

ЭКОНОМИКА И УПРАВЛЕНИЕ НАРОДНЫМ ХОЗЯЙСТВОМ

УДК 004.4

DOI: 10.18384/2310-6646-2021-2-47-55

НЕПАРАМЕТРИЧЕСКИЙ КРИТЕРИЙ РАЗЛИЧИЯ ДЛЯ ДВУХ СВЯЗАННЫХ ВЫБОРОК В ТАБЛИЧНОМ РЕДАКТОРЕ MS EXCEL

Антипина Н. М.¹, Захаров В. Н.¹, Протасов Ю. М.¹, Юров В. М.²

¹Московский государственный областной университет

141014, Московская обл., г. Мытищи, ул. Веры Волошиной, д. 24, Российская Федерация

²Технологический университет

141070, Московская обл.ась, г. Королев, ул. Гагарина, д. 42, Российская Федерация

Аннотация

Цель. Реализовать процедуру различия 2-х связанных выборок с использованием непараметрического критерия в среде текстового редактора MS Excel.

Процедура и методы. Процедура различия двух связанных выборок осуществлена с помощью непараметрического критерия знаков в среде MS Excel.

Результаты. Разработаны алгоритм и программа на языке программирования VBA для выполнения автоматизированных расчётов. Проведена проверка правильности работы программы решением вручную двух практических примеров с использованием электронных таблиц.

Теоретическая и/или практическая значимость. Разработанные алгоритм и программа могут найти практическое применение в системах автоматизированной обработки данных с использованием методов математической статистики там, где отсутствуют программы, выполняющие расчёты по подобным алгоритмам. Наличие исходного кода позволит модернизировать алгоритм и программу в случае необходимости и внедрить программу в пакеты используемых программ, в т. ч. написанные на других языках.

Ключевые слова: связанные выборки, непараметрические критерии, критерий знаков, нулевая гипотеза, биномиальное распределение

NON-PARAMETRIC CRITERION OF DIFFERENCE FOR TWO RELATED SAMPLES IN TABLE EDITOR MS EXCEL

N. Antipina¹, V. Zakharov¹, Yu. Protasov¹, V. Yurov²

¹Moscow Region State University

24, Very Voloshinoy ul., Mytishchi, 141014, Moscow Region, Russian Federation

²Technological University

42, Gagarina ul., Korolev, 141070, Moscow Region, Russian Federation

© СС ВУ Антипина Н. М., Захаров В. Н., Протасов Ю. М., Юров В. М., 2021.

Abstract

Aim. To implement the procedure for distinguishing two related samples using a nonparametric criterion in the MS Excel text editor.

Methodology. The procedure for distinguishing two related samples was carried out using a nonparametric criterion of signs in MS Excel.

Results. An algorithm and a program in the VBA programming language have been developed for performing automated calculations. The correctness of the program was checked by manually solving two practical examples using spreadsheets.

Research implications. The developed algorithm and program can find practical application in automated data processing systems using the methods of mathematical statistics where there are no programs that perform calculations using such algorithms. The presence of the algorithm and source code will make it possible to modernize the algorithm and the program, if necessary, and to implement the program in the packages of the programs used, including in the software packages written in other languages.

Keywords: related samples, nonparametric tests, sign test, null hypothesis, binomial distribution

Введение

Одним из важнейших аналитических инструментов в области поддержки процесса принятия решений являются статистические методы. Они широко используются в политике, бизнесе, научных исследованиях. На практике часто встречаются задачи сравнения результатов обследования некоторого признака до и после определённого воздействия или экспериментальной и контрольной групп. Эти задачи решаются с помощью так называемых критериев различия, которые бывают параметрическими и непараметрическими.

Параметрические критерии различия базируются на известном распределении генеральной совокупности. Чаще всего в роли известного распределения выступает нормальное распределение.

В случае нормального распределения параметрические критерии имеют большую мощность по сравнению с непараметрическими. Однако при существенном отклонении проверяемого распределения от нормального закона распределения предпочтительным является использование непараметрических критериев, т. к. они могут с большей достоверностью отклонять нулевую гипотезу.

