

А.И. Азевич

Онлайн-сервисы и статистическая обработка экспериментальных данных

В статье рассматриваются возможности онлайн-сервисов для обработки статистических данных. Описывается дидактический потенциал этих онлайн-инструментов в обучении студентов применению математических методов в исследовательской работе.

Ключевые слова: научная гипотеза; статистические пакеты; статистические онлайн-калькуляторы; базовые статистики; t -критерий Стьюдента; корреляция.

Подготовка выпускной квалификационной работы — процесс кропотливый и длительный. Он включает в себя несколько взаимосвязанных этапов: определение темы исследования, сбор научно- и фактического материала, разработку методологического аппарата, осмысление содержательной структуры, формирование логических линий, анализ эмпирических данных и т. д.

Именно в процессе анализа данных, полученных опытным путем, у студента чаще всего возникает целая цепочка вопросов. Как грамотно обработать статистический материал? Какой критерий выбрать для подтверждения статистической гипотезы? Какие программные средства использовать в ходе работы над дипломом?

Размышляя об ответах на поставленные вопросы, начнем с научной гипотезы. В исследовании она выступает в роли утверждения, в котором, во-первых, формулируется связь между двумя или несколькими переменными; во-вторых, проектируется мост между экспериментальной методикой и результатом ее внедрения в образовательную практику. Надо сказать, что гипотеза не рождается сама по себе. Чаще всего она приходит не как научное озарение, а как результат напряженной мыслительной работы.

Какие этапы необходимо пройти исследователю, чтобы его научная гипотеза обрела статус надежного методологического инструмента?

Первый из них — выявление некоторой группы фактов, не укладывающихся в известные теории. *Второй* — непосредственно формулировка гипотезы, задача которой объяснить обнаруженные факты. *Третий* — глубокое исследование выдвинутого предположения и выделение из него всевозможных следствий. *Четвертый* — сопоставление выделенных из гипотезы следствий с имеющимися наблюдениями, результатами экспериментов, научными законами. *Пятый* — превращение гипотезы в научное знание или теорию, если подтверждаются все выведенные из гипотезы следствия и не возникает противоречий с ранее известными законами науки [8].

Только преодолев все этапы, исследователь сможет осмысленно и точно выразить научную гипотезу. Приведем несколько наиболее удачных формулировок научных гипотез, представленных в дипломных работах.

Пример 1. Развитие любознательности учащихся младшего школьного возраста станет более успешным, если грамотно организовывать совместную деятельность учителя и детей с использованием современных информационно-коммуникационных технологий.

Пример 2. Методическая система обучения педагогическому веб-дизайну может быть построена, если в ее инструментарий включить новые технологии педагогического проектирования.

Пример 3. Использование современных средств информатизации (в частности сервисов web 2.0 и офлайн-приложений для математической и статистической обработки данных) позволит выявить факторы, положительно влияющие на успеваемость старшеклассников.

Итак, предположим, что цели поставлены, проблема определена, гипотеза сформулирована. Опустим несколько этапов исследовательской работы, сразу же перейдя к обработке экспериментальных данных.

