ГОСУДАРСТВЕННЫЙ ЦЕНТРАЛЬНЫЙ
ОРДЕНА ЛЕНИНА ИНСТИТУТ ФИЗИЧЕСКОЙ
КУЛЬТУРЫ

В.М. Зациорский, З.М. Баранова,
Б. А. Суслаков

ЗАДАЧИ ПО СПОРТИВНОЙ МЕТРОЛОГИИ. НАДЕЖНОСТЬ ТЕСТОВ

Методические разработки для институтов
физической культуры

Москва - 1980

ГОСУДАРСТВЕННЫЙ ЦЕНТРАЛЬНЫЙ ОРДЕНА ЛЕНИНА
ИНСТИТУТ ФИЗИЧЕСКОЙ КУЛЬТУРЫ

В.М. Зациорский, З.М. Баранова,
Б.А. Суслаков

**Утверждены
Ученым советом ГЦОЛИФКа**

ЗАДАЧИ ПО СПОРТИВНОЙ МЕТРОЛОГИИ.
НАДЕЖНОСТЬ ТЕСТОВ

**Методические разработки для
институтов физической культуры**

Москва - 1980

Тест (англ. test -проба, испытание, исследование) в психологии и педагогике, стандартизованные задания, результат выполнения которых позволяет измерить психофизиологические и личностные характеристики, а также знания, умения и навыки испытуемого.

БСЭ, т.25, с.513, М., "Советская энциклопедия", 1976.

Введение

Современное научно-методическое обеспечение тренировочного процесса становится бессмысленным, если исключить отсюда методы педагогического контроля. Среди этих методов достойное и важное место отводится теории спортивных тестов.

Данное учебное пособие акцентировано на вопросах оценки основного свойства тестов - надежности и предназначается как вспомогательное по дисциплине "Спортивная метрология" институтов физической культуры. Пособие предполагает знание элементарных основ статистики.

§ 1. Основные понятия теории спортивных тестов

В спорте тестом называется измерение или испытание, проводимое на спортсмене с целью определения его состояния или способностей. Или, другими словами, тест - это кратковременное, технически сравнительно просто обставленное испытание, проводимое в равных для всех испытуемых условиях, и имеющее вид такого задания, решение которого поддается количественному учету и служит показателем степени развития к данному моменту известной функции у испытуемого.

Процесс испытаний называется тестированием, полученное в итоге измерения числовое значение - результатом тестирования (или результатом теста).

Спортивные тесты, в основе которых лежат двигательные задания, называются двигательными (или моторными) тестами. В этих тестах в качестве результатов могут выступать либо двигательные достижения (время прохождения дистанции, количество повторений, проделанное расстояние и т.п.), либо физиологические и биохимические показатели. В зависимости от этого, а также от задания, которое стоит перед испытуемым, различают три соответствующие группы двигательных тестов.

В тех случаях, когда используется не один, а несколько тестов, имеющих единую конечную цель (например, оценку состояния спортсмена в соревновательном периоде тренировки), такая группа называется комплексом или батареей тестов.

Не всякие измерения могут быть использованы как тесты. Специальные требования к тестам следующие:

I) надежность;

2) стандартность - процедура и условия тестирования должны быть одинаковыми во всех случаях применения теста;

3) наличие системы оценок;

4) информативность.

Тесты, удовлетворяющие требованиям надежности и информативности, называются добротными.

§ 2. Надежность спортивных тестов

Надежностью тестов называют степень совпадения результатов при повторном тестировании одних и тех же людей (или других объектов) в одинаковых условиях. В идеале один и тот же тест, примененный к тем же испытуемым в тех же самых условиях, должен давать одинаковые результаты (если состояние испытуемых не изменилось). Однако результаты тестирования меняются от попытки к попытке даже при самой строгой стандартизации испытаний и точной аппаратуре. Четыре основных причины вызывают это изменение (вариацию):

I) изменение состояния испытуемых;

2) неконтролируемые изменения внешних условий и аппаратуры;

3) изменение состояния лица, проводящего или оценивающего тест;

4) несовершенство теста.

Другими словами, надежность - такая характеристика (свойство) теста, которая позволяет судить о том, насколько внушают доверие полученные при тестировании данные, а также насколько обоснованно ожидание исследователя, что при сохранении известного минимума неизменных условий испытуемые в выборке останутся при повторном испытании примерно на тех же порядковых местах.

Надежность теста можно определить четырьмя различными способами.

I. Повторное проведение испытаний тем же тестом через некоторый промежуток времени, длительность которого определяется характером теста. В этом случае вычисляется так называемый коэффициент стабильности - r_{tt} .