Непараметрические критерии не зависят от вида распределения генеральной совокупности, поэтому их называют ещё свободными от распределения критериями.

К сожалению, богатые возможности непараметрических методов не реализованы в широко используемом для решения экономико-статистических, научно-технических и других задач табличном редакторе MS Excel. Этот пробел был частично устранён в работах Н. М. Антипиной, В. Н. Захарова и Ю. М. Протасова. Так, в работе «Непараметрический критерий различия для двух независимых выборок в табличном редакторе MS EXCEL» [1] задача непараметрического различия решена применительно к 2-м независимым выборкам. В работе «Однофакторный дисперсионный анализ Краскела-Уоллиса в табличном редакторе MS EXCEL» [2] реализован однофакторный дисперсионный анализ Краскела-Уоллиса, позволяющий проводить различие 2-х и более независимых выборок.

Целью данной работы является реализация процедуры различия 2-х связанных выборок с использованием непараметрического критерия знаков в среде текстового редактора MS Excel. Для выполнения автоматизированных расчётов составлена программа на языке программирования VBA [4; 5; 6; 7; 8].

Рассматриваемые в статье связанные выборки получают, когда на каждый элемент совокупности действует исследуемый фактор. Такими факторами могут быть, например, рекламная кампания, обучение, приём лекарства и т. д. При этом производят первое наблюдение до воздействия фактора, а второе – после воздействия.

Процедура различия для двух связанных выборок с помощью критерия знаков

Будем обозначать $a_j, b_j, j = \overline{1, k}$ выборочные данные изучаемых признаков A и B . Нулевую гипотезу H_0 сформулируем как гипотезу о принадлежности признаков A, B к однородной совокупности. В предположении верности гипотезы H_0 значения вариантов a_j и b_j считают взаимозаменяемыми. Поэтому случайные события – появление отрицательных neg и положительных pos разностей $a_j - b_j$ – можно полагать равновероятными.

Т. к. измеряемые признаки предполагаются непрерывными, случайное событие – появление нулевых разностей $a_j - b_j$ – считается невозможным. На практике появление таких разностей возможно вследствие, например, ошибок округления. Поэтому пары наблюдений, соответствующие нулевым разностям $a_j - b_j$, убираются из расчётов.

Появление как положительных, так и отрицательных знаков разностей $a_j - b_j$ является независимым событием. На этом основании считается: число положительных разностей pos и число отрицательных разностей neg имеют биномиальное распределение с параметрами $p = \frac{1}{2}$ и m , где m – число ненулевых разностей $a_j - b_j$.

Таким образом, процедура различия с помощью критерия знаков сводится к проверке нулевой гипотезы:

$$H_0: p = 0,5$$

при альтернативной гипотезе:

$$H_1: p \neq 0,5$$

В критерии знаков статистикой считается так называемое значение нетипичного сдвига:

$$G_{\text{нетип}} = \min(pos, neg) \quad (1)$$

Условие принятия H_0 на уровне значимости ξ можно сформулировать как:

$$\sum_{j=0}^{G_{\text{нетип}}} C_m^j \left(\frac{1}{2}\right)^m > \frac{\xi}{2} \quad (2)$$

В неравенстве (2) левая часть описывает интегральную функцию биномиального распределения нетипичного сдвига $G_{\text{нетип}}$ с параметрами m и $p = \frac{1}{2}$. Здесь C_m^j – число сочетаний из m по j .

Если неравенство (2) не выполняется, нулевая гипотеза отвергается и принимается альтернативная гипотеза H_1 .

Условие принятия H_0 на уровне значимости ξ можно представить иначе:

$$G_{\text{нетип}} > G_{\text{кр}} \quad (3)$$

где $G_{\text{кр}}$ – критическое значение для нетипичного сдвига, определяемое по величине типичного сдвига [3]:

$$G_{\text{тип}} = \max(pos, neg) \quad (4)$$

Примеры использования критерия знаков

Пример 1. Влияние работы на ультразвуковой установке на содержание сахара в крови по выборке из 15 работающих.

