ISSN 1993-5757

SPORT SCIENCE OF UKRAINE ***СПОРТИВНА НАУКА УКРАЇНИ***
2013. №1 (52). P. 9-16. ***2013. №1 (52). С. 9-16.***

УДК 796 + 303.214.2

ОЦІНЮВАННЯ НАДІЙНОСТІ ТЕСТУ ПРИ ТРЕНДІ РЕЗУЛЬТАТІВ ТЕСТУВАННЯ

Ігор ЗАНЕВСЬКИЙ, Людмила ЗАНЕВСЬКА

Львівський державний університет фізичної культури

Анотація. Метою роботи було створення методу врахування тренду результатів тестування при оцінюванні надійності моторних тестів за експериментальними даними. Методику дослідження було побудовано на основі теорії спортивних тестів, математичній теорії надійності моторних тестів, методах математичної статистики: кореляційного аналізу, лінійної регресії, дисперсійного аналізу, методу Шапіро–Уілка. Обчислення проводилися з використанням комп'ютерних програм Excel і Statistica. Запропоновано метод елімінації тренду послідовних результатів тестування. На прикладі стрибків у довжину з місця показано, що варіація похибки в оцінці надійності тесту, яку спричинює тренд результатів, становить 11%. Елімінація тренду дозволяє помітно поліпшити оцінку надійності тесту: з прийнятної (*ICC* = 0,884) до доброї (0,933).

Ключові слова: тест, надійність, тренд результатів, міжкласова кореляція.

ОЦЕНКА НАДЕЖНОСТИ ТЕСТА ПРИ ТРЕНДЕ РЕЗУЛЬТАТОВ ТЕСТИРОВАНИЯ

Игорь ЗАНЕВСКИЙ, Людмила ЗАНЕВСКАЯ

Львовский государственный университет физической культуры

Аннотация. Целью работы было создание метода, учитывающего тренд результатов тестирования при оценивании надёжности моторных тестов по экспериментальным данным. Методика исследования была построена на основе теории спортивных тестов, математической теории надёжности моторных тестов, методах математической статистики: корреляционного анализа, линейной регрессии, дисперсионного анализа, метода Шапиро–Уилка. Вычисления проводились с использованием компьютерных программ Excel и Statistica. Предложен метод элиминации тренда последовательных результатов тестирования. На примере прыжков в длину с места показано, что вариация погрешности в оценке надёжности теста, которая вызывает тренд результатов, составляет 11%. Элиминация тренда позволяет заметно улучшить оценку надёжности теста: с приемлемой (*ICC* = 0,884) до хорошей (0,933).

Ключевые слова: тест, надёжность, тренд результатов, межклассовая корреляция

EVALUATION OF RELIABILITY TESTS FOR TREND OF TEST RESULTS

Igor ZANEVSKYY, Ludmyla ZANEVSKA

Lviv State University of Physical Culture

Abstract. The creation of the method, which considers the trend of the results of testing during the evaluation of the reliability of motor tests according to the experimental data, was the purpose of the work. The procedure of study was built on the basis of the theory of sport tests, mathematical theory of the reliability of motor tests, the methods of the mathematical statistics: correlation analysis, linear regression, analysis of variance, Shapiro-Wilk method. Calculations were conducted with the use of Excel and Statistica computer programs. The method of elimination of the trend of the sequential results of testing is proposed. It is shown based on the example of leaps into the length from the place that a variation in the error in the estimation of the reliability of test, which causes the trend of results, composes 11%. The elimination of trend makes it possible to noticeably improve the estimation of the reliability of test with the acceptable (*ISS* = 0,884) to grow as a good (0,933).

Keywords: test, reliability, trend of results, interclass correlation.

Постановка проблеми. Надійність тесту – це міра подібності результатів повторних тестувань [1]. У теорії спортивних тестів надійність разом із інформативністю, стандартністю процедури тестування та наявністю системи оцінок є необхідною вимогою до будь-якого моторного тесту [2].

