

Всероссийское СМИ

«Академия педагогических идей «НОВАЦИЯ»

Свидетельство о регистрации Эл №ФС 77-62011 от 05.06.2015 г.

(выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций)

Сайт: akademnova.ru

e-mail: akademnova@mail.ru

Ершова Д.А. Разработка алгоритмов и приложения для генерации индивидуальных заданий по статистическим методам // Академия педагогических идей «Новация». Серия: Студенческий научный вестник. – 2020. – №7 (июль). – АРТ 91-эл. – 0,2 п.л. - URL: <http://akademnova.ru/page/875550>

РУБРИКА: ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 004.4, 004.9

Ершова Диана Андреевна

студентка 2 курса магистратуры, факультет «Управления и автоматизации»

Научный руководитель: Воробьев Е.С, к.т.н., доцент

ФГБОУ ВО «Казанский национальный исследовательский

технологический университет»

г. Казань, Российская Федерация

e-mail: diana04081996@mail.ru

РАЗРАБОТКА АЛГОРИТМОВ И ПРИЛОЖЕНИЯ ДЛЯ ГЕНЕРАЦИИ ИНДИВИДУАЛЬНЫХ ЗАДАНИЙ ПО СТАТИСТИЧЕСКИМ МЕТОДАМ

Аннотация: В данной статье описаны алгоритмы основных методов статического анализа и предлагается программная реализация генератора индивидуальных заданий для студентов по дисциплинам с данными разделами по статистике.

Ключевые слова: статический анализ данных, дисперсионный анализ, корреляционный анализ, регрессионный анализ, выборка, динамические ряды.

Ershova Diana Andreevna

2nd year student, Faculty of Management and Automation

Scientific adviser: Vorobiev E.S., Ph.D., Associate Professor

FGBOU VPO "Kazan National Research Technological University"

Kazan, Russian Federation

e-mail: diana04081996@mail.ru

DEVELOPMENT OF ALGORITHMS AND APPLICATIONS FOR THE GENERATION OF INDIVIDUAL TASKS ON STATISTICAL METHODS

Resume: This article describes the algorithms of the basic methods of static analysis and be offers a software implementation of the generator of individual tasks for students in disciplines with these sections on statistics.

Keywords: static data analysis, variance analysis, correlation analysis, regression analysis, sample, dynamic series.

В наше время организация дистанционного образования стала актуальной задачей, автоматизированный процесс обучения позволяет студентам непрерывно получать необходимую информацию, не выходя из дома, но автоматически подготавливать индивидуальные задания не всегда возможно. Каждый преподаватель в своей деятельности постоянно готовит и перерабатывает индивидуальные задания студентам для усвоения или укрепления изученных материалов, а также для проверки закрепленных знаний. Для того чтобы облегчить монотонный труд преподавателя по составлению заданий студентам предлагается автоматизированная программа, которая генерирует задания с уникальными значениями и ответы для преподавателя.

Всероссийское СМИ

«Академия педагогических идей «НОВАЦИЯ»

Свидетельство о регистрации Эл №ФС 77-62011 от 05.06.2015 г.

(выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций)

Сайт: akademnova.ru

e-mail: akademnova@mail.ru

Программа генерирует наборы случайных чисел, содержащие определенные закономерности, которые студент должен обнаружить и подтвердить статистическими расчетами, Задания экспортируются в именные файлы Excel и отправляются лично каждому студенту. На основании сгенерированных данных для преподавателя по каждому заданию рассчитываются ответы, которые хранятся в личном кабинете. Программа реализована в Visual Studio на языке программирования C#.

Сначала студенту необходимо авторизоваться в системе, после этого нажать кнопку «Сгенерировать». После того как студент нажал кнопку «Сгенерировать», кнопка перестает быть активной и появляется прилоудер, в котором показывается степень готовности файла к скачиванию, как показано на рисунке 1.