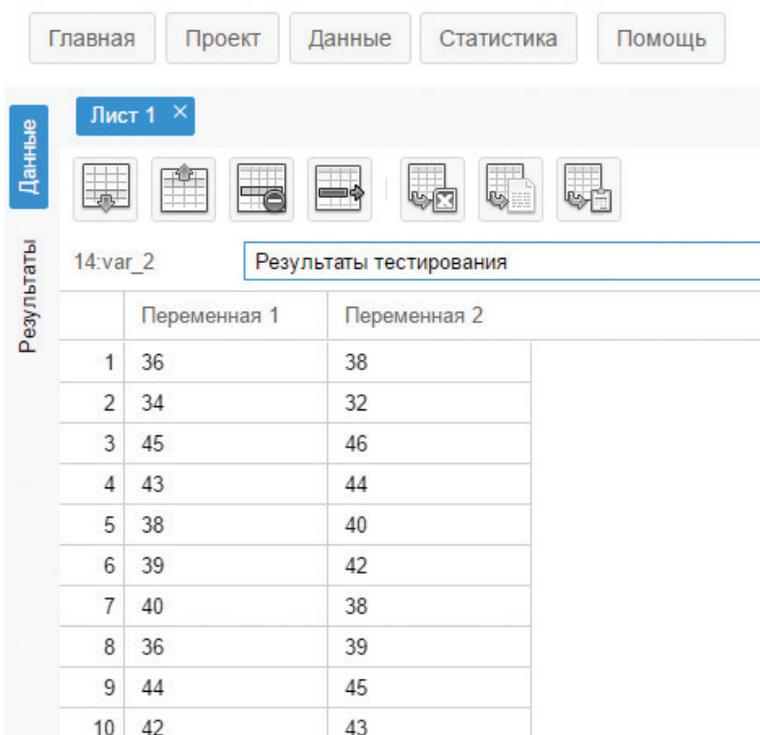
Начнем с программных средств, которые могут быть использованы в ходе подготовки выпускной квалификационной работы. Без них сегодня не обойтись, ведь ручной процесс обработки информации довольно трудоемкий. О каких программах стоит сказать? Это, прежде всего, табличный процессор *MS Excel*, содержащий надстройку «Анализ данных». В нее входит свыше 50 статистических функций. Кроме того, упомянем программу *Stadia* — универсальный пакет, включающий множество функций для обработки статистической информации.

Еще одно эффективное средство для проведения научной аналитики — *SPSS* (Statistical Package for Social Science) — интегрированное семейство продуктов, охватывающих весь процесс анализа данных — от планирования и сбора информации до составления подробных отчетов. *SPSS* отличается гибкостью настроек и множественностью инструментов для решения задач.

Нельзя не упомянуть и о популярной программе *Statistica*. Она включает в себя всевозможные методы анализа, объединенные в статистические модули: Основные статистики и таблицы, Дисперсионный анализ, Множественная регрессия и т. д.

Перечисленные программные пакеты невозможно освоить сразу. Для этого требуется немало времени и сил. Проще обстоит дело со специализированными онлайн-сервисами, которых в Интернете (как в русскоязычном, так и в англоязычном) становится все больше. Они, конечно, не заменят специализированные программные средства, но в некоторых случаях окажут ощутимую помощь студенту и преподавателю.

Чтобы оценить их возможности, рассмотрим один из так называемых *статистических калькуляторов*, находящийся в Интернете по адресу: <http://app.statca.com>. После стандартной процедуры регистрации создадим лист с данными. Кстати, их можно не только сгенерировать на самом сайте, но и загрузить с компьютера. Возможна и обратная процедура — перенос данных с сервиса в табличный процессор MS Excel.



The screenshot shows the web interface of the 'app.statca.com' service. At the top, there are navigation buttons: 'Главная', 'Проект', 'Данные', 'Статистика', and 'Помощь'. Below these is a tab labeled 'Лист 1'. A toolbar contains icons for various statistical functions. The main area shows a data entry field with the label '14: var_2' and the title 'Результаты тестирования'. Below this is a table with 10 rows and 2 columns: 'Переменная 1' and 'Переменная 2'. The table contains the following data:

	Переменная 1	Переменная 2
1	36	38
2	34	32
3	45	46
4	43	44
5	38	40
6	39	42
7	40	38
8	36	39
9	44	45
10	42	43

Рис. 1. Рабочее окно сервиса

На рисунке 1 показано рабочее окно сервиса. В колонках «Переменная 1», «Переменная 2» введены данные тестирования. Предположим, требуется найти основные статистические характеристики случайной величины: минимальное и максимальное значения, размах, медиану, дисперсию, стандартное отклонение и другие сведения. Как это сделать? Очень просто. Достаточно нажать на кнопку *Статистика*, затем в контекстном меню выбрать последовательно команды *Описательная статистика* и *Базовые параметры*. В результате последовательных действий появится таблица с искомыми данными. На рисунке 2 показан ее фрагмент.