Кількісною мірою надійності тесту є коефіцієнт надійності, який визначається відношенням дійсної дисперсії до дисперсії, зареєстрованої при тестуванні [3]. Фактори, які спричинюють зниження величини цього коефіцієнта при оцінюванні надійності тесту за експериментальними даними, – це похибки вимірювань та зміна стану досліджуваних при повторних тестуваннях.

© Заневський І., Заневська Л., 2013

Похибки вимірювань можна звести до належної величини застосовуючи відповідну апаратуру та методики тестування. Зміни стану досліджуваних – втоми й тренувального ефекту − позбутися, ба навіть суттєво зменшити, у принципі неможливо.

У загальній теорії тестів вважається, що дійсний результат тесту можна отримати при проведенні нескінченно великої кількості повторних вимірювань тих самих людей у незмінних умовах [4]. У цьому твердженні закладені внутрішні протиріччя. Якщо першу його частину – нескінченність повторювань тестування – можна на практиці розуміти як достатньо велику їх кількість, то забезпечення стандартних умов тестування не можливе в принципі з огляду на невідворотність зміну стану досліджуваних під час тестування. З одного боку, щоб зменшити вплив зовнішніх неконтрольованих факторів, наприклад, геофізичних, метеорологічних, процедура тестування при дослідженні надійності повинна тривати якомога менше часу. З другого боку, для забезпечення однакових умов тестування потрібно давати достатньо великі інтервали часу на відновлення досліджуваних від втоми. Але й тут можна знайти оптимальні інтервали часу на відновлення.

Тренувальний ефект при повторних тестуваннях, якщо в основу тесту покладено рухове завдання, а саме такі тести називають моторними, зменшити в принципі неможливо. А наявність тренувального ефекту спричинює позитивний тренд результатів повторних тестувань, тобто їхнє систематичне зростання. Так само при повторних тестуваннях виникає тренд результатів, причиною якого є стомлення досліджуваних. Цей тренд є від'ємним, але так само, як і додатний тренд, він спричинює додаткову варіацію результатів й погіршує оцінку рівня надійності тесту.

Роботу виконано відповідно до завдань НДР за темою «Застосування проби Руф'є при визначенні групи здоров'я для занять школярів фізичною культурою» плану науково-дослідної роботи Міністерства освіти і науки, молоді та спорту України на 2013 рік.

Аналіз останніх досліджень і публікацій. ІЗ середини минулого століття спостерігається інтенсивний розвиток загальної теорії тестів [5]. Зокрема, центральна увага приділяється математичним методам для оцінювання надійності та інформативності тестів [6]. Розроблено декілька варіантів моделі надійності на основі міжкласового коефіцієнта кореляції [7, 8].

Оцінювання надійності тесту за експериментальними даними при наявності тренду результатів тестування вважається складною проблемою для теорії і практики застосування моторних тестів. В.М. Зациорський [9] зазначає, що для розв'язання цієї проблеми потрібні складні методи, але не описує цих методів. У доступній літературі нам не вдалося знайти робіт з розв'язання цієї проблеми.

Проблема полягає в тому, що для оцінювання надійності тесту, призначеного для поодинокого тестування, необхідно отримати результати декількох ретестів.

Основна ідея дослідження − елімінувати тренд результатів тестування, зберігаючи немонотонні варіації.

Метою роботи було розроблення методу врахування трендів результатів тестування при оцінюванні надійності моторних тестів за експериментальними даними.

Завдання: оцінити придатність математичного апарату теорії спортивних тестів для врахування тренду результатів тестування; розробити метод елімінації тренду послідовних результатів тестування; визначити похибку в оцінці надійності тесту, яку спричинює тренд результатів.

Методику дослідження було побудовано на основі теорії спортивних тестів [10], математичної моделі надійності моторних тестів [11], із застосуванням методів математичної статистики: кореляційного аналізу, лінійної регресії, дисперсійного аналізу, методу Шапіро–Уілка [12]. Обчислення проводилися з використанням комп'ютерних програм Excel (функції CORREL, TREND і підпрограма з пакету "Tools >> Data Analysis >> Anova: Two Factor without Replicaion") і Statistica (Graphs >> Histograms >> Advanced >> Fit type: Normal >> Shapiro-Wilk test).

