

УДК 519.23

## СРАВНЕНИЕ ДВУХ ПАРНЫХ ВЫБОРОК С ПОМОЩЬЮ ПАКЕТА СТАТИСТИЧЕСКИХ ПРОГРАММ STATA: НЕПАРАМЕТРИЧЕСКИЕ КРИТЕРИИ

© 2014 Г. О. А. Харькова, \*А. М. Гржибовский

Северный государственный медицинский университет, г. Архангельск  
\*Норвежский институт общественного здравоохранения, г. Осло, Норвегия

В предыдущих выпусках практикума рассматривались параметрические и непараметрические критерии, применяемые для сравнения двух, трех и более несвязанных групп, а именно: непарный критерий Стьюдента, однофакторный дисперсионный анализ для независимых выборок, критерий Манна — Уитни и критерий Крускала — Уоллиса, где общим условием была независимость выборок. В случае исследований типа «до — после», когда измерение признаков у одних и тех же участников исследования производится в разные моменты времени, при изучении близнецов, а также в исследованиях типа «случай — контроль» с использованием метода подобранных пар [3], то есть когда речь идет об отсутствии независимости выборок, необходимо применять критерии для парных выборок. В данной статье мы рассмотрим непараметрические критерии для сравнения двух парных групп — парный критерий Вилкоксона (или одновыборочный критерий Вилкоксона) и критерий знаков.

Парный критерий Вилкоксона (Wilcoxon signed rank test) и критерий знаков (Sign test) предназначены для сравнения парных выборок [9]. Для обоих критериев необходимо соблюдение следующих условий:

- количественный или качественный порядковый признак;
- в случае количественного признака не обязательно нормальное распределение разности между значениями изучаемого признака в группах;
- наличие не более двух зависимых групп.

Для проверки соблюдения условий никаких математических манипуляций не требуется. Учитывая второе условие, мы можем применить парный критерий Вилкоксона или критерий знаков, не обращая внимания на распределение изучаемого признака в группах. Однако надо помнить, что непараметрические критерии являются менее чувствительными, чем параметрические [1], поэтому если данные подчиняются закону нормального распределения, то корректнее использовать парный критерий Стьюдента. Помимо количественных переменных, не подчиняющихся закону нормального распределения, эти критерии можно применять и для порядковых данных. Например, пятибалльная оценка в школе, десятибалльная оценка состояния новорожденного по шкале Апгар и т. д., также могут сравниваться с помощью непараметрических критериев [2].

Парными считаются выборки, когда результаты измерения одного и того же показателя у одной и той же группы людей анализируются разными исследователями, например при определении уровня гемоглобина крови у десяти человек двумя разными лабораториями. Исследование типа «случай — контроль» с использованием метода подобранных пар также является показанием для применения парных методов анализа,

В статье рассматривается применение непараметрических критериев для сравнения данных в двух парных группах. Приводятся примеры расчетов критерия Вилкоксона для непарных выборок и критерия знаков как вручную, так и с помощью пакета статистических программ STATA. Разбираются условия применения непараметрических критериев для парных выборок, примеры представления результатов применения парного критерия Вилкоксона и критерия знаков в публикациях.

**Ключевые слова:** статистический анализ, непараметрические критерии, парный критерий Вилкоксона, критерий знаков, STATA

так как контроли для каждого случая подбираются исходя из определенных характеристик случаев [3]. В качестве таких характеристик обычно используют пол, возраст, образование, место жительства и др. Сюда же включаются и проспективные исследования, например, изучение признака до и после операции, в начале и в конце учебного года и т. д.

Основное различие между парным критерием Вилкоксона и критерием знаков заключается в том, что в случае последнего в анализе игнорируется значение изменений переменных и учитывается лишь их направление [4]. Поэтому критерий знаков обладает меньшей статистической мощностью по сравнению с парным критерием Вилкоксона [3].