Переменная	Размер	Минимальное значение	Максимальное значение	Размах	Сумма	Матожидание	Медиана ряда	Дисперсия	Стандартное отклонение
Переменная 1	14	34	45	11	543	38,786	38,5	14,181	3,7658
Переменная 2	13	32	46	14	522	40,154	40	14,641	3,8284

Рис. 2. Описательная статистика. Базовые параметры

С помощью сервиса можно построить гистограмму частот вариационного ряда (рис. 3). Помимо самой гистограммы одновременно будет построена и кривая нормального распределения, что очень удобно. По рисунку легко сравнить, отличается ли распределение частот исследуемого признака от нормального распределения. Как видим, отличия значительные.

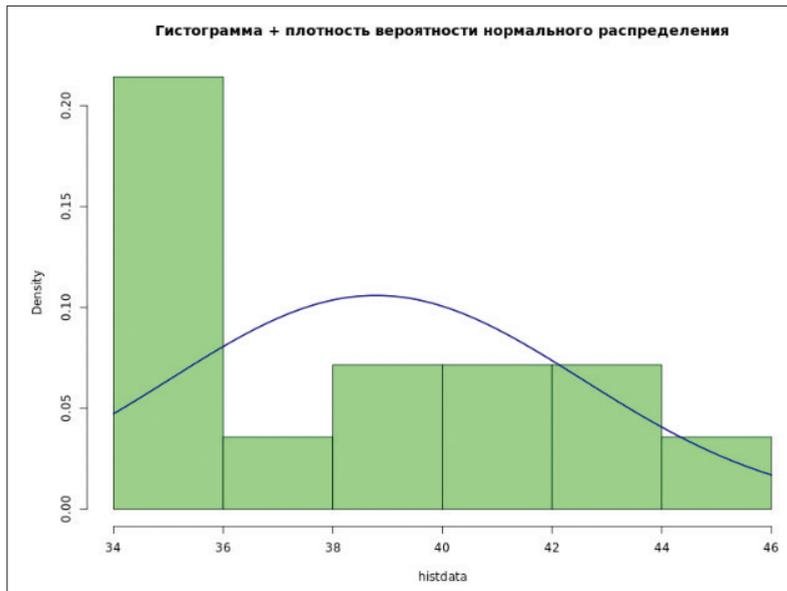


Рис. 3. Гистограмма частот и плотность вероятности нормального распределения

Помимо базовых показателей можно рассчитать и другие. Например, исследовать корреляционную связь между переменными. Для этого вновь нажимаем на кнопки в следующем порядке: *Статистика, Корреляционный анализ, Коэффициент корреляции*. Выполнив указанную цепочку действий, получим таблицу (см. рис. 4). Как видим, корреляционные связи между переменными практически отсутствуют. Это и понятно, ведь массивы чисел взяты совершенно произвольно.

Кoeffициенты



Переменная	Переменная 1	Переменная 2
Переменная 1	1	-0,0031809
Переменная 2	-0,0031809	1

Рис. 4. Коэффициенты корреляции

Когда использовать этот сервис? Бесспорно, он весьма полезен не только на этапе объяснения основных понятий статистики, но и в процессе нахождения тех или иных данных на практическом занятии. С его помощью можно быстро сгенерировать массив случайных чисел, найти базовые характеристики числового ряда и сравнить их с теми же, но уже найденными «вручную».

В ходе работы над дипломом часто требуется найти ошибку среднего, коэффициент вариации и другие показатели случайной величины. К сожалению, на этом ресурсе во вкладке *Базовые статистики* таких функций нет. Тогда стоит воспользоваться другим сервисом, имеющимся на весьма информативном и полезном сайте *Медицинская статистка*. Его адрес в Интернете: <http://medstatistic.ru/index.php>. Рассмотрим методику работы с сервисом.

Войдем в раздел *Калькуляторы*, а затем проследуем по ссылке *Расчет показателей вариационного ряда*. В открывшемся окне укажем объем выборки (пусть он будет равен 30). И уже в новом окне введем значения случайной величины. Нажав на кнопку *Рассчитать показатели*, расположенную в левом нижнем углу страницы, получим необходимые данные (рис. 5). Анализируя их, нетрудно установить, что коэффициент вариации достаточно большой. Это свидетельствует о неравномерности выборки.