Основні результати та дискусія. Одинадцять хлопців, учнів п'ятого класу, виконали по

п’ять стрибків у довжину з місця. Між спробами було кілька хвилин відпочинку. Довжина стрибка визначалася з точністю до одного сантиметра (табл. 1).

Таблиця 1

Результати у стрибках в довжину з місця, см

Досліджуваний	Спроба				
	1	2	3	4	5
1	165	164	161	173	165
2	166	176	167	170	167
3	157	152	152	157	143
4	154	155	154	152	169
5	158	155	159	168	170
6	156	151	153	150	159
7	159	160	160	162	155
8	157	152	156	159	166
9	169	168	166	178	174
10	145	151	155	149	163
11	151	160	151	153	155

Оскільки результати тесту Шапіра–Уілка (табл. 2) вказують на нормальність розподілу результатів у всіх трьох спробах ($p > 0{,}05$), для статистичного опрацювання можна використати параметричні методи.

Таблиця 2

Статистичні параметри результатів тестування

Спроба	*M*	*SD*	*V, %*	*SW-W*	*p*
1	158,0	6,9	4,3	0,967	0,851
2	158,4	8,2	5,2	0,862	0,062
3	157,6	5,5	3,5	0,928	0,388
4	160,9	9,9	6,1	0,924	0,351
5	162,2	8,8	5,4	0,934	0,455

Примітка. Умовні позначення до таблиць і формул подано в додатку до статті.

Між результатами за спробами було виявлено кореляційний взаємозв’язок (табл. 3), тому для статистичного опрацювання доцільно застосувати двофакторний дисперсійний аналіз за кореляції даних. Його результати представлено в табл. 4.

Таблиця 3

Кореляційна таблиця результатів за спробами, *r* \ *p*

Спроби	1	2	3	4	5
1	-	0,077	0,051	0,025	0,350
2	0,750	-	0,055	0,088	0,369
3	0,817	0,807	-	0,035	0,137
4	0,901	0,723	0,866	-	0,241
5	0,374	0,358	0,627	0,482	-

Статистична гіпотеза стосовно походження вибіркових сукупностей, що представляють собою результати групи досліджуваних у п’яти спробах, з однієї генеральної сукупності

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5, \quad (1)$$

може бути прийнята на рівні істотності $p = 0{,}143$ (див. табл. 4). Таким чином, для оцінювання надійності тесту на загальноприйнятому рівні істотності ($p = 0{,}05$) можна застосувати міжкласовий коефіцієнт кореляції:

$$ICC = \frac{MS_B \quad MS_W}{MS_B}, \tag{2}$$

де MS_B – дисперсія між досліджуваними; MS_W – дисперсія похибок.

Таблиця 4

Результати дисперсійного аналізу

Варіація	*SS*	*df*	*MS*	*F*	*p*	*F(0,05)*	*Q, %*
Досліджувані	2225	10	222,5	9,233	<0,001	2,077	66,1
Спроби	176	4	44,0	1,825	0,143	2,606	5,2
Взаємозв'язок	964	40	24,1				28,6
Похибки	1140	44	25,9				33,9
Разом	3365	54	62,3				100,0

За даними дисперсійного аналізу, згідно із формулою (2) величина міжкласового коефіцієнта кореляції (0,884) вказує на прийнятний рівень надійності тесту (табл. 5).