Сахар, мг %		Сдвиг
натощак до работы	после 3 ч работы	
98	51	-47
85	63	-22
103	95	-8
69	52	-17
79	79	0
83	77	-6
68	70	+2
74	69	-5
68	68	0
78	49	-29
67	62	-5
75	75	0
85	51	-34
63	63	0
91	73	-18

Значимо ли на уровне $\xi=0,05$ влияние работы на ультразвуковой установке на содержание сахара в крови работающих?

Решение

1. В столбце «Сдвиг» для каждого работающего определяется изменение сахара в крови. Подсчитываются суммы положительных $pos=1$ и отрицательных $neg=10$ сдвигов. Нулевые сдвиги не учитываются.
2. Находим значение нетипичного сдвига:

$$G_{\text{нетип}} = \min(1,10) = 1$$
3. Находим значение типичного сдвига:

$$G_{\text{тип}} = \max(10,1) = 10$$
4. Находим критическое значение нетипичного сдвига для $\xi = 0,05$ и $m = G_{\text{тип}} = 10$ [3]:

$$G_{\text{кр}} = 1$$
5. Так как $G_{\text{нетип}}$ не превышает $G_{\text{кр}}$, принимается альтернативная гипотеза о статистической значимости изменений сахара в крови.

Достоинство полученного решения – простота, недостаток – неудобство, связанное с необходимостью иметь статистические таблицы. Поэтому решим пример без использования таблиц, вычислив непосредственно левую часть неравенства (2). Для этого воспользуемся статистической функцией MS EXCEL БИНОМ.РАСП (рис. 1), аргументы которой:

число успехов – $G_{\text{нетип}} = \min(1,10) = 1$;

число испытаний – $pos + neg = 11$;

вероятность успеха – $p = 0,5$.

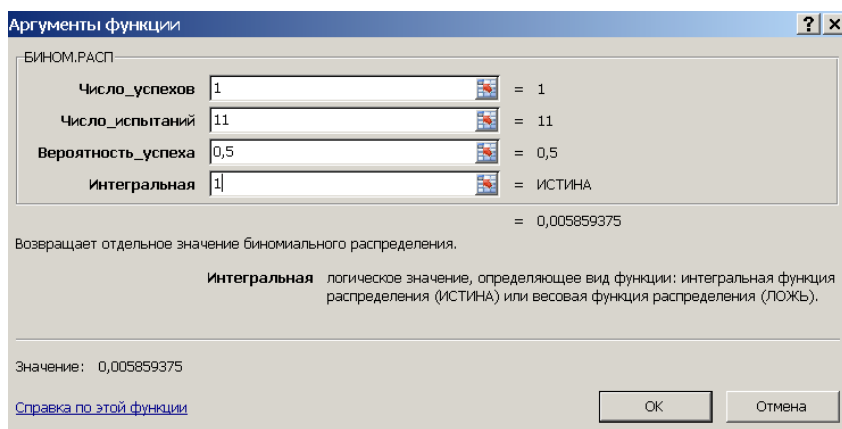


Рис. 1 / Fig. 1. Окно ввода исходных данных / Input Data Window

Т. к. $0,005859 < \frac{0,05}{2} = 0,025$, т. е. неравенство (2) не выполняется, нулевая гипотеза отвергается.

Для автоматизации рассмотренных выше расчётов (и уменьшения времени, затрачиваемого на их проведение) были разработаны алгоритм и соответствующая ему программа на языке VBA [4; 5; 6; 7; 8].