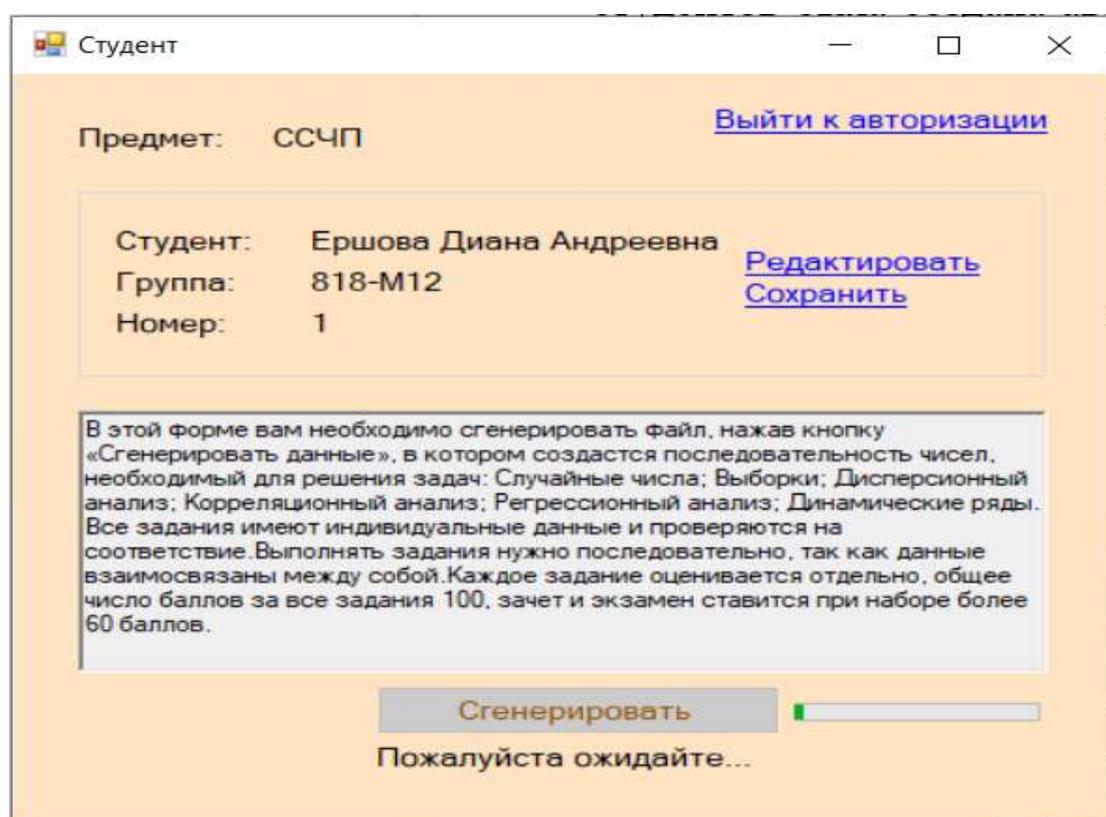


Рисунок 1 – Форма студента

Сгенерированный файл Excel скачивается на личное устройство пользователя, и студент начинает работать с этими данными. Рассмотрим, что собой представляет данный файл и какие задачи на каждом из листов студент должен решить. На рисунке 2 показан титульный лист, который заполирован и является идентификатором файла.

	A	B	C	D	E	F	G	H	I	J	K
1		Рабочая книга по дисциплине									
2		Современные численные методы и пакеты прикладных программ									
3											
4	Студент:	Ершова Диана Андреевна!									
5	Группа:	818-M12									
6	Номер:	1									
7											
8											

Рисунок 2 – Титульный лист задания студента

Далее в файле сгенерированы шесть задач по статическому анализу данных. Рассмотрим их и алгоритмы, по которым они построены.

Первая задача предполагает знакомство студентов с приемами статистической обработки данных средствами Excel и с помощью надстройки «Анализ данных». Программа случайным образом генерируем значения среднего в интервале от 30 до 120 и стандартного отклонения в интервале от 2 до 10, которые используются для генерируемых анализируемых выборок. Пользуясь генератором случайных чисел по нормальному закону распределения генерируем пять столбцов из десяти, двадцати, пятидесяти, ста и тысячи чисел [1]. Полученные выборки помещаем на лист Excel и записываем ответ (среднее и стандартное отклонение) в базу данных для преподавателя. На рисунке 3 можно увидеть, как выглядят сгенерированные данные.

Всероссийское СМИ

«Академия педагогических идей «НОВАЦИЯ»

Свидетельство о регистрации Эл №ФС 77-62011 от 05.06.2015 г.

(выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций)

Сайт: akademnova.ru

e-mail: akademnova@mail.ru

	A	B	C	D	E	F	G	H	I	J	K	L
1	Вам представлены пять наборов случайных чисел с нормальным законом распределения и имеющих одинаковые											
2	среднее значение и стандартное отклонение. Необходимо освоить приемы работы со статистическими функциями											
3	листа и надстройкой "Анализ данных". Оцените на сколько они отличаются друг от друга											
4												
5	5 наборов случайных измерений											
6	Набор 1	Набор 2	Набор 3	Набор 4	Набор 5							
7	99,125	101,125	100,75	100,222	98							
8	101,571	97	103,286	101,875	97,75							
9	98,1111	97,5	96,8571	100,125	98,75							
10	100,857	100	100,333	98	100,75							
11	101,5	98,4286	100,125	99,125	99,5714							
12	102,571	98	99,2222	97,625	98,6667							
13	97,875	101	98,5556	99,4286	97,2222							
14	96,4444	98,2222	99,4444	97,5	101							
15	100,778	97,3333	101,625	101,625	99,625							

Рисунок 3 – Сгенерированные числа для первого метода

Второй задачей является механизм получения выборок из генеральной совокупности случайных чисел и получение знаний о построении выборок. На основании предыдущей работы берем округленные до целых чисел предыдущего среднего значения и стандартное отклонение и на основании этих данных генерируем генеральную совокупность из 1000 элементов (рисунок 4) [2].

	A	B	C	D	E	F	G	H	I	J	K	L
1	Перед вами генеральная совокупность случайных чисел, которая имеет свои статистические характеристики .											
2	Необходими найти их и, используя инструмент Выборка из надстройки "Анализ данных" получить несколько выборок,											
3	оценить их точность и связать её с их объемом. Провести в этих выборках анализ грубых ошибок											
4												
5	Генеральная совокупность:											
6	713,61											
7	706,993											
8	725,898											
9	714,87											
10	715,606											
11	724,953											
12	725,898											
13	728,839											
14	723,693											
15	731,57											

Рисунок 4 – Сгенерированные числа для второго метода

Третья задача – это знакомство с дисперсионный анализ данных, где студент должен из пяти наборов данных выделить те, которые имеют связи между столбцами уровней. Задача генератора заключается в построении пяти наборов данных. В каждом из наборов имеется пять столбцов, отвечающих пяти уровням исследуемого параметра, каждый столбец включает по тринадцать элементов. На первом этапе случайным образом выбираются порядок следования наборов по следующим условиям: два набора со случайным набором чисел без каких-либо влияний исследуемого параметра через его уровни, один набор с линейной зависимостью среднего значения от уровня, один набор с линейной зависимостью стандартного отклонения от уровня и один набор с линейными зависимостями среднего и стандартного отклонения от уровня [3]. Линейная тенденция определяется случайным числом в интервале от 0,2 до 0,9 с положительным или отрицательным знаками. После генерации выборок (рисунок 4), они переносятся на лист Excel и формируются данные для преподавателя (какой набор сгенерирован по указанной выше схеме).

	A	B	C	D	E	F	G	H	I	J	K
1	Предлагается несколько наборов случайных данных, выполненных для разных уровней других параметров $Z_1 - Z_n$,										
2	который так же влияет на выполняемые замеры. Необходимо выяснить какие из внешних параметров влияют на										
3	исследуемую величину.										
4											
5	5 уровней для Z_1										
6	708,254	720,384	723,693	710,459	728,104						
7	723,693	732,515	722,117	724,796	749,057						
8	715,973	712,665	715,501	729,207	708,529						
9	730,625	727,789	727,789	734,406	689,506						
10	731,045	744,094	728,734	724,953	722,222						
11	727,001	725,898	701,086	709,724	717,076						
12	728,839	711,72	710,774	717,391	728,104						
13	722,222	768,907	698,329	746,694	749,057						
14	719,282	715,146	734,721	743,543	731,045						
15	717,076	729,679	738,397	728,104	705,221						

Рисунок 5 – Сгенерированные числа для третьего метода

Всероссийское СМИ

«Академия педагогических идей «НОВАЦИЯ»

Свидетельство о регистрации Эл №ФС 77-62011 от 05.06.2015 г.

(выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций)

Сайт: akademnova.ru

e-mail: akademnova@mail.ru

Четвертая задача – это исследование набора входных параметров на выходную функцию средствами корреляционного анализа. В данном методе генератор готовит десять входных параметров X и один выходной Y , предполагается, что это собранные данные работы какой-то установки на протяжении достаточно большого времени. Требуется определить какие из входных параметров наиболее сильно влияют на выходную функцию и есть ли среди всех параметров сильно коррелированные между собой. В задании сначала генерируются все возможные связи: два столбца имеют среднюю и сильную положительные корреляции на функцию, два других столбца имеют среднюю и сильную отрицательные корреляции на функцию и имеется две пары входных параметров имеющие взаимные корреляции (могут быть как положительные, так и отрицательные) [4]. Программа генерирует необходимый набор данных в виде матрицы 25 строк и 11 столбцов. Для генерации корреляционных зависимостей используется формула генерации случайных чисел от независимой переменной. В базу ответов помещаются номера столбцов X , которые влияют на Y и два пары номеров столбцов, коррелированных X (рисунок 6).