Таблиця 5

Шкала оцінки надійності тесту за величиною міжкласового коефіцієнта кореляції [2]

ICC	0,600 – 0,699	0,700 – 0,799	0,800 – 0,899	0,900 – 0,949	0,950 – 1,000
надійність	сумнівна	низька	прийнятна	добра	відмінна

У рівнянні лінійної регресії (рис. 1) коефіцієнт регресії додатний (1,0788), що свідчить про наявність тренду результатів до зростання. Це можна пояснити тренувальним ефектом тестових стрибків. Можна припустити, що тут наявний також вплив стомлення, але загалом тренувальний ефект переважив. Величина тренду при п'яти послідовних спробах дорівнює 4,3 см, що становить 2,7% від середнього арифметичного всіх результатів (159,4 см). Відповідним чином наявність тренду спричинює похибку у значенні міжкласового коефіцієнта кореляції стосовно оцінки надійності тесту за однією спробою.

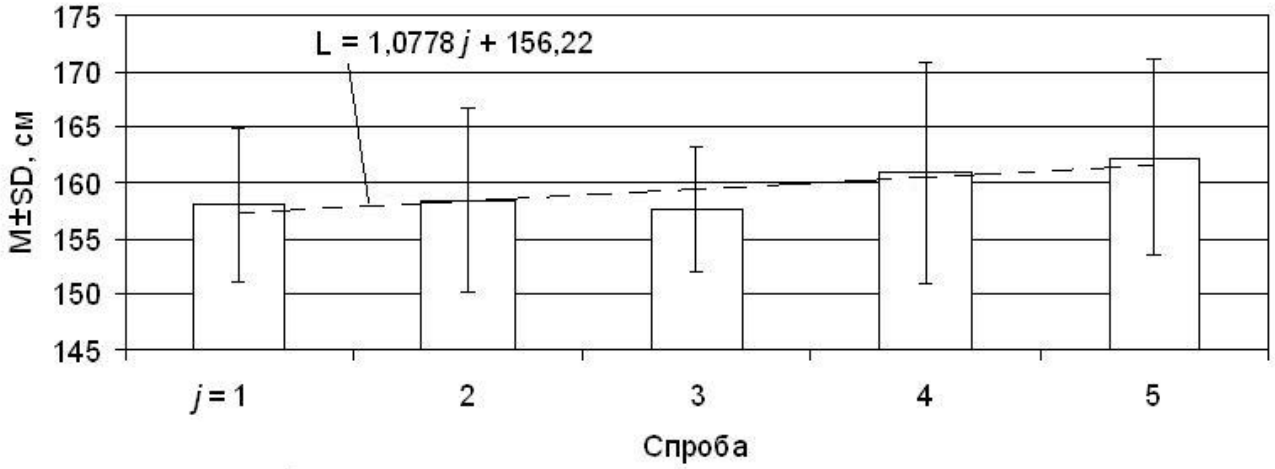

Рис. 1. **Середнє арифметичне (*M*) та середнє квадратичне відхилення (*SD*) групи досліджуваних (*n* = 11) у послідовних спробах із лінією (- - -) та рівнянням регресії**

Для усунення похибки, спричиненої трендом, первинні результати були перетворені за умови відсутності тренду в послідовних спробах при збереженні величини середнього арифме-

тичного результатів досліджуваного в п'яти спробах. Для кожного з одинадцяти досліджуваних із використанням моделі лінійної регресії були записані рівняння тренду результатів тестування:

$$L_{trend} = a + bj, \quad (3)$$

де L_{trend} – значення функції тренду, a – константа рівняння; b – коефіцієнт регресії; j = 1, 2, 3, 4, 5 – порядковий номер спроби (табл. 6). Похибки лінійної апроксимації оцінено абсолютною величиною квадратного кореня коефіцієнта детермінації (|*r*|).

Перетворені результати тестування за умови відсутності тренду обчислювалися за формулою:

$$L_{conv} = L \quad L_{trend} + M_i, \quad (4)$$

де M_i – середні арифметичні значення результатів *i*-того досліджуваного (*i* = 1, 2, …11). Тобто трансформація результатів полягала у відніманні значень функції тренду від збільшених на величину середнього арифметичного первинних даних тестування.

