

МОДЕЛЬ РЕТЕСТОВОЇ НАДІЙНОСТІ, АЛЬТЕРНАТИВНА МОДЕЛІ ВНУТРІКЛАСОВОЇ КОРЕЛЯЦІЇ

Заневський І.П., Заневська Л.Г.

Львівський державний університет фізичної культури

DOI: 10.17309/tmfv.2014.1.1044

Анотація. Метою роботи було створення моделі коефіцієнта надійності тесту, застосування якої дало б можливість у практичній роботі з тестами отримувати значення коефіцієнта надійності в теоретичних границях від нуля до одиниці, як це прийнято у класичній теорії надійності тестів. Основна ідея розв'язання цієї проблеми полягає у прямому визначенні дисперсії істинних результатів тесту, оцінка величини яких приймається рівною середньому арифметичному результатів повторних вимірювань.

Методику дослідження було побудовано на положеннях класичної теорії спортивних тестів, математичної теорії надійності моторних тестів, на моделі внутрікласової кореляції, дисперсійного аналізу, з використанням методу Шапіра-Уїлка. Обчислення проводилися із застосуванням комп'ютерних програм Excel і Mathematica.

Результати дослідження. Показано, що модель внутрікласової кореляції неповною мірою відповідає положенням класичної теорії надійності тестів. Пропонована модель ретестової надійності немає цього недоліку, а відповідний коефіцієнт існує в межах, визначених положеннями класичної теорії надійності ($0 \div 1$). В діапазоні практично важливих величин пропонована модель та модель внутрікласової кореляції призводять до різних оцінок надійності тесту не тільки кількісно ($\approx 20\%$), але й якісно, тобто із суттєвим перевищенням ширини діапазонів оцінки надійності тесту за величиною відповідного коефіцієнта.

Ключові слова: тест, надійність, внутрікласова кореляція.

Постановка проблеми. Тест, зазвичай, вважається надійним, якщо при його застосуванні отримують ті ж самі результати для кожного обстежуваного при різних тестуваннях в однакових умовах. Для практики найбільшої ваги мають три форми прояву надійності: ретестова надійність, надійність паралельних форм тесту та надійність частин тесту [1].

Ретестова надійність – це спосіб визначення надійності тесту, при якому показники точності та усталеності результатів визначаються шляхом повторного дослідження. Надійність у цьому разі встановлюється як міра збігу результатів першого й повторних обстежень або як ступінь збереження рангових місць обстежених у вибірці при ретесті [2]. Іншими словами ретестова надійність – це стабільність результатів повторних тестувань.

За кількісну міру надійності прийнято коефіцієнт надійності, величина якого дорівнює відношенню дисперсії істинних результатів до дисперсії отриманих результатів тесту [3]. Арифметичне значення кореня квадратного з величини коефіцієнта надійності називають індексом надійності, який є теоретичним коефіцієнтом кореляції між істинними й зареєстрованими результатами тесту. За визначенням коефіцієнт надійності та індекс надійності можуть приймати значення від нуля до одиниці.

Оскільки в практичній роботі з тестами істинні значення результатів тесту невідомі, для визначення величини коефіцієнта надійності застосовують опосередковані методи. Найбільш поширеним з них є метод двофакторного дисперсійного аналізу з визначенням коефіцієнта внутрікласової кореляції як оцінки величини коефіцієнта надійності [4]. Величина цього коефіцієнта дорівнює відношенню різниці дисперсії результатів об'єктів тестування і сумарної дисперсії повторних вимірювань та дисперсії взаємодії з об'єктами тестування до дисперсії результатів об'єктів тестування. Верхньою границею величини коефіцієнта внутрікласової кореляції є одиниця, а нижня границя може бути й від'ємним числом, що входить у протиріччя з визначенням коефіцієнта надійності.

За різними оцінками [5-7] вважаються прийнятними для практичного застосування тести, коефіцієнт надійності яких не менший від числа, яке знаходиться в межах 0,5 ... 0,8. Використовують шкалу оцінок надійності тесту із розділенням загального діапазону прийнятних значень коефіцієнта надійності тесту (тобто, принаймні, додатних) на інтервали: відмінна, добра, задовільна, сумнівна й неприйнятна надійність. Оскільки нульове значення коефіцієнта внутрікласової кореляції не є його мінімальним значенням, відповідна шкала оцінок не повністю відповідає визначенню коефіцієнта, прийнятому у класичній теорії надійності тестів. Тобто, від'ємні значення нижньої границі діапазо-

Умовні позначення

F	критерій Фішера
f	стала, яка характеризує обсяг групи досліджуваних та кількість ретестових спроб
k	кількість вимірів (спроб)
M	середнє арифметичне
S^2, MS	дисперсія
n	обсяг групи досліджуваних
p	рівень істотності
Q	частка загальної варіації за квадратами відхилень від середнього арифметичного
r_t	коефіцієнт надійності тесту
r_i	індекс надійності тесту
ICC	внутрікласовий коефіцієнт кореляції
S	сума квадратів відхилення від середнього арифметичного
SW — W	критерій Шапіра-Уїлка
df	число ступенів свободи
Ho	нульова гіпотеза
μ	середнє арифметичне генеральної сукупності

ну величин коефіцієнта внутрікласової кореляції суперечать моделі коефіцієнта надійності тесту, в якій ця границя є нульовою.

Отож у теорії спортивних тестів існує проблема неповної відповідності моделі внутрікласової кореляції теоретичній моделі надійності тесту.

Роботу виконано у рамках завдань НДР по темі «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 (№ держ. реєстрації 0113U000658).