	A	B	C	D	E	F	G	H	I	J	K	L
1	После выбора всех возможных входных параметров $X_1 - X_n$ на выходную величину ставится вопрос, какие из них и как											
2	сильно влияют на выходную величину. Имеется набор из X_n - входных факторов и Y - выходной функции. Необходимо											
3	распределить все фактор X по их уровню влияния на Y и показать какие из факторов имеют взаимное влияние.											
4												
5	Данные для корреляционного анализа											
6	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	Y	
7	11,39745	212,8849	93,30922	23,39992	46,98523	65,58794	-40,7402	79,58637	45,62039	99,63825	116,3691	
8	11,49662	211,5334	94,37937	16,09773	45,93133	66,54835	-41,5142	79,64813	46,79317	97,24633	115,0963	
9	13,68259	223,7397	96,95398	18,75317	47,49535	62,43612	-40,7808	74,74102	48,16915	98,52023	117,7882	
10	12,45723	210,2935	95,62134	19,73887	44,97843	68,39483	-40,8627	76,2126	46,81249	97,90685	118,5891	
11	13,93094	213,4737	95,91834	19,05485	43,82182	68,37046	-38,075	74,87186	48,39948	99,06902	118,9347	
12	11,99331	222,9524	96,82482	18,43158	42,76792	63,27344	-40,7488	78,49116	44,99793	99,56604	115,4358	
13	10,1057	234,6193	96,93386	19,35654	43,07434	65,21864	-41,3351	80,95139	46,11393	96,79864	112,4018	
14	12,95393	208,3284	95,89025	17,86802	41,81502	62,0912	-36,5344	77,85443	47,15247	97,33826	117,0991	
15	16,46443	209,6953	96,51478	19,65822	39,40081	62,06683	-36,9586	78,38745	48,49602	95,63214	118,0006	
16	10,45322	213,7899	95,3701	19,03494	31,23723	60,29596	-45,5981	79,7966	46,81128	98,78333	114,5017	

Рисунок 6 – Сгенерированные числа для четвертого метода

Пятой задачей является регрессионный анализ, студент должен в одном из пяти наборов данных найти столбца которые имеют достаточно сильную регрессионную зависимость и определить тип этой зависимости. Сначала случайными образом выбираем одну из регрессионных функций: линейная, степенная, логарифмическая и экспоненциальная. Затем также случайным образом выбираем пару столбцов, между которыми имеется данная связь [5]. Генерируется четыре набора случайных чисел и пятый набор из выбранной связи генерируем с использованием формулы генерации случайных чисел от независимой переменной. Выводим данные в Excel и заносим номера столбцов в виде независимая – зависимая переменные (рисунок 7).

	A	B	C	D	E	F	G	H	I	J	K	L
1	После выбора всех возможных входных параметров $X_1 - X_n$, на выходную величину ставится вопрос, какие из них и как											
2	сильно влияют на выходную величину. Имеется набор из X_n - входных факторов и Y - выходной функции. Необходимо											
3	распределить все фактор X по их уровню влияния на Y и показать какие из факторов имеют взаимное влияние.											
4												
5	Данные для регрессионного анализа											
6	65,6819	839,645	73,2003	854,901	842,978							
7	81,8823	841,824	72,8294	817,882	818,748							
8	52,9844	861,295	74,3576	830,286	826,969							
9	70,4748	837,209	73,3419	841,103	860,286							
10	59,3137	845,286	74,0751	824,517	846,151							
11	70,4762	839,645	74,6126	846,151	833,026							
12	60,0503	852,209	73,7831	837,209	845,478							
13	61,8009	820,799	72,1937	831,007	861,632							
14	31,3143	839,517	77,7218	877,786	845,927							
15	75,7231	844,132	71,8312	848,619	834,901							

Рисунок 7 – Сгенерированные числа для пятого метода

Шестой задачей является исследование динамических рядов. Студент должен оценить фазу колебания ряда и для определенных зон данной фазы колебания (минимум, максимум, среднее значение или другая зона) подобрать регрессионную зависимость и оценить ее тренды вперед. Для генерации динамических рядов случайным образом выбирается одно из указанных выше регрессионных зависимостей, потом снова случайным образом генерируются параметры синусоидальной зависимости. Далее на основании этих данных формируется возрастающий ряд для независимой переменной и набор