2. Проведение испытаний двумя вариантами одного и того же теста, построенного по одному принципу. В этом случае надежность определяется как коэффициент эквивалентности теста - $r_{tt} = r_{AB}$, где А и В эквивалентные формы одного и того же теста. Коэффициент эквивалентности вычисляется в том случае, когда нельзя дважды применить один и тот же тест, например, из-за легкости его запоминания или усвоения.

3. Внутреннее постоянство (стабильноость) теста определяют как коэффициент корреляции $r_{tt} = r_{t_{1/2}\,t_{1/2}}$ между двумя половинками теста, например, между результатами в четных и нечетных попытках (испытуемых).

4. Разновидностью коэффициента эквивалентности является определяемый коэффициент надежноотв между двумя эквивалентными теотами $r_{tt} = r_{t_1 t_2}$.

5. Кроме надежности часто определяют оценку объективноотв (согласованности) теота.

Объективным очитают такой тест, в котором исключено влияние на оценку лица, проводящего тестирование.

В основном надежность вычисляется как мера взаимосвязи между результатами одного и того же теста либо результатами его эквивалентных форм. В статистике мера теоноты овязи определяется значением коэффициента корреляции либо линейного (Браве-Пирсона) r_{xy} , либо рангового (Спирмена) ρ_s . Считается, что надежность тем больше, чем ближе вычисленный коэффициент корреляции (надежности) к единице.

§ 3. Основные формулы для вычисления коэффициентов надежности

I. В том случае, если результаты тестирования представляются непрерывными переменными (т.е. измерениями, которые могут иметь любое значение внутри некоторой области), а также предполагается, что полученный ряд случайных значений (результатов тестирования) имеет нормальный закон распределения, в качестве коэффициента надежности рассчитывают парный линейный коэффициент корреляции Бравэ-Пирсона по одной из нижеприведенных формул.

$$r_{tt} = r_{XY} = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{N \cdot \sigma_X \cdot \sigma_Y} \quad ; \qquad (1)$$

$$r_{tt} = r_{XY} = \frac{N \cdot \sum X \cdot Y - \sum X \cdot \sum Y}{\sqrt{[N \cdot \sum X^2 - (\sum X)^2][N \cdot \sum Y^2 - (\sum Y)^2]}} \quad ; \qquad (2)$$

$$r_{tt} = r_{XY} = \frac{\sum (X-\bar{X})(Y-\bar{Y})}{\sqrt{[\sum (X-\bar{X})^2][\sum (Y-\bar{Y})^2]}} \quad ; \qquad (3),$$

где X - результаты I-го тестирования, У - результаты повторного тестирования, σ_X и σ_Y - средние квадратические отклонения соответственно I-го и 2-го тестирования, N - объем выборки (число измерений при одном тестировании, число испытуемых, принявших участие в одном тестировании и т.п.).

Достоверность (статистическую значимость) коэффициента распределяют по критерию Стьюдента. Для этого рассчитывают значение t расчетное по формуле:

$$t_{расч.} = r_{tt} \sqrt{\frac{N-2}{1-r_{tt}^2}} \quad , \qquad (4),$$

которое сравнивают с критическим значением $t_{(\alpha, N-2)}$ (точка t - распределения Стьюдента, соответствующая уровню значимости α и числу степеней свободы ($N-2$) и определяемая из таблицы приложения I).

Если $t_{крит.} < t_{расч.}$, то коэффициент надежности r_{tt} считается достоверным с вероятностью не более чем ($1-\alpha$).

Для определения надежности, когда результаты тестирования представляются дискретными переменными (измерения могут давать только отдельные значения), используют ранговый коэффициент корреляции Спирмена ρ_S .

$$r_{tt} = \rho_S = 1 - \frac{6 \cdot \sum d^2}{N(N^2-1)} \quad , \qquad (5),$$

где d - разность рангов первого и второго тестирования, N - объем выборки.

Достоверность коэффициента (5) определяют по критерию Стьюдента. Для этого вычисляют значение

$$t_{расч.} = \rho_S \sqrt{\frac{N-2}{1-\rho_S^2}}\,, \qquad (6),$$

которое сравнивают с критическим значением $t_{(\alpha, N-2)}$ (таблица приложения I). Если $t_{крит.} < t_{расч.}$, то коэффициент надежности, рассчитанный по формуле (5), считается достоверным с вероятностью не более чем $(1-\alpha)$.

На практике часто случается, что тест повторяют на новом контингенте испытуемых, причем число испытуемых (число попыток) увеличивают в несколько раз. В таком случае тест изменяет свои свойства и надежность его определяют по формуле Спирмена-Брауна

$$r_T = \frac{m \cdot r_{tt}}{1+(m-1)\,r_{tt}}\,, \qquad (7),$$

где $\tilde{r}_{tt}$ - коэффициент надежности начального теста, - m - число, показывающее, во сколько раз увеличилось число испытуемых при тестировании.