Парный критерий Вилкоксона и критерий знаков основаны на ранжировании абсолютных величин разности без учета знака. Рассмотрим пример (табл. 1). У девочек младшего школьного возраста провели динамометрию (изучение мышечной силы) правой руки в начале учебного года и в конце. Можно ли утверждать, что в конце учебного года мышечная сила у девочек стала больше по сравнению с началом учебного года?

Таблица 1  
Пример расчета парного критерия Вилкоксона и критерия знаков

Показатель	Номер участника исследования										
	1	2	3	4	5	6	7	8	9	10	11
На начало учебного года	12	14	11	14	10	14	7	12	14	13	14
На конец учебного года	11	18	15	12	14	14	11	15	17	18	18
Разность	-1	4	4	-2	4	0	4	3	3	5	4
Без учета знака	1	4	4	2	4	0	4	3	3	5	4
Ранг	1	7	7	2	7	-	7	3,5	3,5	5	7
T <sup>+</sup>	-	7	7	-	7	-	7	3,5	3,5	10	7
T <sup>-</sup>	1	-	-	2	-	-	-	-	-	-	-
Сумма рангов	T <sup>+</sup> = 52					T <sup>-</sup> = 3					

*Алгоритм расчета парного критерия Вилкоксона:*

1. Создать новую переменную «разность», каждое из значений которой представляет собой  $X_{\text{конец}} - X_{\text{начало}}$ , где  $X_{\text{конец}}$  — значение переменной в конце учебного года, а  $X_{\text{начало}}$  — значение переменной в начале учебного года.

2. Записать абсолютные значения разности без учета знака.

3. Присвоить разностям ранги. Напоминаем о том, что переменные с нулевыми изменениями (участник исследования под номером 6 в нашем примере) в анализе не участвуют, поэтому рангов всего не 11, а 10. Если несколько значений равны, то им присваивается ранг, представляющий собой среднее значение из тех рангов, которые они получили бы, если бы не были равны (например, в отношении значения «3» —  $(3 + 4) / 2 = 3,5$ ; «4» —  $(5 + 6 + 7 + 8 + 9) / 5 = 7$ ).

4. Проставить ранги для положительных разностей (T<sup>+</sup>) и отрицательных (T<sup>-</sup>).

5. Подсчитать сумму рангов для положительных (ΣT<sup>+</sup>) и отрицательных (ΣT<sup>-</sup>) разностей.

6. Меньшая из двух сумм рангов (без учета знака) и будет являться эмпирическим значением парного критерия Вилкоксона (T).

7. Далее рассчитаем z-значение для парного критерия Вилкоксона и уровень статистической значимости [6]:

$$\bar{T} = \frac{n(n + 1)}{4};$$

$$SE_T = \sqrt{\frac{n(n + 1)(2n + 1)}{24}};$$

$$z = \frac{T - \bar{T}}{SE_T},$$

где T — значение парного критерия Вилкоксона;  $\bar{T}$  — среднее значение парного критерия Вилкоксона; SE<sub>T</sub> — стандартная ошибка парного критерия Вилкоксона; n — количество наблюдений, не имеющих нулевых изменений.

$$T = 3; \quad \bar{T} = \frac{10 * (10 + 1)}{4} = 27,5;$$

$$SE_T = \sqrt{\frac{10 * (10 + 1) * (2 * 10 + 1)}{24}} = 9,81;$$

$$z = \frac{3 - 27,5}{9,81} = -2,50$$

Если полученное z-значение больше, чем 1,960 (независимо от знака), то p < 0,05, если z-значение больше, чем 2,576, то p < 0,01 и если z-значение больше, чем 3,291, то p < 0,001. Согласно нашим данным, есть статистически значимые изменения по динамометрии правой руки у девочек младшего школьного возраста в конце учебного года по сравнению с началом (0,01 < p < 0,05).

*Алгоритм расчета критерия знаков:*

1. Создать новую переменную «разность» так же, как для парного критерия Вилкоксона (см. табл. 1). Если для расчета парного критерия Вилкоксона неважно, как осуществляется вычитание «до — после» или «после — до», то для расчета критерия знаков — исключительно «после — до» [4].