Аналіз попередніх досліджень і публікацій. Основи класичної теорії тестів розроблено в середині минулого століття Г. Гулліксоном [8], а пізніше розвинуто Ф.М. Лордом і М.Р. Новіком [9], а також Л. Крокером і Дж. Алгино [10] стосовно теорії психометрії. Початки дисперсійного аналізу та внутрікласової кореляції, які застосовують для оцінки величини коефіцієнта надійності тесту, були закладені наприкінці дев'ятнадцятого століття Р.А. Фішером [11]. Обґрунтуванню різних варіантів методу внутрікласової кореляції для оцінки величини коефіцієнта надійності тесту присвячено

роботи Дж. Бартко [4], Л.Дж. Кронбаха, Г.К. Глєсера, Г. Нанда й Н. Раяратнам [12], а також П.Е. Шпраута із Дж.Л. Фляйссем [13].

Специфіці застосування теорії тестів у спортивній практиці присвячено роботи Х. Бубе зі співавторами (Н. Bube, G. Feck, H. Stubler, F. Trogsch) [14], а також В.Л. Уткіна [15], В.М. Заціорського [16] й У.Г. Гопкінса [17]. Найбільш докладно стан проблеми надійності моторних тестів розглянуто П. О'Донохью [18].

Починаючи з останньої декади минулого століття математична теорія надійності тестів розвивалася у напрямі створення моделей, альтернативних моделі внутрікласової кореляції. Дж.М. Бланд і Д.Г. Альтман запропонували метод оцінки узгодженості двох тестів з використанням стандартного відхилення повторних спроб [19]. А.М. Невілл і Г. Аткинсон розробили метод визначення коефіцієнта надійності для вимірів, проведених у шкалі відношень, на основі аналізу кількості вимірів, які потрапили до 95-процентного інтервалу довіри [20].

На початку нового століття А. Брутон, Дж.Х. Конвей і С.Т. Голгейт підвели ризику під піввіковим періодом інтенсивного розвитку математичних методів наближеної оцінки надійності тестів, представивши порівняльний аналіз запропонованих моделей і визначивши їхні переваги та недоліки [21]. У.Г. Гопкінс доводить, що найточнішим параметром для кількісної оцінки надійності тесту є коефіцієнт варіації результатів повторних спроб [17].

На основі аналізу публікацій останньої декади може скластися враження, що математична теорія надійності тестів досягла завершеності, оскільки головні її здобутки були пов'язані зі створенням відповідного програмного забезпечення загального користування, наприклад, в комп'ютерних пакетах Statistica [22], SPSS [23], Mathematica [24], Mathcad [25] та ін.

У відкритих джерелах інформації останніх трьох декад нами не виявлено публікацій стосовно проблеми, що розглядається.

Метою роботи було створення моделі коефіцієнта надійності тесту, застосування якої дало б можливість у практичній роботі з тестами отримувати значення коефіцієнта надійності в теоретичних границях від нуля до одиниці, як це прийнято у класичній теорії надійності тестів. Основна ідея розв'язання цієї проблеми полягає у прямому визначенні дисперсії істинних результатів тесту, оцінка величини яких приймається рівною середньому арифметичному результатів повторних вимірювань.

Завдання дослідження: показати неповну відповідність моделі внутрікласової кореляції класичній теорії ретестової надійності; розробити модель коефіцієнта надійності, який би загалом відповідав цій теорії; провести кількісний аналіз подібності моделі внутрікласової кореляції та альтернативної моделі.

Методика дослідження ґрунтувалася на положеннях класичної теорії спортивних тестів [5], на математичній теорії надійності моторних тестів [3], на моделі внутрікласової кореляції [11], на дисперсійному аналізі й на методі Шапіра – Уїлка [4]. Для визначення рівних величин коефіцієнта внутрікласової кореляції й альтернативного коефіцієнта надійності тесту було складено алгебричне квадратне рівняння, розв’язки якого були отримані в радикалах.

Обчислення проводилися з використанням комп’ютерної програми Excel (функції CORREL, VAR, AVERAGE, а також підпрограми з пакету “Tools >> Data Analysis >> Anova: Two Factor without Replication”) й програми Mathematica (функція Graphs >> Histograms >> Advanced >> Fit type: Normal >> Shapiro-Wilk test).

Основні результати. Застосування класичної моделі коефіцієнта надійності тесту описано у Morgow [5, с.86] на прикладі вимірювань артеріального тиску з використанням фонендоскопу й механічного тонометра. У десяти пацієнтів систолічний тиск визначався паралельно цим методом і методом безпосереднього вимірювання (табл. 1). Деталі цього другого методу в роботі не подано, але відповідні результати прийнято за істинні, що для теоретичних розважань можна прийняти без вичерпної конкретизації.

Таблиця 1

Систолічний тиск

Пацієнт	мм рт. стовп.		
	Механічний тонометр	Істинна величина артеріального тиску	Похибка
1	103	105	-2
2	117	115	2
3	116	120	-4
4	123	125	-2
5	127	125	2
6	125	125	0
7	135	125	10
8	126	130	-4
9	133	135	-2
10	145	145	0
M*	125	125	0
S ²	133,6	116,7	16,9
SW-W	0,979	0,955	-
p(SW-W)	0,961	0,731	-

* M – середнє арифметичне; S² – дисперсія; SW-W – критерій Шапіра – Уїлка; p(SW-W) – рівень істотності.

Обидві вибірки – тестова й та, що складається з точних значень тиску – отримані з генеральних сукупностей, які підлягають нормальному закону розподілу (p ≥ 0,731). Величини похибок мають випадковий характер, тому їхня середня величина дорівнює нулю, а середні арифметичні тесту та істинних результатів однакові.