Если требуемая надежность известна, то можно определить, во сколько раз необходимо увеличить число испытуемых (попытки) при тестировании, чтобы добиться этой надежности по формуле:

$$m = \frac{r_T(1-r_{tt})}{r_{tt}(1-r_T)}\,, \qquad (8),$$

где $\tilde{r}_T$ - требуемая (желаемая) надежность, $\tilde{r}_{tt}$ - значение коэффициента надежности данного теста.

Когда надежность определена как коэффициент корреляции $r_{t/2,t/2}$ (между четными и нечетными результатами теста), коэффициент внутреннего постоянства (стабильности) теста рассчитывается по формуле:

$$r_T = \frac{2 \cdot r_{t/2,t/2}}{1+r_{t/2,t/2}}\,. \qquad (9),$$

Также для этого можно воспользоваться формулой Г.Кюдера и М. Ричардсона:

$$\gamma_T = \frac{N}{N-1}\left[1 - \frac{\bar{x}(1-\bar{x}/N)}{\sigma_x^2}\right], \qquad (10),$$

где $\bar{x}$ - среднее арифметическое значение результатов тестирования и σ_x - среднее квадратическое отклонение.

В случае, когда результаты тестирования представляются дихотомическими данными (0,1; да, нет; выполнил, не выполнил), коэффициент надежности (эквивалентности) определяется с помощью коэффициента корреляции φ, предложенного впервые К.Пироном в 1901 г. Результаты испытания группы по двум тестам сводят в следующую таблицу (табл. 1).

Таблица 1

		Тест X: выполн.	Тест X: не выпол.	Итог
Тест	не вып.	a	в	a + в
	вып.	c	d	c + d
Итог		a + c	в + d	N

В этой таблице $N = a + b + c + d$ - общее число участников испытаний, a - число выполнивших испытания по тесту X и не выполнивших по тесту У и т.д. Коэффициент φ рассчитывается по формуле 11

$$\varphi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(b+d)(a+c)}} \qquad (11).$$

Отметим, что $\varphi = +1$ тогда, когда (a + в) = (в + d). Достоверность этого коэффициента проверяется с помощью значения Z

$$Z_{расч.} = \varphi \cdot \sqrt{N}, \qquad (12),$$

которое сравнивается со значениями единичного нормального распределения (приложение 3). Если $Z_{расч.} > Z_{крит.}$ для данной вероятности P, то тест считается надежным (эквивалентным) с вероятностью не более P.

Все определения, рассмотренные выше, предполагают, что тестирование проводится один или два раза. В случае многократных испытаний надежность теста удобнее определять по методике дисперсионного анализа. Смысл такого подхода следующий: предположим, что имеется таблица результатов тестирования (табл.2).

Таблица 2

Испытания / Испытуемые	Результаты			
	I-е	2-е	3-е	К
I	X_{11}	X_{21}	....	X_{K1}
2	X_{12}	X_{22}	....	
.	.	.	....	.
N	X_{1N}	X_{2N}	....	X_{KN}

В этом случае выделяют следующие виды вариации (сумма квадратов отклонений от среднего $S = \sum(x-\bar{x})^2$.

$S_{факт.}$ – факторная или межгрупповая вариация результатов в испытаниях от одного тестирования к другому у одних и тех же испытуемых (между столбцами).

$S_{ост.}$ – остаточная или внутригрупповая вариация, обусловленная изменением результатов тестирования от одного испытуемого к другому (между строчками). Надежность теота в таких испытаниях определяют по следующей формуле:

$$r_{tt} = \frac{S_{факт.}}{S_{общ.}} = \frac{S_{общ} - S_{ост.}}{S_{общ}}, \qquad (13),$$

где $S_{общ.} = S_{факт.} + S_{ост.}$ – есть общая вариация. Так же определяют индекс надежности, как $r_{t\infty} = \sqrt{r_{tt}}$, который в полном комплексе дисперсионного анализа называется внутриклассовым коэффициентом корреляции.

§ 4. Решение практических задач

Пример I. Студенты технического вуза тестировалиcь в беге на 100 м. Испытания проводились два раза с интервалом в одну неделю. Результаты испытаний в стандартных уcловиях получились следующими (табл. 3).

Таблица 3

Испытуемые / Испытания	Результаты						
	I	2	3	4	5	6	7
I-е							
X	I4.3	I4.2	I3.3	I3.4	I4.0	I2.9	I3.5
2-е							
У	I4.5	I4.0	I3.7	I3.2	I4.0	I3.2	I3.3

Требуется определить надежность теста для данного контингента, оценить достоверность и ответить на вопрос - достаточен ли объем выборки.