2. Подсчитать суммы нулевых, положительных и отрицательных сдвигов в строке «разница»:

$$\sum_{\text{нулевых сдвигов}} = 1$$

$$\sum_{\text{положительных сдвигов}} = 8$$

$$\sum_{\text{отрицательных сдвигов}} = 2$$

Далее в анализ учитываются только суммы положительных и отрицательных сдвигов, а сумма нулевых сдвигов игнорируется [4, 5].

3. Сумма сдвигов, получившаяся наибольшей, называется типичным сдвигом ( $n$ ), а сумма сдвигов, получившаяся наименьшей, соответственно нетипичным сдвигом ( $G_{эмп}$ ). Согласно нашим данным,  $n = 8$ , а  $G_{эмп} = 2$ . В случае, когда величины типичного и нетипичного сдвигов оказываются равными ( $n = G_{эмп}$ ), критерий знаков неприменим.

4. Оценка статистической значимости различий по критерию знаков производится по соответствующей таблице критических значений (табл. 2) [4]. В столбце, обозначенном  $n$ , приведены величины типичных сдвигов, а в столбцах, имеющих обозначения, соотносящиеся с уровнями значимости (0,05 и 0,01), представлены величины нетипичных сдвигов  $G$ . Если  $p < 0,05$ , то принимается альтернативная гипотеза (свидетельствующая о том, что есть различия) при соответствующем уровне значимости; в данном случае либо при  $p = 0,05$ , либо при  $p = 0,01$ .

Таблица 2

Критические значения критерия знаков ( $n$ ) при  $p = 0,05$  и  $p = 0,01$

n	p		n	p		n	p		n	p	
	0,05	0,01		0,05	0,01		0,05	0,01		0,05	0,01
5	0	—	27	8	7	49	18	15	92	37	34
6	0	0	28	8	7	50	18	16	94	38	35
7	0	0	29	9	7	52	19	17	96	39	36
8	1	0	30	10	8	54	20	18	98	40	37
9	1	0	31	10	8	56	21	18	100	41	37
10	1	0	32	10	8	58	22	19	110	45	42
11	2	1	33	11	9	60	23	20	120	50	46
12	2	1	34	11	9	62	24	21	130	55	51
13	3	1	35	12	10	64	24	22	140	59	55
14	3	2	36	12	10	66	25	23	150	64	60
15	3	2	37	13	10	68	26	23	160	69	64
16	4	2	38	13	11	70	27	24	170	73	69
17	4	3	39	13	11	72	28	25	180	78	73
18	5	3	40	14	12	74	29	26	190	83	78
19	5	4	41	14	12	76	30	27	200	87	83
20	5	4	42	15	13	78	31	28	220	97	92
21	6	4	43	15	13	80	32	29	240	106	101
22	6	5	44	16	13	82	33	30	260	116	110
23	7	5	45	16	14	84	33	30	280	125	120
24	7	5	46	16	14	86	34	31	300	135	129
25	7	6	47	17	15	88	35	32			
26	8	6	48	17	15	90	36	33			

При  $n = 8$   $G_{кр} = 1$  ( $p = 0,05$ ) или  $G_{кр} = 0$  ( $p = 0,01$ ). Следовательно, принимается нулевая гипотеза ( $p > 0,05$ ), так как  $G_{кр} < G_{эмп}$ . Это говорит о том, что нет статистически значимых изменений по динамометрии правой руки у девочек младшего школьного возраста в конце учебного года по сравнению с началом.

Рассмотрим использование парного критерия Вилкоксона и критерия знаков в программе STATA [7, 8]. Возьмем тот же самый пример. У девочек младшего школьного возраста провели динамометрию правой руки в начале учебного года (переменная  $dyn\_1$ ) и в конце (переменная  $dyn\_2$ ). Для того чтобы исключить возможность применения параметрического парного критерия Стьюдента, необходимо проверить

распределение разности между значениями изучаемого признака. С этой целью создается новая переменная  $difference$  путем открытия диалогового окна Create new variable через Data — Create or change data (рис. 1).

