

*Аникин Данила Алексеевич, студент бакалавриата 2 курс,
МИРЭА-Российский технологический университет (РТУ МИРЭА)*

*Свищёв Андрей Владимирович, старший преподаватель кафедры
практической и прикладной информатики*

МИРЭА-Российский технологический университет (РТУ МИРЭА),

Россия, г. Москва, Институт информационных технологий, Россия, г. Москва

МЕТОД ВЫБОРА СТАТИСТИЧЕСКОГО КРИТЕРИЯ ДЛЯ ПРОВЕДЕНИЯ А/В ТЕСТИРОВАНИЯ

Аннотация: Данная статья посвящена вопросу выбора статистического критерия для проведения А/В тестирования. Корректный и безошибочный А/В тест способен уберечь компании от потери средств из-за потенциального отказа от продукта, для этого необходимо верно составить условия тестирования и проанализировать полученные данные. В результате исследования статистических методов анализа данных, были выявлены рекомендации по принятию решения, какой из критериев следует выбрать в зависимости от специфики данных, полученных в ходе проведения теста.

Ключевые слова: прикладная статистика, математическая статистика, А/В тест, статистические критерии, анализ данных.

Abstract: This article is devoted to the issue of choosing a statistical criterion for conducting A / B testing. A correct and error-free A / B test can save a company from losing funds due to a potential rejection of the product, for this it is necessary to correctly draw up testing conditions and analyze the data obtained. As a result of the study of statistical methods of data analysis, recommendations were identified for deciding which of the criteria should be selected depending on the specifics of the data obtained during the test.

Key words: applied statistics, mathematical statistics, A / B test, statistical criteria, data analysis.

В настоящее время любое изменение, принятое руководством крупной компании, может стать определяющим в её дальнейшей судьбе: верно принятые решения способны смести всех конкурентов и принести невиданную ранее прибыль, однако существует риск принятия неверных решений, которые способны привести к полному разорению, оттоку потребителей и банкротству. Есть ли возможность обезопасить компанию в условиях современной конкуренции, когда пользователь продукта всегда может найти аналог всему, начиная с продуктов питания, заканчивая информационными сервисами?

Такая возможность есть, и она заключается в проведении предварительного тестирования. Именно в информационной среде существует широкое поле возможностей для сбора реакции пользователей на то или иное изменение в продукте, к примеру, при изменении дизайна сайта можно проследить время, которое пользователь стал на нём проводить, как активно стали пользоваться платным функционалом при изменении условий и много другое, в мире неинформационного бизнеса собрать показатели такого рода гораздо сложнее, поэтому далее в статье будет рассматриваться вопрос тестирования именно в контексте информационных технологий. Существует несколько методик тестирования: многовариантный тест, метод многоруких бандитов, но наиболее распространенный и общеизвестный вид тестирования программных продуктов – это A/B тест.

A/B тест, также известный как сплит тест (от английского split testing) – это метод исследования пользовательского опыта использования продукта. A/B тестирование состоит из следующих этапов: исследование, выдвижение гипотезы, тестирование и анализ полученных результатов. Рассмотрим каждый из них более внимательно [1].

Прежде чем проводить какие-либо изменения в продукте, необходимо понять, нужны ли они вообще, именно за это отвечает этап исследования. В первую

очередь, необходимо определить цели, которые стоят перед продуктом, и ключевые показатели эффективности. Стоит проанализировать персональный пользовательский опыт использования продукта, понять, какие изменения действительно необходимы и принесут пользу.

После локализации проблемных зон продукта необходимо выдвинуть гипотезу, каким способом улучшить те или иные показатели. Этот этап очень важен, так как цель тестирования – подтвердить или опровергнуть гипотезу, следовательно, необходимо выдвинуть её изначально. Гипотеза должна быть измеримой и поддаваться проверке. Также нужно предположить, как именно изменения должны сказаться на ключевых показателях. Следует отметить, что за одно тестирование проверяют только одну гипотезу, так как большее количество даст результаты, на основании которых невозможно сделать однозначные выводы.

На этапе проведения тестирования нельзя обделить вниманием процесс выделения должной выборки: если вносимое изменение рискованно, имеет смысл проверить его на небольшом количестве пользователей. Также следует обратить внимание на то, что для составления качественной базы данных необходимо продолжительное время, в течение которого нивелируется большинство факторов, которые могли бы повлиять на результаты.