Алгоритм

1. Ввести значения признаков A и B в столбцы A и B электронной таблицы, начиная с $A1$ и $B1$ – вниз по столбцам – до старта программы.
2. Ввести значение k (количество строк в столбцах данных) в ячейку $D1$ электронной таблицы до старта программы.
3. Стартовать программу и после старта ввести в диалоговом окне значение ksi (уровень значимости). Далее все действия выполняются в автоматическом режиме.
4. Задать начальные значения $neg:=0$; $pos:=0$.
5. Начало цикла по j для $j:=1, 2, \dots, k$.
6. Вычислить разность и записать в ячейку:
 $Cells(j, 3) = Cells(j, 2) - Cells(j, 1)$,
 где $Cells$ – ячейки электронной таблицы и в скобках справа указаны адреса (номер строки, номер столбца).
7. Если $Cells(j, 3) < 0$, выполнить рекуррентное суммирование:
 $neg:=neg + 1$.
8. Если $Cells(j, 3) > 0$, выполнить рекуррентное суммирование:
 $pos:=pos + 1$.
9. Конец цикла по j .
10. Записать в ячейки $F2$ и $G2$ вычисленные значения pos и neg .
11. Вычислить $min(pos, neg)$ и записать в ячейку $F5$.
12. Вычислить сумму $pos + neg$ и записать в ячейку $F8$.
13. Вычислить $max(pos, neg)$ и записать в ячейку $G5$.
14. Вычислить $ksi/2$ и записать в ячейку $I8$.
15. Записать в ячейку $E1$ текст: «Условие (2) о том, что значение функции БИНОМ.РАСП больше, чем $ksi/2$ ».
16. Если $G8 > I8$, в ячейку $E12$ записать: «выполнено», иначе в ячейку $E12$ записать: «не выполнено».

Разработанный алгоритм и составленная по этому алгоритму компьютерная программа на языке VBA [4; 5; 6; 7; 8] были использованы для решения практических примеров.

Решение Примера 1. Данные для расчётов были предварительно занесены в таблицу MS Excel в диапазон ячеек A1:B15 (значения признаков A и B) и ячейку D1 (n – количество строк в столбцах A и B). Следует щёлкнуть кнопкой CommandButton и запустить программу на исполнение.

Исходные данные, результаты расчётов и кнопка запуска показаны на рис. 2.

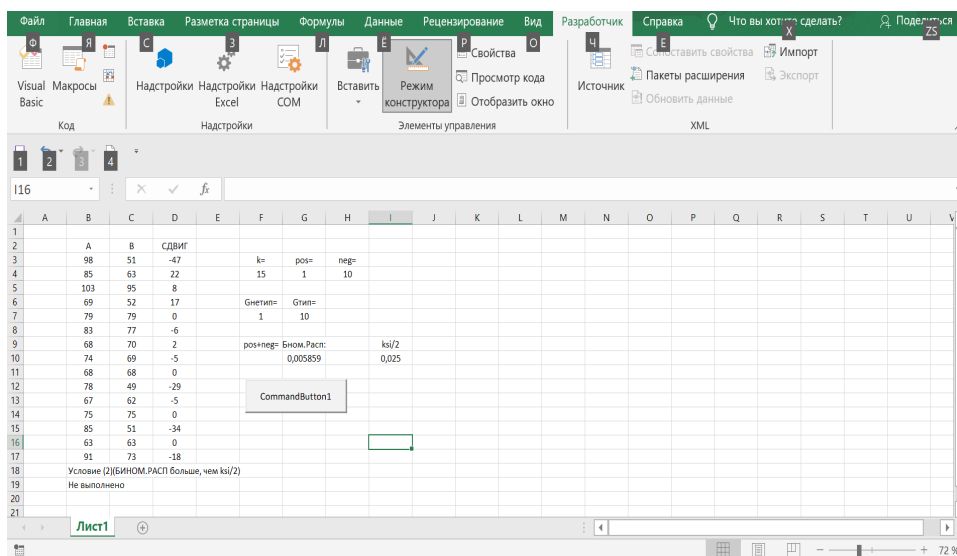


Рис. 2 / Fig. 2. Нулевая гипотеза отклоняется / The null hypothesis is rejected

Столбец исходных данных A содержит значения изучаемого признака до воздействия фактора, а столбец B – значение изучаемого признака после воздействия.

Программа считает СДВИГ и количество положительных и отрицательных разностей. В данном примере эти значения равны 1 и 10. Под “ $pos + neg$ ”, “БИНОМ.РАСПИ=”, “ $ksi/2$ ” показаны 3 вычисленные величины:

- число испытаний (т. е. $pos + neg$),
- значение статистической функции БИНОМ.РАСПИ;
- $ksi/2$ (где ksi – уровень значимости ξ , и в данном примере: $\xi = 0,05$).