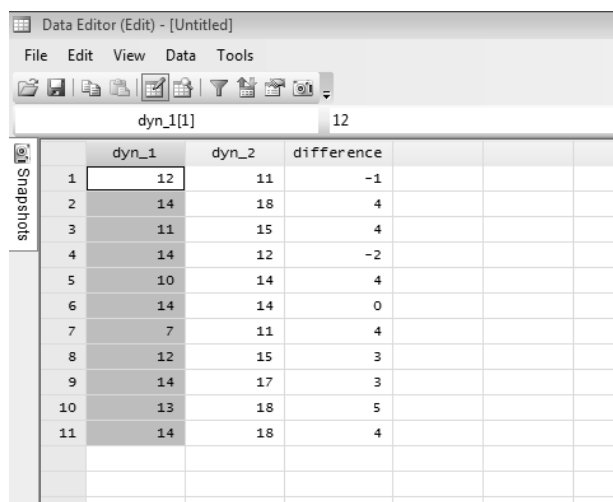
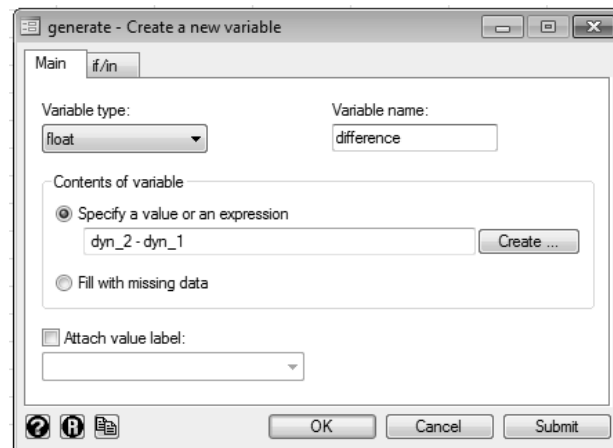


Рис. 1. Диалоговое окно для создания новой переменной с помощью математических вычислений (слева) и результат создания новой переменной  $difference$  (справа)

Во вкладке Main в окошке Variable name даем название новой переменной  $difference$ , в окошке Content of variable ставим « $\ominus$ » в Specify a value or an expression и пишем соответствующее математическое вычисление  $dyn\_2 - dyn\_1$ , затем Ok (см. рис. 1). В результате с нашими переменными  $dyn\_1$  и  $dyn\_2$  появляется новая переменная  $difference$ , каждое из значений которой получилось путем разницы значения динамометрии в конце учебного года и в начале.

Посмотрим распределение новой переменной  $difference$  с помощью критерия Shapiro-Wilk, который открывается при помощи меню Statistics — Summaries, tables and tests — Distributional plots and tests — Shapiro-Wilk normality test. В диалоговом окне Shapiro-Wilk normality test (рис. 2) во вкладке Main в окошке Variables выбираем переменную  $difference$  и нажимаем Ok. Уровень статистической значимости ( $Prob > z = 0.00210$ ) свидетельствует о

негауссовом распределении разности между значениями изучаемого признака, то есть новой переменной difference (рис. 3). В связи с этим корректнее применять непараметрические критерии — парный критерий Вилкоксона или критерий знаков.

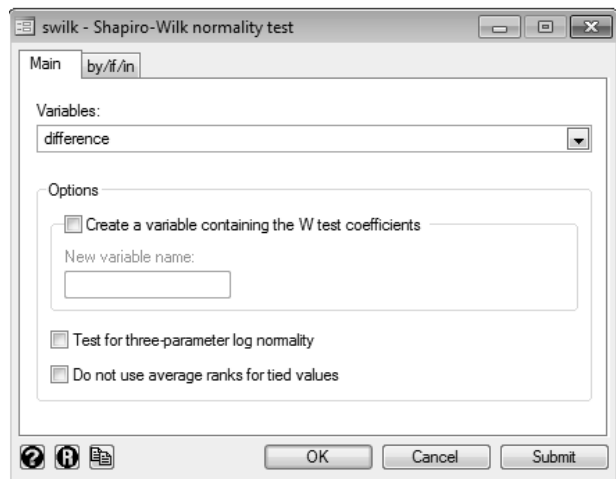


Рис. 2. Диалоговое окно критерия Shapiro-Wilk

```
. swilk difference
```

Shapiro-Wilk W test for normal data					
Variable	Obs	W	V	z	Prob>z
difference	11	0.75041	4.041	2.862	0.00210