Коефіцієнт надійності тесту обчислено за формулою:

$$r_t = \frac{S_t^2}{S_o^2} = \frac{S_o^2 - S_e^2}{S_o^2}, \quad (1)$$

де S₀² – дисперсія результатів тесту (за вимірами механічного тонометра); S_t² – дисперсія істинних результатів; S_e² – дисперсія похибок вимірювань.

Для чисельних величин дисперсії (див. табл. 1) формула (1) дає результат r_t = 0,874, який вказує на прийнятну надійність цього тесту [7]. Величина індексу надійності дорівнює додатному значенню квадратного кореня з величини коефіцієнта надійності:

$$r_i = \sqrt{r_t}. \quad (2)$$

З огляду на характер розподілу результатів тесту та істинних значень, величина індексу дорівнює величині парного лінійного коефіцієнта кореляції Брауе-Пірсона (r_t = 0,935), величина якого вказує на високу статистичну достовірність існування кореляційного взаємозв’язку (p < 0,001), а отже – й на високу надійність тесту.

Модель внутрікласової кореляції

Внутрікласовий коефіцієнт кореляції застосовано для оцінки надійності тесту серії штрафних кидків у баскетболі [6, с. 68]. Відповідні результати трьох спроб шести юних баскетболістів подано у табличній формі (табл. 2).

Величина внутрікласового коефіцієнта кореляції визначається за такою формулою [13]:

$$ICC = \frac{MS_B - MS_W}{MS_B}, \quad (3)$$

де MS_B – дисперсія між спортсменами; MS_W – дисперсія похибок, тобто спільної варіації результатів між спортсменами і варіації взаємозв’язку результатів у спробах та результатів між спортсменами.

Для визначення величини дисперсій проведено двофакторний дисперсійний аналіз за кореляції даних, результати якого зведено в таблиці 3.

Величина внутрікласового коефіцієнта кореляції за формулою (3) із використанням результатах дисперсійного аналізу (див. табл. 3) вказує на низький рівень надійності тесту (ICC = 0,425), хоча статистична гіпотеза стосовно походження

Таблиця 2
Результати трьох серій штрафних кидків юних баскетболістів (низька надійність тесту)

Спортсмен	Спроба			М
	1	2	3	
1	5	6	5	5,3
2	9	8	7	8,0
3	3	4	3	3,3
4	7	5	5	5,7
5	9	2	9	6,7
6	7	3	7	5,7

ненадійного тесту можна розглянути гіпотетичні результати штрафних кидків юних баскетболістів, зведені в табл. 4.

Нульова величина внутрікласового коефіцієнта кореляції буде мати місце за такої умови:

$$n(k-1)SS_R = (n-1)(SS_T - SS_B), \quad (5)$$

де n – кількість досліджуваних (спортсменів); k – кількість вимірів (спроб); SS_B – сума квадратів відхилення від середнього результату між спортсменами; SS_T – сума квадратів відхилення від середнього між всіма вимірами загалом.

Відповідні результати двофакторного дисперсійного аналізу зведено в табл. 5.

Таблиця 3
Результати дисперсійного аналізу трьох серій штрафних кидків юних баскетболістів (див. табл. 2)

Джерело варіації	SS*	df	MS	F	p	F0,05	Q, %
Спортсмени (B)	35,8	5	7,2	1,940	0,174	3,326	42,0
Спроби	12,4	2	6,2	1,687	0,234	4,103	14,6
Взаємозв'язок	36,9	10	3,7	-	-	-	43,3
Похибки (W)	49,3	12	4,1	-	-	-	58,0
Разом (T)	85,1	17	5,0	-	-	-	100,0

* SS – сума квадратів відхилень від середнього арифметичного; df – число ступенів свободи; F – критерій Фішера – Снедекора; Q – відсоток внеску до загальної варіації результатів.

Таблиця 4
Гіпотетичні результати трьох серій штрафних кидків при зовсім ненадійному тесті

Спортсмен	Спроба			М
	1	2	3	
1	7	5	5	5,7
2	9	8	7	8,0
3	3	4	5	4,0
4	9	5	5	6,3
5	9	2	8	6,3
6	7	2	7	5,3

вибіркових сукупностей, що представляють собою результати спортсменів у спробах, з однієї генеральної сукупності (μ – середнє арифметичне генеральної сукупності),

$$H_0: \mu_1 = \mu_2 = \mu_3, \quad (4)$$

може бути прийнята із прийнятним рівнем істотності $p = 0,234$ (див. табл. 3).

Якщо результати всіх спроб кожного зі спортсменів були б однакові ($MS_W = 0$), величина внутрікласового коефіцієнта кореляції дорівнювала б одиниці, що свідчило б про абсолютну надійність тесту. У протилежному випадку, коли б дисперсія результатів між спортсменами дорівнювала за величиною дисперсії похибок ($MS_B = MS_W$), величина внутрікласового коефіцієнта кореляції дорівнювала б нулю, що свідчило б про абсолютну ненадійність тесту. Як приклад близького до абсолютно

величина внутрікласового коефіцієнта кореляції за формулою (3) при результатах дисперсійного аналізу (див. табл. 5) вказує на майже повну ненадійність тесту ($ICC = 0,006$), а статистична гіпотеза стосовно походження вибірових сукупностей, що представляють собою результати групи досліджуваних у спробах, з однієї генеральної сукупності (4) не може бути прийнята на достатньо високому рівні істотності: $p = 0,056$ (див. табл. 5).