Р е ш е н и е. Так как результаты тестирования - непрерывные переменные, то для определения надежности рассчитаем парный коэффициент корреляции Бравэ-Пирсона (формула 3).

$$r_{tt} = r_{XY} = \frac{\sum(X-\bar{X})(Y-\bar{Y})}{\sqrt{\left[\sum(X-\bar{X})^2\right]\left[\sum(Y-\bar{Y})^2\right]}} .$$

Результаты расчетов сведем в таблицу (табл. 4).

Таблица 4

Испытуемые	Результаты						
	X	У	(X-X)	(У-У)	$(X-X)^2$	$(У-у)^2$	(X-X)·(У-У)
I	I4.3	I4.5	0.7	0.8	0.49	0,64	0.56
2	I4.2	I4.0	0.6	0.3	0.36	0.09	0.I8
3	I3.3	I3.7	-0.3	0	0.09	0	0
4	I3.4	I3,2	-0.2	-0.5	0.04	0.25	0.I
5	I4.0	I4.0	0.4	0.3	0.I6	0.09	0.I2
6	I2.9	I3,2	-0.7	-0.5	0.49	0.25	0.35
7	I3.5	I3.3	-0.I	-0.4	0.0I	0.I6	0.04
	95.6	95.9			I.64	I.48	I.35
	$\bar{X} = 13.6$	$\bar{Y} = 13.7$					

$$r_{tt} = r_{XY} = \frac{1.35}{\sqrt{1.64 \cdot 1.48}} = \frac{1.35}{\sqrt{2.42}} = 0.865 .$$

Для оценки надежности по формуле (4) рассчитаем

$$t_{расч.} = r_{tt}\sqrt{\frac{N-2}{1-r_{tt}^2}} = 0{,}865\sqrt{\frac{7-2}{1-0{,}865^2}} = 3{,}854.$$

Из таблицы распределения Стьюдента для доверительного уровня $\alpha = 0{,}05$ и числа степеней свободы 7-2 = 5 выпишем значение $t_{крит.}(\alpha=0{,}05;5)$ - 2.571

Так как $t_{расч.} > t_{крит.}$ (3.854 > 2.571), можно заключить, что для данного контингента тест статистически значим с вероятностью не более 95% $(1-\alpha = 1-0{,}05 = 0{,}95)$ и вопрос о выборке отпадает.

Если же мы потребуем более высокой значимости этого теста, т.е. $t_{крит.}$ возьмем для уровня $\alpha = 0{,}01$. Тогда $t_{крит.}(\alpha=0{,}01;5)$ = 4.032 и получаем, что $t_{крит.} > t_{расч.}$ Следовательно, значимости теста с вероятностью 99% уже не проявляется.

Из таблицы нижних доверительных границ коэффициента корреляции (приложение 3) можно определить, что для числа наблюдений N = 7 и $\alpha = 0{,}01$, $r_{XY} = 0{,}874$. Взяв эту оценку как желаемый коэффициент надежности для нашего теста, определим, во сколько раз нужно увеличить выборку, чтобы добиться желаемой надежности. Для этого воспользуемся формулой (8)

$$m = \frac{r_T(1-r_{tt})}{r_{tt}(1-r_T)} = \frac{0.874\ (1-0.865)}{0.865\ (1-0.874)} = \frac{0.118}{0{,}109} = 1.082$$

Таким образом, для получения значимого с вероятностью не более 99% теста число испытуемых необходимо взять равным

7 x 1.082 = 7.574

Округлив до большего целого, получаем, что тестировать необходимо 8 человек из данного контингента испытуемых.

П р и м е р 2. Для оценки фактора силовой выносливости часто используется простой тест - подтягивание на перекладине. Это испытание проводилось на студентах I курса ИФК (специализация - футбол). Было проведено два тестирования с интервалом один месяц.

Получили следующие результаты (табл. 5).

Таблица 5

Испытания \ Испытуемые		Результаты (кол.раз)												
		I	2	3	4	5	6	7	8	9	I0	II	I2	I3
I-е	X	I0	7	I2	7	I7	9	6	I4	I0	II	I4	8	I3
2-е	У	I0	9	I0	8	I4	II	8	I3	I2	I3	I2	8	I5

Требуется определить надежность этого теста для данной специализации, оценить достоверность и определить достаточность объема выборки.

Р е ш е н и е. Учитывая, что результаты тестирования - дискретные переменные, для определения надежности рассчитаем ранговый коэффициент корреляции Спирмена (формула 5)

$$r_{tt} = \rho_s = 1 - \frac{6 \cdot \Sigma d^2}{N(N^2-1)} ;$$

Результаты расчетов представим в табл. 6.