Таблиця 6

Параметри функції тренду результатів тестування

Досліджуваний	Коефіцієнти рівняння регресії		
	a	*b*	|*r*|
1	162,9	0,812	0,285
2	170,9	-0,480	0,183
3	158,8	-2,275	0,631
4	149,1	2,558	0,597
5	150,7	3,760	0,888
6	152,0	0,579	0,255
7	161,4	-0,729	0,471
8	150,9	2,329	0,733
9	165,0	1,951	0,653
10	142,7	3,328	0,794
11	153,9	0,024	0,010

Статистичні параметри перетворених результатів тестування за умови відсутності тренду при послідовних тестуваннях представлено в табл. 7. Загалом вони досить мало відрізняються від первинних результатів за компактністю (див. табл. 2) при тих самих загальних статистиках: середньому арифметичному (*M* = 159,4 см) й стандартному відхиленні (*SD* = 6,7 см) усіх п'ятдесяти п'яти результатів, як і передбачалося за умовами перетворення. Зберігся й закон нормального розподілу перетворених даних (p = 0,387 − 0,895).

Загалом у результатах тестування досліджуваної групи учнів спостерігається тренд зростання (див. рис. 1). У восьми з одинадцяти досліджуваних (*i* = 1, 4, 5, 6, 8, 9, 11 і 12) тренд додатний ($b > 0$), у трьох (*i* = 2, 3 і 7) – від'ємний ($b < 0$, див. табл. 6). Можна припустити одночасну дію двох факторів: тренувального ефекту, який спричинює підвищення результатів послідовних спроб, і стомлення, яке призводить до зниження результатів. Наприклад, у третього досліджуваного переважає ефект тренування при повторних тестуваннях (рис. 2), а в п'ятого переважає стомлення (рис. 3).

Результати дисперсійного аналізу перетворених даних подано в табл. 8. Статистична гіпотеза стосовно походження вибіркових сукупностей, що представляють собою результати групи досліджуваних у спробах, з однієї генеральної сукупності (1) може бути прийнята на рівні істотності p = 0,541 (див. табл. 8).

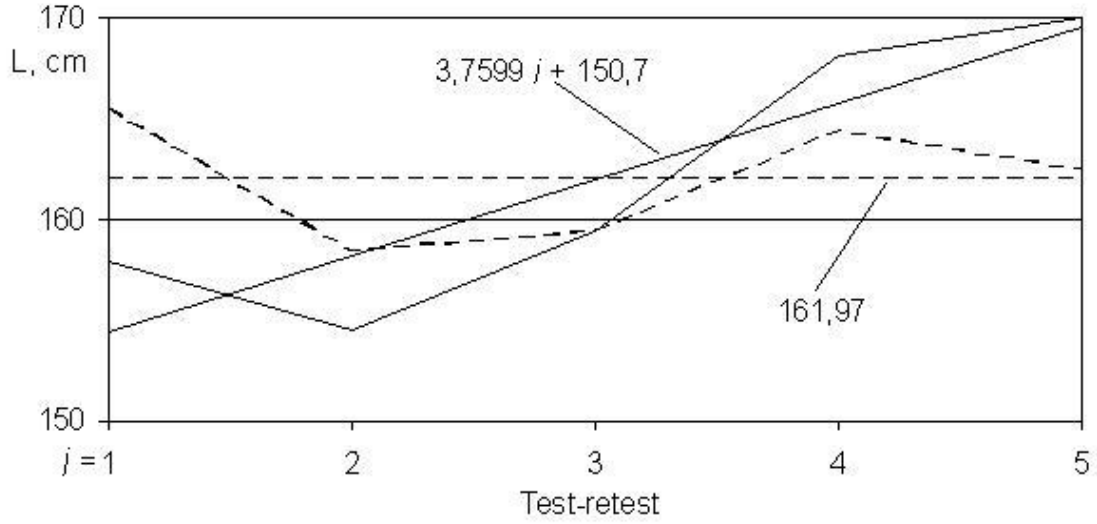

Рис. 2. **Результати досліджуваного (*i* = 5) з перевагою ефекту тренування (—) й перетворені дані за елімінації цього додатного тренду (---): прямими показано лінії регресії**