Рассчитать показатели

Число единиц наблюдения (n): 30
Средняя арифметическая (M): 38.13
Медиана (Me): 38
Стандартное квадратичное отклонение (σ): 3.64
Коэффициент вариации (Cv): 9.55%
Средняя ошибка средней арифметической (m): 0.66

Рис. 5. Расчет показателей вариационного ряда

Далее изучим, как можно использовать сервис для нахождения статистических критериев, в частности *t-критерия Стьюдента* для независимых выборок. Для этого вернемся к *Примеру 1* и, используя формулировку *научной*

гипотезы, определим соответствующую этому примеру *статистическую гипотезу*. Для этого условно создадим таблицу оценок тестирования учеников контрольной и экспериментальной групп по 10-балльной шкале (табл. 1). Кстати, на этом же сайте можно найти условия выбора того или иного статистического критерия. Это ценная информация в ходе исследовательской работы.

Таблица 1

Оценки тестирования учеников контрольной и экспериментальной групп по 10-балльной шкале

Контрольная группа		Экспериментальная группа	
1	9	1	9
2	8	2	8
3	7	3	6
4	5	4	7
5	6	5	7
6	7	6	7
7	7	7	8
8	8	8	9
9	9	9	10
10	10	10	9
11	5	11	8
12	6	12	5
13	7	13	8
14	4	14	4
15	3	15	5
16	4	16	6
17	5	17	6
18	7	18	9
19	8	19	10
20	7	20	7
21	6	21	7
22	5	22	6
23	6	23	6
24	7	24	9
25	8	25	8
26	9	26	10
27	5	27	7
28	6	28	8
29	7	29	8
30	9	30	10
31	6	31	7

Сформулируем нулевую гипотезу: *различия между средними значениями двух выборок статистически незначимы*. И альтернативную гипотезу: *различия между средними значениями статистически значимы*. Уяснив, какая

из них справедлива, подведем итоги экспериментальной работы. А затем подтвердим эффективность разработанной методики применения информационно-коммуникационных технологий для развития познавательной активности детей или скорректируем ее, продолжив дальнейшие исследования.

Итак, пользуясь данными таблицы, с помощью все того же сервиса найдем значения среднего и ошибку репрезентативности (среднюю ошибку средней арифметической) для каждой группы учащихся (они нужны для определения эмпирического t -критерия Стьюдента). Среднее значение для первой выборки — 6,65, для второй оно равно 7,74 (с точностью до второго знака после запятой). Соответственно ошибка репрезентативности первой выборки равна 0,30, второй — 0,29. Подставляем полученные значения в окно калькулятора и получаем следующие данные (рис. 6).

Эмпирический t -критерий Стьюдента на уровне значимости 0,05 равен 2,61, что превышает соответствующее критическое значение t -критерия, равное 2,0 (это в сервисе, кстати, тоже показано). Следовательно, различия средних контрольной и экспериментальной групп статистически значимы. Верна альтернативная гипотеза. Тем самым можем констатировать, что разработанная методика эффективна.

Введите данные сравниваемых совокупностей

(дробные значения вводятся через точку, например: 2.35)

для расчета значений M и m воспользуйтесь калькулятором "Показатели вариационного ряда"

Средняя величина 1 (M_1):	6.65
Средняя ошибка средней арифметической 1 (m_1):	0.30
Количество измерений в первой группе (n_1):	31
Средняя величина 2 (M_2):	7.74
Средняя ошибка средней арифметической 2 (m_2):	0.29
Количество измерений в первой группе (n_2):	31

Значение t -критерия Стьюдента: 2.61

Различия статистически значимы ($p < 0,05$)

Число степеней свободы $f = 60$

Критическое значение t -критерия Стьюдента = 2, при уровне значимости $\alpha = 0,05$

Рис. 6. Значение t -критерия Стьюдента

Как видим, статистические калькуляторы значительно облегчают вычисления и сокращают время нахождения нужных данных. Они с успехом могут быть использованы как в ходе непосредственной обработки экспериментальных данных, так и на практических занятиях, посвященных использованию математических методов в научных исследованиях.