Наиболее ответственным шагом является заключительный этап – анализ полученных результатов. Он настолько важен из-за возможности обесценить всю предыдущую работу по выявлению проблемы, работы над её решением и проведения тестирования. Также на этапе анализа может быть совершена ошибка в том, что выдвинутая ранее гипотеза верна, но были допущены ошибки при её фактической реализации. На основании ошибочного анализа данных невозможно принять верное решение.

Для анализа результатов тестирования используется статистическая математика, а именно – статистические критерии анализа. В некоторых случаях данные визуализируют для облегчения восприятия, особенно часто подход

визуализации встречается в веб-разработке, где есть уже готовые сервисы для сбора информации о поведении пользователей на сайте.

Статистический критерий – это математическое правило, в соответствии с которым принимается или отвергается та, или иная статистическая гипотеза с заданным уровнем значимости. Выбор критерия зависит от типа данных, полученных в ходе тестирования [5].

Существует три типа данных:

1. Номинативный – данные, которые невозможно измерить, но которые можно описать, это характерные признаки элементов выборки. Допустимая мера среднего: частота.

2. Ранговый – данные, которые представляют собой последовательность, полученную при расположении в порядке возрастания. Допустимые меры среднего: частота и медианное значение. Пример: рейтинг приложения.

3. Количественный – данные, которые получают при измерениях и которые не зависят от субъективной оценки, их можно распределить на интервальной шкале. Допустимые меры среднего: частота, медианное и среднее значение. Пример: время, которое пользователь проводит в приложении.

На основании вышеприведенной классификации необходимо определить, какой тип данных поддается анализу именно в нашем случае. Если мы анализируем параметрические данные, то нам подходят: U-критерий Манна-Уитни, критерий Уилкоксона, критерий Пирсона, критерий Колмогорова-Смирнова. Если анализируем непараметрические данные, тогда подходят: t-критерий Стьюдента, дисперсионный анализ [3].

Далее дефинируем количество выборок, по которым проходит сравнение и выбираем подходящий критерий. Для параметрических данных с двумя выборками: t-критерий Стьюдента. Для параметрических данных с тремя и более выборками: дисперсионный анализ. Для непараметрических данных с двумя выборками подойдут критерий Манна-Уитни, критерий Уилкоксона, критерий

Колмогорова-Смирнова. Для непараметрических данных с тремя и более выборками: критерий Краскела-Уоллиса, критерий Пирсона [2].

На основании выше представленных правил была составлена таблица, упрощающая восприятие этой концепции.

Таблица 1 – Методика выбора статистического метода

Тип анализируемых данных	Параметрические		Непараметрические	
	2	3 и более	2	3 и более
Статистический критерий	t-критерий Стьюдента	Дисперсионный анализ	критерий Манна-Уитни	критерий Краскела-Уоллиса,
			критерий Уилкоксона	
			критерий Колмогорова-Смирнова	

В результате выбора критерия можно провести анализ данных посредством программного обеспечения, сформировать р-значение и на его основе проверить статистическую значимость проведенного эксперимента, иными словами, выяснить, была ли разница между выборками или же различия выборок сформированы по воле случая. Если р-значение меньше уровня значимости, то аналитик отвергает нулевую гипотезу, если больше, то это сделать невозможно. Дальнейший анализ зависит исключительно от конкретной ситуации [4].

По большому счету, из А/В теста можно сделать следующие выводы: наиболее эффективен вариант А, наиболее эффективен вариант В или между вариантами нет разницы. На основании полученного результата принимается решение о модернизации продукта и, благодаря проведенному ранее тестированию, оно не представляет серьезных рисков для компании, потому что реакция пользователей на нововведение уже получена и обработана.

Библиографический список:

1. Кохави Р., Тан Д., Сюй Я. Доверительное А/В-тестирование 2021 г., 298 с.
2. Рекомендации по стандартизации. Прикладная статистика: Правила проверки согласия опытного распределения с теоретическим. Часть II: Непараметрические критерии. — М.: Госстандарт РФ, 2002.
3. Ниворожкина Людмила Ивановна, Арженовский Сергей Валентинович Статистические методы анализа данных. Учебник. Студентам ВУЗов. 2019 г., 333 с.
4. Сара Бослаф. Статистика для всех. 2017 г. 586 с.
5. Хастри Тревор, Тибришани Роберт. Основы статистического обучения: интеллектуальный анализ данных, логический вывод и прогнозирование. 2020 г., 768 с.