Ці два гіпотетичні випадки ($ICC = 1$ і $ICC = 0$) можна розглядати як теоретичні границі реальних властивостей моторних тестів у сенсі їхньої надійності. Але у практиці застосування моторних тестів часто трапляються випадки, коли величина внутрікласового коефіцієнта кореляції стає від'ємною. Наприклад, для того ж таки тесту штрафних кидків юних баскетболістів із гіпотетичними результатами (табл. 6).

Відповідні результати двофакторного дисперсійного аналізу зведено в табл. 7 ($MS_B < MS_W$).

Пропонована модель коефіцієнта надійності тесту

Від'ємне значення внутрікласового коефіцієнта кореляції ($ICC = -1,300$), обчислене за формулою (3) при результатах дисперсійного аналізу з табл. 7, суперечить основам класичної теорії надійності тестів, згідно з якою коефіцієнт надійності не може бути від'ємним числом (за визначенням: $0 \leq r_{tt} \leq 1$).

Цього протиріччя можна уникнути, приймаючи модель коефіцієнта надійності тесту, в якій за

Таблиця 5

Результати дисперсійного аналізу трьох серій штрафних кидків (див. табл. 4)

Джерело варіації	SS	df	MS	F	p	F0,05	Q, %
Спортсмени (B)	26,3	5	5,3	1,492	0,276	3,326	29,5
Спроби	27,4	2	13,7	3,896	0,056	4,103	30,9
Взаємозв'язок	35,2	10	3,5	-	-	-	39,6
Похибки (W)	62,7	12	5,2	-	-	-	70,5
Разом (T)	88,9	17	5,2	-	-	-	100,0

Таблиця 6

Гіпотетичні результати трьох серій штрафних кидків, коли величина внутрікласового коефіцієнта кореляції від'ємна

Спортсмен	Спроба			M
	1	2	3	
1	7	5	5	5,7
2	5	6	6	5,7
3	3	4	5	4,0
4	9	5	5	6,3
5	9	2	8	6,3
6	7	2	7	5,3

істинні значення взяти середні арифметичні індивідуальних результатів повторних спроб, оскільки середнє арифметичне – це найкраще наближення для оцінки істинної величини у генеральній сукупності. Вихідні дані й проміжні результати для основного прикладу трьох серій штрафних кидків юних баскетболістів з табл. 2, були опрацьовані відповідно до цієї моделі й зведені в таблиці 8.

Величина коефіцієнта надійності ($r_{tt} = 0,420$), обчисленого за формулою (1), вказує на низький рівень надійності цього тесту, однак величина індексу надійності ($r_t = 0,648$) свідчить про досить високу статистичну достовірність існування кореляційного взаємозв'язку ($p < 0,004$).

Особливо показовими є результати аналізу надійності тесту з табл. 4, для яких отримано близьке до нуля значення внутрікласового коефіцієнта кореляції ($ICC = 0,006$). Відповідні результати трьох серій штрафних кидків представлено таблицею 9.

Величина коефіцієнта ($r_{tt} = 0,295$), обчисленого за формулою (1), вказує на низький рівень надійності цього тесту, однак величина індексу надійності тесту ($r_t = 0,544$), подібно до попереднього прикладу (див. табл. 8), дозволяє із достатньо високою статистичною достовірністю визнати існування кореляційного взаємозв'язку ($p < 0,020$).

Для випадку, коли величина внутрікласового коефіцієнта кореляції від'ємна (див. табл. 6), зведені в одну вибірку результати трьох серій штрафних кидків подано в таблиці 10.

Величина коефіцієнта, обчисленого за формулою (1) при величинах дисперсій з табл. 10, вказує на низький рівень надійності цього тесту ($r_{tt} = 0,153$). Однак величина індексу надійності тесту, обчисленого за формулою (2), яка дорівнює величині парного лінійного коефіцієнта кореляції істинних та зафіксованих результатів тесту ($r_t = 0,392$), вказує на певну статистичну достовірність існування кореляційного взаємозв'язку ($p = 0,108$).

Гіпотетичний варіант абсолютно надійного тесту, коли у всіх трьох спробах результати кожного зі спортсменів залишаються незмінними, наведено в таблиці 11. У цьому випадку мають місце однакові значення всіх трьох коефіцієнтів: $ICC \equiv r_{tt} \equiv r_t = 1$.

Величини коефіцієнта надійності тесту, який складався з трьох серій штрафних кидків юних баскетболістів, обчислених за моделлю внутрікласового коефіцієнта кореляції й за альтернативною моделлю, а також відповідні значення індексу надійності зведено в таблиці 12.

Дискусія. Результати оцінки надійності тесту, отримані із застосуванням пропонованої моделі, органічно відповідають основам класичної теорії

Таблиця 7

Результати дисперсійного аналізу трьох серій штрафних кидків (див. табл. 6)

Джерело варіації	SS	df	MS	F	p	F0,05	Q, %
Спортсмени (B)	11,1	5	2,2	0,581	0,714	3,326	15,3
Спроби	23,1	2	11,6	3,023	0,094	4,103	31,9
Взаємозв'язок	38,2	10	3,8	-	-	-	52,8
Похибки (W)	61,3	12	5,1	-	-	-	84,7
Разом (T)	72,4	17	4,3	-	-	-	100,0

Таблиця 8
Результати трьох серій штрафних кидків юних баскетболістів (див. табл. 2), зведені в одну вибірку