Таблица 6

Тесты	X	Y	d_x	d_Y	$d_x - d_Y$	d^2
I	I0	I0	7.5	8.5	I	I
2	7	9	II.5	I0	I.5	2.25
3	I2	I0	5	8.5	-3.5	I2.25
4	7	8	II.5	I2	-0.5	0.25
5	I7	I5	I	I	0	0
6	9	II	9	7	2	4
7	6	8	I3	I2	I	I
8	I4	I3	2.5	3.5	-I	I
9	I0	I2	7.5	5.5	2	4
I0	II	I3	6	3.5	2.5	6.25
II	I4	I2	2.5	5.5	-3	9
I2	8	8	I0	I2	-2	4
I3	I3	I4	4	2	2	4

$\Sigma d^2 = 49$

$$r_{..} = \rho_S = 1 - \frac{6 \cdot 49}{13(13^2-1)} = 1 - \frac{294}{2184} = 0.865.$$

Достоверность этого коэффициента проверим по критерию Стьюдента, для чего по формуле (6) определим расчетное

$$t_{расч.} = \rho_S \sqrt{\frac{N-2}{1-\rho_S^2}} = 0{,}865 \sqrt{\frac{13-2}{1-0{,}865^2}} =$$

$$= 0{,}865 \cdot 6{,}609 = 5{,}716.$$

Из таблицы распределения Стьюдента для доверительного уровня α = 0.05 и числа степеней свободы $N-2$ = I3-2 =II выписываем

$$t_{крит.}(\alpha = 0{,}05;\ 11) = 2{,}201.$$

Для уровня $\alpha = 0{,}01$, $t_{крит.}(\alpha = 0{,}01;\ 11) = 3{,}106$

Видно, что $t_{расч.} > t_{крит.}$ $(5{,}716 > 3{,}106)$. Следовательно, тест можно считать статистически значимым с вероятностью не более 99% (I-α), и вопрос об объеме выборки отпадает.

Пример 3. При отборе в детской спортивной школе предлагается выполнить два контрольных упражнения или теста.

Результаты тестирования фиксируются так: выполнено, не выполнено. Испытанию было подвергнуто 69 человек. Необходимо проверить предположение об эквивалентности тестов.

Данные испытаний представим в виде таблицы (табл. 7).

Таблица 7

Упражнения			
2-е	I-е		
	да	нет	
нет	I7a	I9 в	36
да	25c	8 d	33
	42	27	69

Надежность или эквивалентность рассчитаем через коэффициент корреляции φ , по формуле (II).

$$\varphi = \frac{ad - b\cdot c}{\sqrt{(a+b)(c+d)(b+d)(a+c)}} = \frac{17\cdot 8 - 25\cdot 19}{\sqrt{36\cdot 33\cdot 27\cdot 42}} =$$

$$= \frac{-339}{1160{,}68} = -0{,}292.$$

Достоверность этого коэффициента проверим по формуле (I2)

$$Z_{расч.} = -0{,}292\sqrt{69} = -2{,}42$$

В единичном нормальном распределении величина $|Z_{расч.}| = 2{,}42$ превышает величину I.96. Следовательно, можно говорить о статистической значимости коэффициента, характеризующего эквивалентность двух тестов с вероятностью не ниже 95%.

Пример 4. Для студентов различных специализаций ИФК были разработаны четыре варианта одной контрольной работы по теоретической дисциплине. Необходимо оценить эквивалентность или равнотрудность этих вариантов и определить надежность.

Случайным методом отобранные данные (оценки) за контрольные работы по всем четырем вариантам в группах различных специализаций представлены в табл. 8.

Таблица 8

Варианты / Испытуемые	I	II	III	IV
	Оценки (в баллах)			
I	4	3	4	5
2	3	5	3	4
3	5	4	5	4
4	4	3	4	3
5	4	3	3	5
6	3	3	4	3
7	3		3	
8			5	
Средние баллы по вариантам	$\bar{X}_I = 3.7$	$\bar{X}_2 = 3.5$	$\bar{X}_3 = 3.9$	$\bar{X}_4 = 4$
Числен. групп (n)	7	6	8	6

Так как в данном случае испытание проводилось четыре раза, оценку надежности произведем с помощью статистического метода – однофакторного дисперсионного анализа. Для этого проведем следующий ряд вычислений:

I. Рассчитаем общий средний балл всего испытания:

$$\bar{X}_0 = \frac{\bar{x}_1 n_1 + \bar{x}_2 n_2 + \cdots + \bar{x}_4 n_4}{n_1 + n_2 + \cdots + n_4} = \frac{3.7\cdot 7 + 3.5\cdot 6 + 3.9\cdot 8 + 4\cdot 6}{7 + 6 + 8 + 6} =$$

$$= \frac{102}{27} = 3.8$$

2. Рассчитаем общую вариацию всех оценок относительно общего среднего балла:

$$S_{общ} = \sum\sum (x - \bar{x}_0)^2 = (4 - 3.8)^2 + (3 - 3.8)^2 + (5 - 3.8)^2 + \ldots + (5 - 3.8)^2 + (3 - 3.8)^2 = 16.68$$

3. Рассчитаем межгрупповую или факторную вариацию:

$$S_{факт} = \sum (\bar{x} - \bar{x}_0)^2 n = (3.7 - 3.8)^2 \cdot 7 + (3.5 - 3.8)^2 \cdot 6 + (3.9 - 3.8)^2 \cdot 8 + (4 - 3.8)^2 \cdot 6 = 0.57$$

4. Учитывая равенство $S_{общ} = S_{факт} - S_{ост.}$, рассчитаем внутригрупповую или остаточную вариацию:

$$S_{ост} = S_{общ} - S_{факт} = 16.68 - 0.57 = 16.11$$

5. Рассчитаем соответствующие дисперсии: общую, факторную, остаточную:

$$\sigma^2_{общ} = \frac{S_{общ}}{n_1 + n_2 + n_3 + n_4 - 1} = \frac{16.68}{27 - 1} = 0.64$$

$$\sigma^2_{факт} = \frac{S_{факт}}{4 - 1} = \frac{0.57}{3} = 0.28$$

$$\sigma^2_{ост} = \frac{S_{ост}}{27 - 4} = \frac{16.11}{23} = 0.70$$

Для оценки надежности воспользуемся формулой:

$$\eta_{tt} = \frac{S_{факт}}{S_{общ}} = \frac{0.57}{16.68} = 0.034$$

Как видно, надежность весьма мала, что вызывает некоторое недоумение. Очевидно, что в этом случае при оценке надежности необходим другой подход.

В том случае, если $\sigma^2_{факт.} < \sigma^2_{ост.}$, говорят, что ошибка выборки меньше ошибки испытания и поступают следующим образом:

I) Проверяют гипотезу $H_0: (\bar{x}_1 = \bar{x}_2 = \bar{x}_3 = \bar{x}_4)$, (статистического равенства групповых средних оценок), для чего вычисляют

$$F_{расч.} = \frac{\sigma^2_{ост.}}{\sigma^2_{факт.}} = \frac{0,7}{0,28} = 3,68$$

$F_{расч.}$ сравнивают с критическим значением распределения Фишера, $F_{крит.}$ для соответствующего доверительного уровня α и числом степеней свободы $K_I=27-4=23$ и $K_2=4-I=3$. Из приложения эти значения равны:

$$F_{крит.}(\alpha=0,05; K=23; K_2=3) = 8,64 \quad (P=1-\alpha=0,95)$$

$$F_{крит.}(\alpha=0,01; K_1=23; K_2=3) = 26,6 \quad (P=1-\alpha=0,99)$$

Следовательно, раз $F_{расч.} < F_{крит.}(\alpha=0,05) < F_{крит.}(\alpha=0,01)$, то нет оснований отклонять гипотезу $H_0: (\bar{x}_1=\bar{x}_2=\bar{x}_3=\bar{x}_4)$;

2) Оценку коэффициента надежности производят по следующей формуле:

$$r_{tt} = 1 - \frac{1}{F_{расч.}} = 1 - \frac{1}{3.68} = 0,728.$$

Тогда теоретически коэффициент корреляции или индекс надежности определяют, как:

$$r_{t\infty} = \sqrt{r_{tt}} = \sqrt{0,728} = 0,853.$$

Таким образом, полученный коэффициент надежности $r_{tt} = 0.728$ говорит о высокой эквивалентности всех вариантов контрольной работы. Достоверность этого коэффициента доказана в п. I о использованием критерия Фишера.

§ 5. Задачи

I. Проведено два испытания с интервалом в I неделю у 7 испытуемых по измерению силы кисти правой руки. Получены следующие результаты (табл. 9):

Таблица 9

Испытания \ Испытуемые	Результаты (в кг)						
	I	2	3	4	5	6	7
I-е	53	50	46	49	47	55	5I
2-е	5I	52	48	46	48	54	5I

Проанализировать результаты тестирования, выбрать метод вычисления коэффициента надежности, рассчитать коэффициент надежности и определить его достоверность.

2. Проведено пять испытаний с интервалом в I неделю у 7 испытуемых по измерению силы кисти правой руки. Получены следующие результаты (табл. IO).