случайных чисел для зависимой переменной, среднее значение которой формируется от текущего значения независимой переменной, влияние регрессионной и синусоидальной зависимостей [5]. Генерируется набор из двух переменных объемом от 50 до 200 элементов в каждом. В базу данных для преподавателя вносится тип регрессионной зависимости и параметры синусоиды. Необходимо определить общую тенденцию по сгенерированному ряду, то есть колебательная составляющая синус и регрессионная составляющая, необходимо определить регрессионную составляющую, которая существует, выбирается определенные фазы синусоиды, точки средних значений, группа точек максимальных значений, минимальных и как раз для этих групп точек строятся график. Точки выделяются определенным образом, накладываются определенные ограничения на X – обычно это время, например, если наложить ограничение на часовой интервал с восьми утра до десяти утра можно определить степень заинтересованности пользователя в определенном приложении и по найденным значениям возможно построить регрессионные ряды (рисунок 8).

	A	B	C	D	E	F	G	H	I	J	K	L
1	После выбора всех возможных входных параметров $X_1 - X_n$ на выходную величину ставится вопрос, какие из них и как											
2	сильно влияют на выходную величину. Имеется набор из X_n - входных факторов и Y - выходной функции. Необходимо											
3	распределить все фактор X по их уровню влияния на Y и показать какие из факторов имеют взаимное влияние.											
4												
5	Динамический ряд:											
6	22,62											
7	24,24											
8	25,86											
9	27,48											
10	29,1											
11	30,72											
12	32,34											
13	33,96											
14	35,58											
15	37,2											

Рисунок 8 – Сгенерированные числа для пятого метода

Всероссийское СМИ

«Академия педагогических идей «НОВАЦИЯ»

Свидетельство о регистрации Эл №ФС 77-62011 от 05.06.2015 г.

(выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций)

Сайт: akademnova.ru

e-mail: akademnova@mail.ru

После того как, студент сгенерировал данные, в личном кабинете преподавателя появляются ответы (рисунок 9).

ФИО	Сп. числа	Выборки	ДА	КА	РА	ДР
Ершова Диана Андреевна	Хсп: 64.2234, Спр: ...	Х: 64, Sr: 6	1 5 3 2 4	4 5 1 10 3 ...	1-3f=3	1-2f=4
*						

Рисунок 9 – Ответы в форме преподавателя.

Рассмотренные выше методы являются основами в статическом анализе. В дальнейшем генератор предлагается расширить для создания заданий в рамках дисциплины «Анализ больших данных». С помощью данной программы и разработанных алгоритмов студенты могут изучать методы статического анализа. А преподавателю не нужно вычислять ответы вручную для каждого студента, что экономит время и силы.

Список использованной литературы:

1. Джон Форман. Много цифр: Анализ больших данных при помощи Excel / Джон Форман; Пер. с англ. А. Соколовой. – М.: Альпин Паблишер, 2016. – 461 с.
2. Гоц А.Н., Горнушкин Ю.Г. Погрешности измерений при экспериментальных исследованиях двигателей внутреннего сгорания: Учеб. пособие / Владим. гос.ун-т. – Владимир, 2003. – 64 с.
3. Арьков В.Ю., Жевак А.В., Абдулнагимов А.И.. Статистический анализ средствами пакета MS Excel:Лабораторный практикум по дисциплине «Статистика» / Уфимск. гос. авиац. техн. ун-т; Сост.: В.Ю. Арьков, А.В. Жевак, А.И. Абдулнагимов. – Уфа, 2008. – 51 с.

Всероссийское СМИ

«Академия педагогических идей «НОВАЦИЯ»

Свидетельство о регистрации Эл №ФС 77-62011 от 05.06.2015 г.

(выдано Федеральной службой по надзору в сфере связи, информационных технологий и массовых коммуникаций)

Сайт: akademnova.ru

e-mail: akademnova@mail.ru

4. Спири́н Н.А., Лавров В.В., Зайнуллин Л.А., Бондин А.Р., Бурыкин А.А. Методы планирования и обработки результатов инженерного эксперимента: Учебное пособие / Н.А. Спири́н, В.В. Лавров, Л.А. Зайнуллин, А.Р. Бондин, А.А. Бурыкин; Под общ. ред. Н.А. Спирина. — Екатеринбург: ООО «УИНЦ», 2015 — 290 с.
5. Вадзинский Р. Статистические вычисления в среде Excel. Библиотека пользователя. — СПб.: Питер, 2008. — 608 с.

Дата поступления в редакцию: 01.07.2020 г.

Опубликовано: 01.07.2020 г.

© Академия педагогических идей «Новация». Серия «Студенческий научный вестник», электронный журнал, 2020

© Ершова Д.А., 2020