С учётом полученных результатов программа автоматически выводит сообщение: «Условие не выполнено». Это означает, что нулевая гипотеза H_0 о несущественности влияния фактора отклоняется.

Таким образом, на уровне значимости $\xi = 0,05$ можно считать статистически значимыми изменения сахара в крови натощак у 15 сотрудников, работающих на ультразвуковых установках, до работы и через 3 ч после начала работы.

Пример 2. Отдел маркетинга торговой фирмы провёл рекламную кампанию (РК) в связи с поступлением в продажу нового товара. В исследовании участвовало 14 магазинов фирмы. Необходимо выяснить на уровне значимости 0,05 эффективность РК для увеличения объёма продаж.

Объём продаж (тыс. у.е.) «до»	Объём продаж (тыс. у.е.) «после»
30	34
39	39
35	26
34	33
40	34
35	40
22	25
22	23
32	33
23	24
16	15
34	27
33	35
34	37

Решение

В данном примере результаты расчёта оказались другими (рис. 3).

	A	B	C	D	E	F	G	H	I
1									
2		A	B	СДВИГ					
3		30	34	4		k=	pos=	neg=	
4		39	39	0		14	8	5	
5		35	26	-9					
6		34	33	-1		Gнетип=	Gтип=		
7		40	34	-6		5	8		
8		35	40	5					
9		22	25	3		pos+neg=Зном.Расп:		ksi/2	
10		22	23	1		13	0,290527	0,025	
11		32	33	1					
12		23	34	11					
13		16	15	-1					
14		34	27	-7					
15		33	35	2					
16		34	37	3					
17		Условие (2) (БИНОМ.РАСП больше, чем ksi/2)							
18		Выполнено							

Рис. 3 / Fig. 3. Нулевая гипотеза принимается / The null hypothesis is accepted

Количество чисел в столбцах A и B задаётся значением k , где $k=14$. Остальные значения, посчитанные по программе, сведены в следующую таблицу:

pos	neg	Gнетип	Gтип	pos + neg	БИНОМ.РАСП	ksi/2
8	5	5	8	13	0,290527	0,025

Из двух последних столбцов видно, что условие (2) выполняется. Это означает, что нулевая гипотеза H_0 однородности признаков A и B принимается.

Таким образом, по результатам анализа объёмов продаж в 14 магазинах торговой фирмы можно сделать заключение на уровне значимости $\xi = 0,05$ о неэффективности проведённой рекламной кампании.

Заключение

Рассмотрены 2 практических примера. Разработанные алгоритм и программа могут найти применение в системах автоматизированной обработки данных с использованием методов математической статистики там, где отсутствуют программы, выполняющие расчёты по подобным алгоритмам.

Статья поступила в редакцию 12.02.2021.

ЛИТЕРАТУРА

1. Антипина Н. М., Захаров В. Н., Протасов Ю. М. Непараметрический критерий различия для двух независимых выборок в табличном редакторе MS EXCEL // Информационно-технологический вестник. 2019. № 1. С. 48–58.
2. Антипина Н. М., Захаров В. Н., Протасов Ю. М. Однофакторный дисперсионный анализ Краскела-Уоллиса в табличном редакторе MS EXCEL // Информационно-технологический вестник. 2019. № 4. С. 46–54.
3. Закс Л. Статистическое оценивание. М: Статистика, 1976. 320 с.
4. Захаров В. Н. Инструменты моделирования бизнес-процессов // Вестник Московского государственного областного университета. Серия: Экономика. 2016. № 3. С. 19–28.
5. Слепцова Л. Д. Программирование на VBA в Excel 2010. М.: Диалектика, 2010. 432 с.
6. Комолова Н., Клименко А. Программирование на VBA в Excel 2019. СПб.: BHV, 2020. 495 с.
7. Александр М., Куслейка Р. Excel 2016. Профессиональное программирование на VBA. М.: Диалектика, 2018. 784 с.
8. Швыдкой С. А. Учебник по функциям Excel и программированию в среде VBA. М., 2019. 104 с.