Спортсмен	Результат тесту		
	Зафіксований	Істинний	Похибка
1	5,0	5,3	-0,3
2	9,0	8,0	1,0
3	3,0	3,3	-0,3
4	7,0	5,7	1,3
5	9,0	6,7	2,3
6	7,0	5,7	1,3
1	6,0	5,3	0,7
2	8,0	8,0	0,0
3	4,0	3,3	0,7
4	5,0	5,7	-0,7
5	2,0	6,7	-4,7
6	3,0	5,7	-2,7
1	5,0	5,3	-0,3
2	7,0	8,0	-1,0
3	3,0	3,3	-0,3
4	5,0	5,7	-0,7
5	9,0	6,7	2,3
6	7,0	5,7	1,3
M	5,8	5,8	0
S ²	5,007	2,105	2,902
SW-W	0,934	0,966	-
p(SW-W)	0,225	0,328	-

Таблиця 9
Результати трьох серій штрафних кидків юних баскетболістів (див. табл. 4), зведені в одну вибірку

Спортсмен	Результат тесту		
	Зафіксований	Істинний	Похибка
1	7,0	5,7	1,3
2	9,0	8,0	1,0
3	3,0	4,0	-1,0
4	9,0	6,3	2,7
5	9,0	6,3	2,7
6	7,0	5,3	1,7
1	5,0	5,7	-0,7
2	8,0	8,0	0,0
3	4,0	4,0	0,0
4	5,0	6,3	-1,3
5	2,0	6,3	-4,3
6	2,0	5,3	-3,3
1	5,0	5,7	-0,7
2	7,0	8,0	-1,0
3	5,0	4,0	1,0
4	5,0	6,3	-1,3
5	8,0	6,3	1,7
6	7,0	5,3	1,7
M	5,9	5,9	0,0
S ²	5,232	1,546	3,686
SW-W	0,922	0,898	-
p(SW-W)	0,138	0,052	-

надійності (1). Стає можливим отримати кількісну характеристику надійності для будь-яких теоретично можливих результатів тестування, зокрема тоді, коли значення внутрікласового коефіцієнта кореляції від'ємне. Наприклад, величина коефіцієнта надійності для результатів тесту з табл. 6 ($r_{tt} = 0,153$) має природний смисл на протипагу, величині внутрікласового кореляційного аналізу ($ICC = -1,300$), отримано за результатами дисперсійного аналізу (див. табл. 7).

Співвідношення між величинами внутрікласового коефіцієнта кореляції й коефіцієнта надійності, визначеного за запропонованою моделлю, можна отримати з використанням таких залежностей:

$$MS_B = \frac{(SS_B + SS_W)Q_B}{100(n-1)} ;$$

$$MS_W = \frac{(SS_B + SS_W)(100 - Q_B)}{100n(k-1)} ; r_{tt} = \frac{Q_B}{100}, \quad (6)$$

де $Q_B = \frac{SS_B}{SS_B + SS_W} \cdot 100\%$ – частка варіації результатів між об'єктами тестування (тобто між спортсменами) у загальній варіації (див. табл. 3, 5, 7). За формулами (2), (3) і (6) побудовано відповідні графіки для $n = 6$ і $k = 3$ (рис. 1).

Таблиця 10
Результати трьох серій штрафних кидків у юних баскетболістів (див. табл. 6), зведені в одну вибірку

Спортсмен	Результат тесту		
	Зафіксований	Істинний	Похибка
1	7,0	5,7	1,3
2	5,0	5,7	-0,7
3	3,0	4,0	-1,0
4	9,0	6,3	2,7
5	9,0	6,3	2,7
6	7,0	5,3	1,7
1	5,0	5,7	-0,7
2	6,0	5,7	0,3
3	4,0	4,0	0,0
4	5,0	6,3	-1,3
5	2,0	6,3	-4,3
6	2,0	5,3	-3,3
1	5,0	5,7	-0,7
2	6,0	5,7	0,3
3	5,0	4,0	1,0
4	5,0	6,3	-1,3
5	8,0	6,3	1,7
6	7,0	5,3	1,7
M	5,6	5,6	0,0
S ²	4,261	0,654	3,608
SW-W	0,944	0,943	-
p(SW-W)	0,345	0,323	-

Таблиця 11

Гіпотетичні результати трьох серій штрафних кидків юних баскетболістів за абсолютної надійності тесту

Спортсмен	Спроба			М
	1	2	3	
1	5	5	5	5
2	9	9	9	9
3	3	3	3	3
4	7	7	7	7
5	9	9	9	9
6	7	7	7	7

($ICC=r_{tt}=f$). У межах між цими величина внутрікласового коефіцієнта кореляції більша від величини пропонованого коефіцієнта ($ICC>r_{tt}$). Притому, різниця між величиною цих двох коефіцієнтів є максимальною й дорівнює $(1-\sqrt{f})^2$, коли $Q_B=\sqrt{f}\times 100\%$, що складає (для прикладу, який розглядається) 19,5% величини коефіцієнта надійності. При $Q_B<f\times 100\%$ величина внутрікласового коефіцієнта кореляції

стрімко зменшується, обертаючись на нуль при

$$Q_B = \frac{f}{1+f} 100\%.$$

При $Q_B = 0$ його величина прямує до від'ємної нескінченності ($ICC \rightarrow -\infty$).

Таблиця 12

Результати аналізу надійності тесту з трьох серій штрафних кидків

Результати тесту	S_t^2	S_0^2	S_e^2	r_{tt}	r_t	P	ICC
Табл. 2	6,557	4,943	4,826	0,420	0,648	0,004	0,425
Табл. 4	5,232	1,546	3,686	0,295	0,544	0,020	0,006
Табл. 6	4,261	0,654	3,608	0,153	0,392	0,108	-1,300
Табл. 11	2,196	2,196	0	1	1	0	1

З урахуванням (6) формулу (1) можна записати у такому вигляді:

$$r_{tt} = \frac{(n-1)MS_B}{(n-1)MS_B + n(k-1)MS_W}, \quad (7)$$

що дає можливість встановити кількісне співвідношення величини r_{tt} й ICC. Шляхом алгебричних перетворень рівнянь (3), (6) і (7) отримано залежність величини ICC від величини r_{tt} :

$$ICC = 1 - \frac{1-r_{tt}}{r_t} f, \quad (8)$$

де $f = \frac{n-1}{n(k-1)}$ – стала, яка характеризує обсяг групи досліджуваних та кількість ретестових спроб.