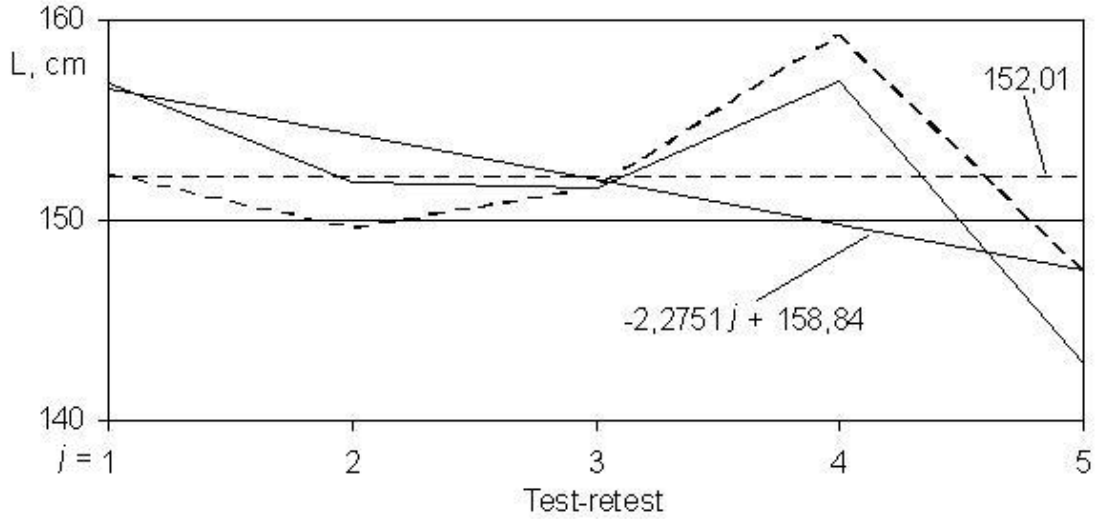

Рис. 3. **Результати досліджуваного (*i* = 3) з перевагою ефекту стомлення (—) й перетворені дані за елімінації цього від'ємного тренду (---): прямими показано лінії регресії**

Таблиця 7

Статистичні параметри перетворених результатів тестування

Спроба	*M*	*SD*	*V*, %	*SW-W*	*p*
1	160,2	7,0	4,4	0,945	0,577
2	159,5	7,9	4,9	0,930	0,406
3	157,6	5,5	3,5	0,928	0,387
4	159,8	10,0	6,3	0,949	0,635
5	160,1	6,4	4,0	0,971	0,895

Таким чином, для оцінювання надійності тесту на відповідному рівні істотності можна знову застосувати міжкласовий коефіцієнт кореляції (2), величина якого (0,933), за даними диспер-

сійного аналізу перетворених результатів, є більшою порівняно з первинним його значенням (0,884). Зростання величини цього коефіцієнта помітне не тільки кількісно, але й якісно, оскільки змінюється оцінка надійності тесту – з прийнятної на добру (див. табл. 5).

Таблиця 8

Результати дисперсійного аналізу перетворених даних

Варіація	*SS*	*df*	*MS*	*F*	*p*	*F(0,05)*	*Q, %*
Досліджувані	2226	10	222,6	14,536	<0,001	2,077	77,1
Спроби	48	4	12,0	0,786	0,541	2,606	1,7
Взаємозв'язок	613	40	15,3				21,2
Похибки	661	44	15,0				22,9
Разом	2887	54	53,5				100,0

Причину такої суттєвої похибки при оцінюванні надійності тесту треба шукати у відносному збільшенні варіації результатів у спробах через наявність тренду на 3,5% і варіації взаємозв'язку – на 7,4% (див. табл. 4 і 8). Загальна величина варіації похибки в дійсності становить 22,9% (див. табл. 8), тобто на 11% менше, ніж обчислена за результатами тестування без елімінації їхнього тренду (див. табл. 4)

Висновки. За міжкласовим коефіцієнтом кореляції не можна виявити вплив тренду результатів на надійність тесту. Наявність тренду в результатах повторних тестувань суттєвим чином, а саме порядку одного інтервалу градації, знижує оцінку надійності тесту. Повної елімінації впливу тренду можна досягнути застосуванням лінійної апроксимації. Наприклад, для стрибків у довжину з місця варіація похибки в оцінці надійності тесту, яку спричинює тренд результатів, становить 11%. Елімінація тренду результатів дозволяє помітно поліпшити оцінку надійності тесту з прийнятної (ICC = 0,884) до доброї (0,933).

