тремя гостини-

Формулируем нулевую гипотезу: между недовольством гостей и фактором гостиницы взаимосвязи нет. Вычислим теоретические частоты. Если нулевая гипотеза является истинной, то доля гостей, которые недовольны

расположением гостиницы, одинакова для каждой гостиницы и равна 27,3%

$$d_1 = \frac{71}{260} = 0,273$$

Доля гостей, которые в качестве причины указали неудобные номера, равна 14,6%

$$d_2 = \frac{38}{260} = 0,146$$

Аналогично вычисляются доля гостей, которые недовольны качеством обслуживания $d_3 = 33,5\%$ и доля гостей, указавших в качестве причины недовольства питание, $d_4 = 24,6\%$.

Теоретические частоты получаем, умножая доли на число гостей, недовольных обслуживанием в каждой из гостиниц.

Вычисления χ^2 -статистики представлены в следующей таблице.

<i>Таблица 5</i>				
f	f^t	$f - f^t$	$(f - f^t)^2$	$(f - f^t)^2 / f^t$
14	24,57	-10,57	111,7249	4,552
19	20,75	-1,75	3,0625	0,148
38	25,66	12,34	152,2756	5,934
15	13,14	1,86	3,4596	0,263
15	11,1	3,9	15,21	1,370
8	13,72	-5,72	32,7184	2,385
25	30,15	-5,15	26,5225	0,880
32	25,46	6,54	42,7716	1,68
30	31,49	-1,49	2,2201	0,071
36	22,14	13,86	192,0996	8,677
10	18,69	-8,69	75,5161	4,040
18	23,13	-5,13	26,3169	1,138
Итого	–	–	–	$\chi^2 = 31,138$

Число степеней свободы в данном случае равно $(4-1) \cdot (3-1) = 6$, при уровне значимости $\alpha = 0,05$ и числе степеней свободы 6 находим из таблицы критическое значение статистики $\chi^2_{\text{крит}} = 12,592$, расчетное значение $\chi^2 = 31,138$. Критическое значение меньше расчетного, поэтому

нулевая гипотеза об отсутствии связи между причиной недовольства гостей и конкретными гостиницами отклоняется. Между причинами недовольства гостей и характеристиками гостиниц существует связь. Сравнивая эмпирические и теоретические значения можно отметить, что клиенты гостиницы 1 более всего недовольны питанием, гостиницы 2 – качеством обслуживания, а клиентов гостиницы 3 больше всего не устраивает расположение гостиницы.

Отметим, что применение критерия χ^2 Пирсона дает достаточно точные результаты при объеме выборки больше 50 и в том случае, когда все теоретические частоты больше или равны 5.

Рассмотренный в статье критерий χ^2 относится к так называемым непараметрическим критериям, которые не требуют, чтобы исходные данные были распределены по тому или иному закону. Этот критерий успешно применяется в медицине, биологии, педагогике, психологии. Критерий χ^2 позволяет сделать вывод в отношении качественных данных, которые представлены категориями, поэтому активно используется при анализе результатов опросов и анкетирования.

В заключение предлагаем задачи для самостоятельного решения:

Задача 1.

С целью изучения качества обслуживания постояльцев трех отелей было проведено анкетирование. В таблице сопряженности признаков приведены ответы постояльцев на вопрос: «Довольны ли Вы качеством уборки номера?». Вычислить χ^2 -статистику. Можно ли утверждать, что существует статистически значимая разница между степенью удовлетворенности клиентами отелей качеством уборки номеров, если уровень значимости равен 0,05? Сделать выводы.

Довольны ли Вы качеством уборки номера?	Отель 1	Отель 2	Отель 3
Да	93	124	75
Нет	40	103	82

Задача 2.

В санаторно-курортном учреждении было проведено исследование отдыхающих с целью изучения влияния наследственности на возникновение гипертонической болезни. В таблице сопряженности признаков приведены результаты исследования. Можно ли утверждать, что существует статистически значимая связь между наследственностью и наличием гипертонической болезни, если уровень значимости равен 0,05?

Кто-либо из родителей болен гипертонией	Состояние отдыхающих	
	Больны гипертонией	Здоровы
Да	25	22
Нет	40	73

Задача 3.

Среди клиентов гостиницы был проведен опрос с целью выяснения наличия связи между полом туриста и его выбором желаемого типа завтрака. Гостей просили ответить на вопрос: «Какой тип завтрака Вы предпочитаете?» Таблица сопряженности признаков, содержащая результаты опроса, приведена ниже.

Пол	Какой тип завтрака Вы предпочитаете?	
	Шведский стол	Заказ по меню
Мужской	75	42
Женский	30	70

Литература

1. Теория статистики. Учебник под ред. Г.Л. Громыко. М.: ИНФРА-М, 2005.

Можно ли утверждать, что между полом туриста и предпочитаемым им завтраком существует статистически значимая связь, если уровень значимости равен 0,05? Сделать выводы.

Задача 4.

Среди пассажиров авиакомпании было проведено анкетирование. Ответы на вопрос: «Как Вы проводите время в полете?» - приведены в таблице сопряженности признаков.

Как Вы проводите время в полете?	Возраст		
	до 30	30-50	старше 50
Читаю журналы и книги	34	78	40
Сплю, отдыхаю	18	51	48
Общаюсь с попутчиками	55	66	30

Можно ли утверждать, что между возрастом авиапассажира и способом проведения свободного времени в полете существует статистически значимая связь, если уровень значимости равен 0,05? Сделать выводы.

Таким образом, применение данной методики позволяет глубоко и системно проанализировать отношение гостей к качеству гостиничных продуктов и услуг. Вместе с тем данный анализ основывается на вполне объективных и измеримых критериях.

2. Левин Д., Стефан Д., Кребиль Т., Беренсон М. Статистика для менеджеров с использованием Microsoft Excel. М.: Издательский дом «Вильямс», 2005.