REFERENCES

1. Antipina N. M., Zakharov V. N., Protasov Yu. M. [Nonparametric test of difference for two independent samples in the MS EXCEL spreadsheet editor]. In: *Informatsionno-tekhnologicheskii vestnik* [Information Technology Bulletin], 2019, no. 1, pp. 48–58.
2. Antipina N. M., Zakharov V. N., Protasov Yu. M. [Kruskal-Wallis one-way ANOVA in MS EXCEL spreadsheet editor]. In: *Informatsionno-tekhnologicheskii vestnik* [Information Technology Bulletin], 2019, no. 4, pp. 46–54.
3. Zaks L. *Statisticheskoe otsenivanie* [Statistical estimation]. Moscow, Statistika Publ., 1976. 320 p.
4. Zakharov V. N. [Business process modeling tools]. In: *Vestnik Moskovskogo gosudarstvennogo oblastnogo universiteta. Seriya: Ekonomika* [Bulletin of Moscow Region State University. Series: Economics], 2016, no. 3, pp. 19–28.
5. Sleptsova L. D. *Programmirovaniye na VBA v Excel 2010* [VBA programming in Excel 2010]. Moscow, Dialektika Publ., 2010. 432 p.
6. Komolova N., Klimentko A. *Programmirovaniye na VBA v Excel 2019* [VBA Programming in Excel 2019]. St. Petersburg, BHV Publ., 2020. 495 p.
7. Aleksander M., Kusleika R. Excel 2016. *Professionalnoye programmirovaniye na VBA* [Excel 2016. Professional VBA Programming]. Moscow, Dialektika Publ., 2018. 784 p.
8. Shvydkoy S. A. *Uchebnik po funktsiyam Excel i programmirovaniyu v srede VBA* [Excel Functions and VBA Programming Tutorial]. Moscow, 2019. 104 p.

ИНФОРМАЦИЯ ОБ АВТОРАХ

Антипина Наталья Михайловна – кандидат педагогических наук, профессор, заведующая кафедрой прикладной математики и информатики Московского государственного областного университета;

e-mail: nantipina@yandex.ru

Захаров Владимир Николаевич – кандидат технических наук, доцент, доцент кафедры прикладной математики и информатики Московского государственного областного университета;

e-mail: 232zakharov@rambler.ru

Протасов Юрий Михайлович – кандидат технических наук, доцент, доцент кафедры прикладной математики и информатики Московского государственного областного университета;

e-mail: protasov_54@mail.ru

Юров Владимир Михайлович – кандидат технических наук, доцент, доцент кафедры управления качеством и стандартизации Технологического университета;

e-mail: urow5@mail.ru

INFORMATION ABOUT THE AUTHORS

Nataliya M. Antipina – Cand. Sci. (Education), Prof., Departmental Head, Department of Applied Mathematics and Computer Science, Moscow Region State University;

e-mail: nantipina@yandex.ru

Vladimir N. Zakharov – Cand. Sci. (Engineering), Assoc. Prof., Department of Applied Mathematics and Informatics, Moscow Region State University;

e-mail: 232zakharov@rambler.ru

Yury M. Protasov – Cand. Sci. (Engineering), Assoc. Prof., Department of Applied Mathematics and Informatics, Moscow Region State University;

e-mail: protasov_54@mail.ru

Vladimir M. Yurov – Cand. Sci. (Engineering), Assoc. Prof., Department of Quality Management and Standardization, Technological University;

e-mail: urow5@mail.ru

ПРАВИЛЬНАЯ ССЫЛКА НА СТАТЬЮ

Непараметрический критерий различия для двух связанных выборок в табличном редакторе MS Excel / Н. М. Антипина, В. Н. Захаров, Ю. М. Протасов, В. М. Юров // Вестник Московского государственного областного университета. Серия: Экономика. 2021. № 2. С. 47–55.

DOI: 10.18384/2310-6646-2021-2-47-55

FOR CITATION

Antipina N. M., Zakharov V. N., Protasov Yu. M., Yurov V. M. Non-parametric Criterion of Difference for Two Related Samples in Table Editor MS Excel. In: *Bulletin of Moscow Region State University. Series: Economics*, 2021, no. 2, pp. 47–55.

DOI: 10.18384/2310-6646-2021-2-47-55