Ці два коефіцієнти, обчислені за формулами (7) і (8), є однаковими за величиною у двох випадках: коли $Q_B = 100\%$ ($ICC=r_{tt}=1$) і коли $Q_B=f\times 100\%$

лежність коефіцієнта надійності тесту від частки варіації результатів між об'єктами тестування (Q_B) на цілому діапазоні є лінійною (6).

Для практики застосування моторних тестів важливим можна вважати діапазон $Q_B>f\times 100\%$. Для розглянутого тесту штрафних кидків – це $41,7 \div 100\%$, коли коефіцієнт надійності знаходиться в межах $0,417 \div 1$. Максимальна різниця (0,126) між оцінками коефіцієнта надійності за двома моделями, що розглядаються, має місце, коли $ICC = 0,771$, а $r_{tt} = 0,645$. Згідно зі шкалою оцінки надійності тесту за величиною внутрікласового коефіцієнта кореляції (табл. 13), ширина діапазону загальноприйнятих рівнів надійності дорівнює 0,05 або 0,10. Обчислена різниця (0,126) суттєво перевищує ширину відповідних діапазонів ($0,6 \div 0,7 \div 0,8$). За моделлю внутрікласового коефіцієнта кореляції надійність тесту має бути визнана низькою, а за альтернативною моделлю – сумнівною. Тобто в діапазоні практично важливих величин ці дві моделі призводять до різних оцінок надійності тесту не тільки кількісно, але й якісно.

Окрім розглянутої моделі надійності тесту із використанням коефіцієнта внутрікласової кореляції, в теорії тестів розроблено ще п'ять моделей [3], які застосовуються переважно у психології для визначення надійності експертних оцінок. Вибір певної моделі внутрікласової кореляції залежить від типу дисперсійного аналізу (одно- чи двофакторний) та способу виведення експертних оцінок (індивідуальний, груповий чи комбінований) [13].

Класичний приклад оцінки шести об'єктів бригадою чотирьох експертів представлено Дж. Бартко [4] (табл. 14).

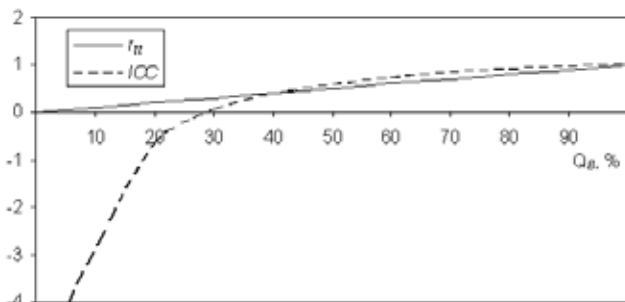


Рис. 1. Графіки залежностей величини коефіцієнта надійності тесту, обчисленими за моделлю внутрікласової кореляції (ICC) й за пропонованою моделлю r_{tt} .

Таблиця 13

Шкала оцінки надійності тесту за величиною внутрікласового коефіцієнта кореляції [6]

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

Відповідні величини коефіцієнтів внутрікласової кореляції (ICC), альтернативного коефіцієнта надійності (r_{tt}), а також відносна різниця між ними у відсотках,

$$\frac{ICC - r_{tt}}{ICC + r_{tt}} 200\%, \quad (9)$$

зведені в табл. 15.

Різниця у величині коефіцієнта надійності, визначеного у рамках пропонованої моделі (r_{tt}) з величиною коефіцієнтів внутрікласової кореляції (ICC) у прикладі, що розглядається (див. табл. 15), складає 157,5%. Два з внутрікласових коефіцієнтів (0,17 і 0,29) є меншими за величиною від пропонованого коефіцієнта надійності, решта чотири (0,71; 0,44; 0,62 і 0,91) – більшими. Модель внутрікласової кореляції (ICC (1, 4) = 0,443), аналогічна прикладу ретестової надійності для результатів трьох серій штрафних кидків юних баскетболістів (див. табл. 12), дає значення коефіцієнта надійності на 28,6% завищене, порівняно зі значенням, отриманим із застосуванням пропонованої моделі ($r_{tt} = 0,33$).

Висновки. Модель внутрікласової кореляції неповною мірою відповідає положенням класичної теорії надійності: величина відповідного коефіцієнта ретестової надійності, при її погіршенні, прямує до від'ємної нескінченності. Пропонована модель ретестової надійності позбавлена цього недоліку, а відповідний коефіцієнт надійності існує в межах, визначених положеннями класичної теорії надійності ($r_{tt} = 0 \div 1$).

Найкраще наближення до істинного результату – це середнє арифметичне послідовних тестувань. В діапазоні практично важливих величин пропонована модель та модель внутрікласової кореляції призводять до різних оцінок надійності тесту не тільки кількісно ($\approx 20\%$), але й якісно, тобто із суттєвим перевищенням ширини діапазонів оцінки надійності тесту за величиною відповідного коефіцієнта.

Список літератури

- Jacobs L.C. Test Reliability, 1991. [Електронний ресурс]. – Режим доступу: <http://www.indiana.edu/~best/bweb3/test-reliability/>
- Cherry K. What Is Test-Retest Reliability, 2003. [Електронний ресурс]. – Режим доступу: <http://psychology.about.com/od/tindex/g/test-retest-reliability.htm>
- Machowski A. Rzetelnosc testow psychologicznych. Dwa ujecia modelowe. – Warszawa – Poznan: Wydawnictwo Naukowe PWN, 1993. – 82 s.