Рис. 3. Результаты проверки распределения переменной difference с помощью критерия Shapiro-Wilk

Для использования парного критерия Вилкоксона в STATA следует зайти в диалоговое окно Wilcoxon matched-pairs signed-rank test, которое открывается при помощи меню Statistics — Summaries, tables, and tests — Nonparametric tests of hypothesis — Wilcoxon matched-pairs signed-rank test (рис. 4).

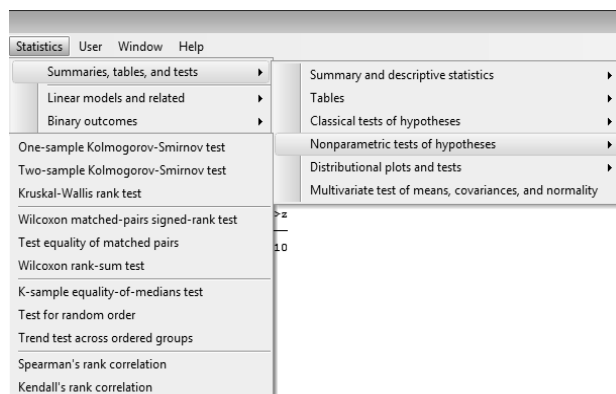


Рис. 4. Алгоритм поиска парного критерия Вилкоксона или критерия знаков в программе STATA

Во вкладке Main в окошке Variable выбирается изучаемая переменная dyn\_2; в окошке Expression пишется переменная dyn\_1, затем Ok (рис. 5). Результаты сравнения групп в динамике с помощью парного критерия Вилкоксона представлены на рис. 6.



Рис. 5. Диалоговое окно Wilcoxon matched-pairs signed-rank test

```
. signrank dyn_2 = dyn_1
```

Wilcoxon signed-rank test			
sign	obs	sum ranks	expected
positive	8	60	32.5
negative	2	5	32.5
zero	1	1	1
all	11	66	66

unadjusted variance	126.50
adjustment for ties	-2.63
adjustment for zeros	-0.25
adjusted variance	123.63

```
Ho: dyn_2 = dyn_1
      z = 2.473
      Prob > |z| = 0.0134
```

Рис. 6. Результат сравнения динамометрии правой руки у девочек младшего школьного возраста в начале учебного года и в конце (с помощью парного критерия Вилкоксона)

В таблице рис. 6 представлено количество (obs) положительных (positive), отрицательных (negative) и нулевых (zero) изменений, а также сумма рангов для каждого из них (sum ranks). В отличие от ручного способа расчета и расчета с помощью статистической программы SPSS, STATA для вычисления значения парного критерия Вилкоксона учитывает нулевые изменения, отсюда различия между расчетом, приведенным в начале статьи, и результатами применения программы STATA. Под таблицей представлено значение  $z = 2,473$  и уровень статистической значимости ( $\text{Prob} > |z| = 0.0134$ ), который позволяет нам отклонить нулевую гипотезу при критическом уровне значимости 5 %.

Для использования критерия знаков в STATA следует зайти в диалоговое окно Test equality of matched pairs, которое открывается при помощи меню Statistics — Summaries, tables, and tests — Nonparametric tests of hypothesis — Test equality of matched pairs (см. рис. 4). Диалоговое окно у критерия знаков (Sign test of matched pairs) имеет такой же вид и параметры, как и у парного критерия Вилкоксона (Wilcoxon matched-pairs signed-rank test) (рис. 7).