Практичні рекомендації. Функцію тренду, у першому наближенні, доцільно апроксимувати рівнянням лінійної регресії у формі (3). Для елімінації тренду до первинних даних потрібно додати середнє арифметичне відповідних результатів і відняти значення функції тренду (4).

Перспективи досліджень у цьому напрямку пов'язані з можливістю нелінійної апроксимації тренду, а також із необхідністю апробації пропонованого методу на широкій гамі моторних тестів.

Додаток

Умовні позначення

a – константа рівняння регресії

b – коефіцієнт рівняння регресії

df – число ступенів вільності

F – критерій Фішера

H_0 – нульова статистична гіпотеза

i – порядковий номер досліджуваного

ICC – коефіцієнт міжкласової кореляції

j – порядковий номер спроби

L – довжина стрибка з місця

L_{conv} – перетворені результати тестування за умови відсутності тренду

L_{trend} – позначення функції тренду

M – середнє арифметичне

MS – дисперсія

n – обсяг групи досліджуваних
p – рівень істотності
Q – частка варіації
r – коефіцієнт кореляції
SD – середнє квадратичне відхилення
SS – сума квадратів відхилення від середнього арифметичного
$SW\text{-}W$ – критерій Шапіра-Уілка
V – коефіцієнт варіації
μ – середнє арифметичне генеральної сукупності

Список літератури

1. *Miller T.* NSCA's Guide to Tests and Assessments / T. Miller. − Champaign, IL: Human Kinetics, 2011. − 368 p.
2. *Уткин В. Л.* Измерения в спорте / В. Л.Уткин. – М. : ГЦОЛИФК, 1978. – 200 с.
3. *Wood T.* Measurement Theory and Practice in Kinesiology / Wood T., Zhu W. – Champaign, IL: Human Kinetics, 2012. − 376 p.
4. *Бубэ Х.* Тесты в спортивной практике / Х. Бубэ, Г. Фэк, Х. Трогаш. – М. : Физкультура и спорт, 1968. − 239 с.
5. *Michael W. B.* Development of Statistical Methods Especially Useful in Test Construction and Evaluation / Michael W. B. // Review of Educational Research. – 1956. − Vol. 26, №. 1. − P. 89-109.
6. *Глас Дж.* Статистические методы в педагогике и психологи / Дж. Глас, Дж. Стэнли. – М.: Прогресс, 1976. − 494 с.
7. *Shrout P. E.* Intraclass correlations: Uses in assessing rater reliability / Shrout P. E., Fleiss J. L. // Psychological Bulletin. – 1979. − Vol 86(2) − P. 420–428 .
8. *Bartko J. J.* The intraclass correlation coefficient as a measure of reliability / Bartko J. J. // Psychological Reports. – 1966. − Vol. 19, is. 1. – P. 3–11.
9. Основы спортивной метрологи / под ред. В. М. Зациорского – М. : Физкультура и спорт, 1979. – 68 с.
10. Measurement and Evaluation in Human Performance / Jr. Morrow, A. Jackson, J. Dich D. Mood − Champaign, IL: Human Kinetics, 2006. − 440 p.
11. *Thomas J.* Research Methods in Physical Activity / J. Thomas, J. Nelson, S. Silverman. – Champaign. − IL : Human Kinetics, 2009. − 472 p.
12. *Vincent W. J.* Statistics in Kinesiology / W.J. Vincent. – Champaign, IL : Human Kinetics, 2005. − 294 p.

Стаття надійшла до редколегії 05.01.2013