Таблиця експертних оцінок [4]

Об'єкт	Експерт			
	1	2	3	4
1	9	2	5	8
2	6	1	3	2
3	8	4	6	8
4	7	1	2	6
5	10	5	6	9
6	6	2	4	7

Таблиця 14

Таблиця 15

Коефіцієнт надійності тесту за різних моделей внутрікласової кореляції

Позначення	Величина	Різниця (9), %
ICC (1, 1)	0,17	-64,0
ICC (2, 1)	0,29	-12,9
ICC (3, 1)	0,71	73,1
ICC (1, 4)	0,44	28,6
ICC (2, 4)	0,62	61,1
ICC (3, 4)	0,91	93,5
r_{tt}	0,33	–

Практичні рекомендації. Для обчислення дисперсій результатів тесту доцільно застосувати функцію Excel VAR.

Перспективи досліджень у даному напрямку пов'язані з можливістю розповсюдження пропонованої моделі коефіцієнта надійності на тести, результати яких визначаються у ранговій шкалі, а також з необхідністю апробації пропонованої моделі на ширшій гамі моторних тестів.

References

- Jacobs L.C. (1991). Test Reliability, [Electronic resource]. — Access: <http://www.indiana.edu/~best/bweb3/test-reliability/>
- Cherry K. (2003). What Is Test-Retest Reliability, [Electronic resource]. — Access: <http://psychology.about.com/od/tindex/g/test-retest-reliability.htm>
- Machowski A. (1993). Rzetelnosc testow psychologicznych. Dwa ujecia modelowe. Warszawa – Poznan: Wydawnictwo Naukowe PWN, 82 s.

4. Bartko J.J. The intraclass correlation coefficient as a measure of reliability // *Psychological Reports*. – 1966. – Volume 19, Issue, pp. 3-11.
5. Morrow Jr., Jackson A., Dich J., Mood D. Measurement and Evaluation in Human Performance. – Champaign, IL: Human Kinetics, 2006. – 440 p.
6. Спортивная метрология. Учеб.для ин-тов физ. культ. / Под ред. В.М. Зацюрского. – М.: Физкультура и спорт, 1982. – 256 с.
7. Морозов С. М. Засоби контролю діагностичних якостей психологічних тестів. — К.: ІСДЮ, 1994. — 68 с.
8. Gulliksen H. Theory of mental tests. – New York: J. Wiley, 1950, – 486 p.
9. Lord F.M., Novick R.M. Statistical theories of mental test scores. – Reading, MA: Addison-Wesley, 1986. – 192 p.
10. Crocker L., Algina, J. Introduction to classical and modern test theory. – New York: Holt, Rinehart, and Winston, 1986. – 218 p.
11. Fisher R.A. Statistical Methods for Research Workers. – London: Oliver and Boyd, 1925. – 276 p.
12. Cronbach L.J., Gleser G.C., Nanda G., Rajratnam N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. – New York: J. Wiley, 1972. – 152 p.
13. Shrout P.E., Fleiss J.L. Intraclass correlations: Uses in assessing rater reliability // *Psychological Bulletin*. – 1979. – Vol 86(2) – pp. 420-428 .
14. Тесты в спортивной практике (пер. с нем. / сост. Х. Бубэ и др.). – М.: Физкультура и спорт, 1968. – 239 с.
15. Уткин В.Л. Измерения в спорте (введение в спортивную метрологию). – М.: ГЦОЛИФК, 1978. – 200 с.
16. Зацюрский В.М. Основы спортивной метрологии. – М.: Физкультура и спорт, 1979. – 68 с.
17. Hopkins W.G. Measures of Reliability in Sports Medicine and Science // *Sports Medicine*, 2000, 30 (1), 1-15.
18. O'Donoghue P. Research methods for sports performance analysis. – New York, NY: Routledge, 2010. – 278 p.
19. Bland J.M., Altman D.G. Statistical methods for assessing agreement between two methods of clinical measurement // *Lancet*, 1986 Feb; 8: 307-310.
20. Nevill A.M., Atkinson G. Assessing agreement between measurements recorded on a ratio scale in sports medicine and sports science // *British Journal of Sports Medicine*, 1997; 31: 314-318.
21. Bruton A, Conway J.H., Holgate S.T. Reliability: What is it and how is it measured? // *Physiotherapy*, 2000, 86, 2, 94-99.
22. Statistica www.statsoft.com
23. SPSS www.ibm.com/software/analytics/spss/
24. Mathematica <http://www.wolfram.com/mathematica/>
25. Mathcad <http://www.ptc.com/product/mathcad/>
4. Bartko J.J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, Volume 19, Issue, pp. 3-11.
5. Morrow Jr., Jackson A., Dich J., & Mood D. (2006). Measurement and Evaluation in Human Performance. *Champaign, IL: Human Kinetics*, 440 p.
6. Zatsiorskiy V.M. (1982). Sportivnaya metrologiya. Ucheb.dlya in-tov fiz.kult. M.: Fizkultura i sport, 256.
7. Morozov S. M. (1994). Zasobi kontrolyu dlagnostichnih yakostey psihologichnih testiv. K.: ISDO, 68.
8. Gulliksen H. (1950). Theory of mental tests. *New York: J. Wiley*, 486 p.
9. Lord F.M., & Novick R.M. (1986). Statistical theories of mental test scores. *Reading, MA: Addison-Wesley*, 192 p.
10. Crocker L. & Algina, J. (1986). Introduction to classical and modern test theory. *New York: Holt, Rinehart, and Winston*, 218 p.
11. Fisher R.A. (1925). Statistical Methods for Research Workers. *London: Oliver and Boyd*, 276 p.
12. Cronbach L.J., Gleser G.C., Nanda G. & Rajratnam N. (1972). The dependability of behavioral measurements: Theory of generalizability for scores and profiles. *New York: J. Wiley*, 152 p.
13. Shrout P.E. & Fleiss J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, Vol 86(2), 420-428 .
14. Bube H. (1968). Testyi v sportivnoy praktike (per. s nem.). M.: Fizkultura i sport, 239 c.
15. Utkin V.L.(1978). Izmereniya v sporte (vvedenie v sportivnuyu metrologiyu). M.: GTsoLIFK, 200 c.
16. Zatsiorskiy V.M. (1979). Osnovyi sportivnoy metrologii. M.: Fizkultura i sport, 68 c.
17. Hopkins W.G. (2000). Measures of Reliability in Sports Medicine and Science. *Sports Medicine*, 30 (1), 1-15.
18. O'Donoghue P. (2010). Research methods for sports performance analysis. *New York, NY: Routledge*, 278 p.
19. Bland J.M. . Altman D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 8: 307-310.
20. Nevill A.M. & Atkinson G. (1997). Assessing agreement between measurements recorded on a ratio scale in sports medicine and sports science. *British Journal of Sports Medicine*, 31: 314-318.
21. Bruton A, Conway J.H. & Holgate S.T. (2000). Reliability: What is it and how is it measured? *Physiotherapy*, 86, 2, 94-99.
22. Statistica www.statsoft.com
23. SPSS www.ibm.com/software/analytics/spss/
24. Mathematica <http://www.wolfram.com/mathematica/>
25. Mathcad <http://www.ptc.com/product/mathcad/>