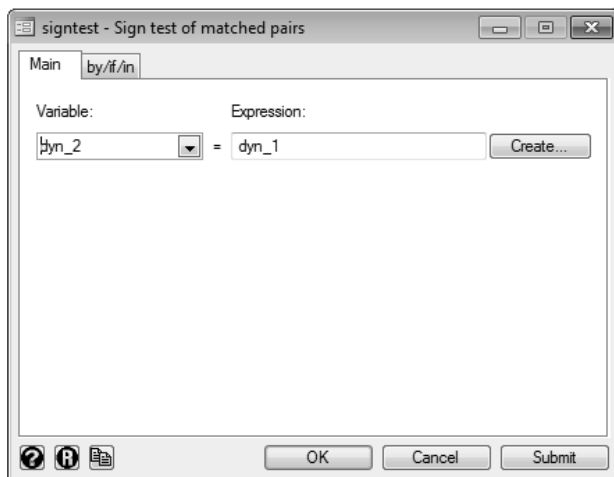


Рис. 7. Диалоговое окно Sign test of matched pairs

Результат критерия знаков в программе STATA представлен как односторонне направленный статистический критерий (One-sided tests), так и двусторонне направленный (Two-sided tests) (рис. 8). В литературе и в статьях чаще можно встретить представление результатов с двусторонне направленным статистическим критерием. Согласно нашим данным, нулевую гипотезу мы отклонить не можем, следовательно, мы принимаем нулевую гипотезу об отсутствии статистически значимых изменений по динамометрии правой руки у девочек младшего школьного возраста в конце учебного года по сравнению с началом ( $p = 0,109$ ).

```
. signtest dyn_2 = dyn_1

Sign test

      sign | observed expected
-----|-----
 positive |      8         5
 negative |      2         5
   zero   |      1         1
-----|-----
    all   |     11        11

One-sided tests:
Ho: median of dyn_2 - dyn_1 = 0 vs.
Ha: median of dyn_2 - dyn_1 > 0
Pr(#positive >= 8) =
  Binomial(n = 10, x >= 8, p = 0.5) = 0.0547

Ho: median of dyn_2 - dyn_1 = 0 vs.
Ha: median of dyn_2 - dyn_1 < 0
Pr(#negative >= 2) =
  Binomial(n = 10, x >= 2, p = 0.5) = 0.9893

Two-sided test:
Ho: median of dyn_2 - dyn_1 = 0 vs.
Ha: median of dyn_2 - dyn_1 != 0
Pr(#positive >= 8 or #negative >= 8) =
  min(1, 2*Binomial(n = 10, x >= 8, p = 0.5)) = 0.1094
```

Рис. 8. Результат сравнения динамометрии правой руки у девочек младшего школьного возраста в начале учебного года и в конце (с помощью критерия знаков)

При представлении результатов применения парного критерия Вилкоксона рекомендуется указывать медиану, квантили первый и третий, z-значение и абсолютную величину достигнутого уровня значимости

( $p$ ); при критерии знаков – все то же самое, кроме z-значения. Пример представления результатов с применением парного критерия Вилкоксона в табл. 3.

Таблица 3  
Динамические особенности динамометрии правой руки у девочек младшего школьного возраста в разные периоды наблюдения, Me (Q<sub>1</sub>; Q<sub>3</sub>)

Признак	Период наблюдения		z	p
	Начало учебного года	Конец учебного года		
Динамометрия правой руки	13 (11; 14)	15 (12; 18)	2,473	0,013

Следует отметить, что при применении парного критерия Вилкоксона отклоняется нулевая гипотеза об отсутствии различий, а при применении критерия знаков – принимается, то есть в первом случае делается вывод о наличии различий между мышечной силой в начале и в конце учебного года, а во втором – об отсутствии его. Это свидетельствует о том, что критерий знаков обладает меньшей статистической мощностью по сравнению с парным критерием Вилкоксона, поэтому даже при анализе одних и тех же данных можно получить различные результаты и сделать разные выводы, что лишний раз подтверждает необходимость выбора статистического критерия для анализа данных еще на этапе планирования исследования.