Таблица IO

Испытания \ Испытуемые	Результаты (в кг)						
	I	2	3	4	5	6	7
I-е	5I	47	42	48	46	55	53
2-е	50	48	47	45	43	56	55
3-е	5I	46	48	49	47	54	5I
4-е	48	49	49	48	45	5I	54
5-е	49	50	49	47	46	54	54

Проанализировать результаты тестирования, выбрать метод вычисления коэффициента надежности, рассчитать коэффициент надежности и определить его достоверность.

3. Проведено пять испытаний о интервалом в I неделю по измерению силы кисти правой руки у 7 испытуемых, в каждом испытании проводились 3 попытки. Получены следующие результаты (табл. II):

Таблица II

Испытания и попытки / Испытуемые	Результаты (в кг)														
	I-е			2-е			3-е			4-е			5-е		
	I	2	3	I	2	3	I	2	3	I	2	3	I	2	3
I	46	46	47	47	47	46	48	46	46	49	47	47	47	46	48
2	43	44	45	45	44	44	44	47	44	45	45	44	43	46	46
3	45	43	42	43	43	45	44	45	43	46	45	43	43	46	45
4	5I	50	5I	5I	5I	50	52	50	52	5I	52	5I	52	5I	53
5	54	52	52	52	53	53	54	54	5I	54	5I	55	54	53	52
6	56	55	57	57	55	56	57	57	56	56	57	55	56	56	57
7	49	50	49	49	50	5I	5I	5I	50	52	5I	52	52	50	52

Проанализировать результаты тестирования, выбрать метод вычисления коэффициента надежности, рассчитать коэффициент надежности и определить его достоверность.

4. В группе из 20 студентов проведены испытания на исполнение тройного прыжка с места. Первое испытание проводилось после общей разминки, второе в середине тренировочного занятия после выполнения специальных упражнений и третье в конце тренировочного занятия. Интервалы между занятиями - 30 мин (табл. I2).

Таблица I2

Испытания / Испытуемые	Результаты (в кг)		
	I-е	2-е	3-е
I	8,05	7,80	7,60
2	7,34	7,I0	7,08
3	7,37	7,62	7,I2
4	7,77	7,92	7,34
5	7,04	7,32	6,98
6	7,I7	6,95	6,75
7	6,50	6,6I	6,40
8	8,I5	7,92	7,90
9	6,98	6,54	6,72
I0	6,97	6,75	6,58
II	6,57	6,78	6,37
I2	7,I8	7,32	6,93

Продолжение

Испытания / Испытуемые	Результаты (в кг)		
	I-е	2-е	3-е
I3	7,45	7,5I	6,87
I4	6,98	7,08	6,74
I5	7,I5	7,I2	6,83
I6	8,I9	7,93	7,58
I7	7,I7	7,23	7,0I
I8	7,42	7,I5	7,I2
I9	7,35	7,28	6,98
20	8,24	8,09	7,73

Необходимо определить: в каком испытании этот тест (упражнение) имеет наилучшую стабильность? Какова достоверность этой стабильности? Как ответить на вопрос - влияет ли фактор физической нагрузки в данном случае на результат в тройном прыжке с места?

5. У I0 испытуемых измерили результат прыжка в высоту с места в 5 попытках и еще раз повторили эту серию испытаний через неделю (табл. 13).

Таблица I3

Попытки / Испытуемые	Результаты (в см)									
	I	2	3	4	5	I	2	3	4	5
I	60	6I	60	62	60	6I	60	6I	60	6I
2	55	54	53	54	54	60	62	62	60	6I
3	62	62	63	6I	6I	6I	63	6I	6I	63
4	64	64	62	63	64	66	64	66	65	65
5	63	62	62	6I	63	63	63	62	63	64
6	64	65	66	66	65	65	65	67	67	68
7	65	6I	63	65	64	66	66	67	68	68
8	55	54	53	54	55	56	56	55	54	55
9	56	56	57	58	56	57	57	56	55	59
I0	54	5I	52	53	54	53	54	52	53	55

Проанализировать результаты тестирования, выбрать метод вычисления коэффициента надежности (стабильности), рассчитать коэффициент стабильности и определить его достоверность.

6. При выполнении обязательного упражнения в фигурном катании I4 участникам судьи выставили следующие оценки (табл.I4):

Таблица I4

Участники / Судьи	Оценки в баллах													
	I	2	3	4	5	6	7	8	9	I0	II	I2	I3	I4
А.	5,0	5,I	5,3	5,2	5,5	5,6	5,7	5,8	5,9	6,0	5,4	5,2	5,I	5,5
В.	5,I	5,I	4,9	5,I	5,0	5,3	5,6	5,8	5,9	5,9	5,5	5,3	5,4	5,0

Проанализировать результаты оценок, выбрать метод вычисления коэффициента объективности, рассчитать его и определить достоверность.