МОДЕЛЬ РЕТЕСТОВОЙ НАДЕЖНОСТИ, АЛЬТЕРНАТИВНАЯ МОДЕЛИ ВНУТРИ КЛАССОВОЙ КОРРЕЛЯЦИИ

Заневский И.Ф., Заневская Л.Г.

Львовский государственный университет физической культуры

Реферат. Статья: 9 с., 5 табл., 49 источников.

Целью работы было создание модели коэффициента надёжности теста, применение которой сделало бы возможным в практической работе с тестами получать значения коэффициента надёжности в теоретических пределах от нуля до единицы, как это принято в классической теории надёжности тестов. **Методика исследования** основывалась на положениях классической теории тестов, математической теории надёжности моторных тестов, на модели внутриклассовой корреляции, дисперсионного анализа, с использованием метода Шапиро-Уилка. Вычисления проводились с использованием компьютерных программ Excel и Mathematica. Показано, что модель внутриклассовой корреляции не в полной мере отвечает

положениям классической теории надёжности тестов. Предложенная модель ретестовой надёжности не имеет этого недостатка, а соответствующий коэффициент существует в пределах, определённых положениями классической теории надёжности ($0 \div 1$). В диапазоне практически важных величин предлагаемая модель и модель внутриклассовой корреляции приводят к разным оценкам надёжности теста не только количественно ($\approx 20\%$), но и качественно, то есть с существенным превышением ширины диапазонов оценки надёжности теста по величине соответствующего коэффициента.

Ключевые слова: тест, надёжность, внутриклассовая корреляция.

MODEL OF THE RELIABILITY OF THE TEST, ALTERNATIVE MODELS INSIDE CLASS CORRELATION

I.P. Zanevskyy, L.H. Zanevska

Lviv State University of Physical Culture

Report. Article: 9 p., 7 tables., 47 sources.

The aim was to create a model of the reliability of the test, the application of which would in the practical work with the tests to receive the value of the coefficient of reliability in the theoretical framework from zero to one, as is customary in classical theory of reliability tests. The basic idea of solving this problem was in determining the true variance of the test results, the value of which is equal to the arithmetical average of the results of the repeated measurements. **Methodology of the research** was based on provisions of the classical theory of sports tests, mathematical theory of reliability of motor tests, on the model of intraclass correlation, ANOVA, using the method of Shapiro-Wilk. Calculations were carried out using Excel

and Mathematica computer programs. It was shown that the intraclass correlation model is incomplete with the line of the classical theory of reliability tests. The proposed model of retest reliability does not have this shortage, and the corresponding coefficient is defined in terms of the classical theory of reliability ($0 \div 1$). In the range of almost important quantities of the proposed model and the intraclass model correlations lead to different estimates of the reliability of the test not only quantitatively ($\approx 20\%$), but also qualitatively, i.e., with significant excess width of the ranges of test estimation of reliability coefficient of the corresponding size.

Keywords: test, reliability, intraclass correlation.

Інформація про авторів:

Заневський Ігор Пилипович: anvitvl@ukr.net;
Львівський державний університет фізичної культури,
вул. Костюшко, 11, м. Львів, 79009, Україна.

Заневська Людмила Григорівна: anvitvl@ukr.net;
Львівський державний університет фізичної культури,
вул. Костюшко, 11, м. Львів, 79009, Україна

Цитуйте статтю як: Заневський І.П., Заневська Л.Г. Модель ретестової надійності, альтернативна моделі внутрікласової кореляції/ Заневський І.П., Заневська Л.Г. // Теорія та методика фізичного виховання. — 2014. — № 1. — С. 43—52. DOI: <http://dx.doi.org/10.17309/tmfv.2014.1.1044>

Стаття надійшла до редакції: 15.03.2014 р.