**Список литературы**

1. Гланц С. Медико-биологическая статистика / пер. с англ. М. : Практика, 1998. 459 с.
2. Гржибовский А. М. Выбор статистического критерия для проверки гипотез // Экология человека. 2008. № 11. С. 48–57.
3. Гржибовский А. М. Одномерный анализ повторных измерений // Экология человека. 2008. № 4. С. 51–60.
4. Ермолаев О. Ю. Математическая статистика для психологов. 2-е изд. испр. М. : Московский психолого-социальный институт Флинта, 2003. 336 с.
5. Сидоренко Е. В. Методы математической обработки в психологии. СПб. : ООО «Речь», 2004. 350 с.
6. Field A. Discovering statistics using SPSS. 2-nd edition. SAGE Publication, London, 2005. 781 p.
7. Hamilton L. C. Statistics with STATA: Updated for Version 10. USA, 2009. 400 p.
8. Stata: Release 12. Statistical Software. College Station, TX: StataCorp LP, 2011.
9. Wilcoxon F. Individual comparisons by ranking methods // Biometrics Bulletin. 1945. Vol. 1, N 6. P 80–83.

**References**

1. Glantz S. *Primer of biostatistics* [Translated into Russian]. Moscow, Praktika Publ., 1998, 459 p.
2. Grjibovski A. M. Choosing a statistical test for hypothesis testing. *Ekologiya cheloveka* [Human Ecology]. 2008, 11, pp. 48-57. [in Russian]
3. Grjibovski A. M. Univariate analysis of repeated measurements. *Ekologiya cheloveka* [Human Ecology]. 2008, 4, pp. 51-60. [in Russian]
4. Ermolaev O. Yu. *Matematicheskaya statistika dlya psikhologov* [Mathematical statistics for psychologists]. Moscow, Flints psychological and sociological institute, 2003, 336 p.

5. Sidorenko E. V. *Metody matematicheskoi obrabotki v psikhologii* [Methods of mathematical data processing in psychology]. Saint Petersburg, Rech Publ., 2004, 350 p.

6. Field A. *Discovering statistics using SPSS*. 2nd edition. SAGE Publication, London, 2005, 781 p.

7. Hamilton L. C. *Statistics with STATA: Updated for Version 10*. USA, 2009, 400 p.

8. *Stata: Release 12*. Statistical Software. College Station, TX, StataCorp LP, 2011.

9. Wilcoxon F. Individual comparisons by ranking methods. *Biometrics Bulletin*. 1945, 1 (6), pp. 80-83.

#### NON-PARAMETRIC TESTS FOR TWO PAIRED SAMPLES USING STATA

O. A. Kharkova, \*A. M. Grjibovski

*International School of Public Health, Northern State Medical University, Arkhangelsk, Russia*

*\*Department of International Public Health, Norwegian Institute of Public Health, Oslo, Norway*

In this paper we describe the use of non-parametric tests

for comparisons of two paired samples. The authors present examples of manual calculation of Wilcoxon's signed rank test and sign test as well as the use of STATA for calculation of the abovementioned statistical tests. We present assumptions for use of non-parametric tests for paired samples and recommendations on how to present the results in scientific publications.

**Keywords:** statistical analysis, non-parametric tests, Wilcoxon signed rank test, sign test, STATA

#### Контактная информация:

*Гржибовский Андрей Мечиславович* – доктор медицины, профессор, старший советник Норвежского института общественного здравоохранения, г. Осло, Норвегия; директор Архангельской международной школы общественного здоровья ГБУО ВПО «Северный государственный медицинский университет» Министерства здравоохранения Российской Федерации, г. Архангельск

Адрес: Nasjonalt folkehelseinstitutt, Pb 4404 Nydalen, 0403 Oslo, Norway

Тел.: +47 22048319, +47 45268913

E-mail: andrej.grjibovski@gmail.com