Литература

1. *Азевич А.И.* Несколько компьютерных программ // Математика в школе. 2002. № 10. С. 44.
2. *Азевич А.И.* Особенности преподавания курса «Использование современных информационных и коммуникационных технологий в обучении» на факультете специальной педагогики // Вестник Московского городского педагогического университета. Серия «Информатика и информатизация образования». 2010. № 19. С. 62–68.
3. *Азевич А.И.* Роль персонального сайта преподавателя в формировании информационной обучающей среды // Современные исследования и инновации. 2012. № 11 (19). С. 28.
4. *Азевич А.И.* Онлайн-сервисы как средство формирования контента сайта преподавателя // Инновации в системе высшего образования: материалы IV Всероссийской научно-методической конференции. Челябинск: Челябинский институт экономики и права им. М.В. Ладощина, 2013. С. 50.
5. *Азевич А.И.* Моделирование средствами MS Excel в деятельности учителя физической культуры // Вестник Российского университета дружбы народов. Серия «Информатизация образования». 2015. № 1. С. 28–34.
6. *Азевич А.И., Сыч С.П.* Формирование ИКТ-компетентности в ходе реализации связей вузовских дисциплин // Вестник Московского городского педагогического университета. Серия «Информатика и информатизация образования». 2010. № 20. С. 73–81.
7. *Азевич А.И., Алексеева С.И.* Педагогический эксперимент и средства описательной статистики // Вестник Российского университета дружбы народов. Серия «Информатизация образования». 2016. № 2. С. 57–65.
8. *Иващенко Ф.И.* Практикум по методологии психологического исследования: учеб. пособие для пед. спецвузов. Минск: ФУАинформ, 2003. С. 26.

Literatura

1. *Azevich A.I.* Neskol'ko komp'yuterny'x programm // Matematika v shkole. 2002. № 10. S. 44.
2. *Azevich A.I.* Osobennosti prepodavaniya kursa «Ispol'zovanie sovremenny'x informacionny'x i kommunikacionny'x tehnologij v obuchenii» na fakul'tete special'noj pedagogiki // Vestnik Moskovskogo gorodskogo pedagogicheskogo universiteta. Seriya «Informatika i informatizaciya obrazovaniya». 2010. № 19. S. 62–68.
3. *Azevich A.I.* Rol' personal'nogo sajta prepodavatelya v formirovanii informacionnoj obuchayushhej sredy' // Sovremenny'e issledovaniya i innovacii. 2012. № 11 (19). S. 28.
4. *Azevich A.I.* Onlajn-servisy' kak sredstvo formirovaniya kontenta sajta prepodavatelya // Innovacii v sisteme vy'sshego obrazovaniya: materialy' IV Vserossijskoj nauchno-metodicheskoy konferencii. Chelyabinsk: Chelyabinskij institut e'konomiki i prava im. M.V. Ladoshina, 2013. S. 50.
5. *Azevich A.I.* Modelirovanie sredstvami MS Excel v deyatel'nosti uchitelya fizicheskoy kul'tury' // Vestnik Rossijskogo universiteta druzhby' narodov. Seriya «Informatizaciya obrazovaniya». 2015. № 1. S. 28–34.
6. *Azevich A.I., Sy'ch S.P.* Formirovanie IKT-kompetentnosti v xode realizacii svyazej vuzovskix disciplin // Vestnik Moskovskogo gorodskogo pedagogicheskogo universiteta. Seriya «Informatika i informatizaciya obrazovaniya». 2010. № 20. S. 73–81.

7. *Azevich A.I., Alekseeva S.I.* Pedagogicheskij e'ksperiment i sredstva opisatel'noj statistiki // Vestnik Rossijskogo universiteta družby' narodov. Seriya «Informatizaciya obrazovaniya». 2016. № 2. S. 57–65.

8. *Ivashhenko F.I.* Praktikum po metodologii psixologicheskogo issledovaniya: ucheb. posobie dlya ped. speczvuzov. Minsk: FUAinform, 2003. S. 26.

A.I. Azevich

Online Services and Statistical Processing of Experimental Data

The article considers the possibilities of online services for processing statistical data. The didactic potential of these online tools in teaching students the use of mathematical methods in research work is described.

Keywords: scientific hypothesis; statistical packages; statistical online calculators; basic statistics; Student's *t*-test; correlation.