7. Надежность теста, оценивающего уровень физической подготовленности спортсменов одной квалификации r_{tt} = 0,45. Во сколько раз нужно увеличить число испытуемых (попыток), чтобы надежность теста стала r_{T} = 0,75?

8. При выполнении обязательного упражнения в фигурном катании 4 спортсменами 8 судей выставили им следующие оценки (табл. I5):

Таблица I5

Судьи / Участники	Оценки в баллах							
	I	2	3	4	5	6	7	8
I	5,3	5,I	5,2	5,2	5,0	5,I	5,3	5,4
2	5,3	5,0	5,0	5,I	4,9	5,0	5,2	5,I
3	5,0	4,9	4,8	4,8	4,9	5,0	5,I	4,8
4	5,2	5,I	5,2	5,3	5,0	5,I	5,2	4,9

Проанализировать результаты оценок, выбрать метод вычисления коэффициента объективности для данной триады судей, рассчитать коэффициент объективности и оценить его достоверность.

9. В контрольных упражнениях по тяжелой атлетике "рывок" и "толчок" фиксировались результаты: по три попытки за одно тренировочное занятие в течение двух дней у 8 спортсменов I разряда (табл. I6).

Таблица I6

Попытки, упражнения / Участники	Результаты (в кг)											
	"Рывок"						"Толчок"					
	I-й день			2-й день			I-й день			2-й день		
	I	2	3	I	2	3	I	2	3	I	2	3
I	I37,5	I36	I37	I30	I32	I30	I72,5	I70	I72	I68	I70	I72
2	I00	98	I02	96	94	96	I30	I30	I3I	I29	I30	I30
3	I45	I42	I42	I46	I44	I45	I70	I70	I68	I68	I69	I70
4	I07,5	I05	I07	I04	I05	I03	I32,5	I32	I32,5	I30	I33	I32
5	I50	I48	I46,5	I48	I47	I47	I75	I74	I74	I73	I74	I85
6	II0	I08	I08	II2	II0	I08	I40	I40	I4I	I40	I38	I39
7	I20	II8	II6	II8	II4	I20	I45	I43	I44	I43	I43	I44
8	92,5	94	92	94	92	92	II5	II5	II2	II2	II2	II4

Проанализировать результаты контрольных упражнений, выбрать метод вычисления коэффициента надежности, рассчитать их и оценить достоверность. Сравнить надежность двух упражнений.

I0. После того, как в тесте, проверяющем быстроту простой реакции, число заданий увеличили в 5 раз, надежность теста стала $r_T = 0{,}83$. Определить надежность первоначального теста (r_{tt})?

II. Для оценки техники толкания ядра (и для ее совершенствования) используют ядра различного веса. В группе из I0 человек фиксировались результаты толкания ядра (о места) весом 5 кГ и I0 кГ в течение 4-х дней с интервалом в I день. Результаты(в м)получились следующими (табл. I7):

Таблица I7

Вес ядра, дни / Испытуемые	Результаты (в кг)							
	5				I0			
	I	2	3	4	I	2	3	4
I	9,I2	I0,82	9,30	II,05	5,82	5,6I	6,II	5,70
2	9,88	8,92	I0,00	9,70	6,63	6,I7	5,92	5,65
3	I0,I0	I0,I5	8,90	9,40	6,37	6,89	6,34	6,32
4	8,65	9,I5	8,70	8,90	5,93	6,I5	6,II	5,85
5	9,38	9,86	9,65	9,80	5,79	6,49	5,92	5,94
6	8,98	9,98	I0,I2	I0,I0	5,7I	5,I8	5,30	5,72
7	8,I9	8,90	8,70	7,90	5,6I	6,I7	5,84	5,95
8	I0,23	8,25	9,35	9,60	6,90	6,92	6,70	6,3I
9	I0,25	9,34	9,40	8,90	6,80	6,I5	6,25	6,78
I0	II,05	9,37	9,65	9,92	9,25	7,20	7,I5	7,24

Проанализировать результаты, выбрать метод вычисления коэффициента надежности, рассчитать их и оценить достоверность. Какое из двух контрольных упражнений (тестов) н дежнее?

I2. Надежность теста, проверяющего зрительную память испытуемых, равна $r_{tt} = 0,54$. Число задании в тесте увеличили в 7 раз. Какова стала надежность этого теста (r_T)?

I3. При поступлении в ИФК на экзамене по легкой атлетике в упражнении - толкании ядра-70 абитуриентов показали следующий средний спортивный результат:

$\bar{x}_I = 8,42$ м и $\bar{x}_I = 0,70$ м

Через год учебы, те же студенты (не специализировавшиеся в толкании ядра) показали следующий средний результат:

$\bar{X}_2 = 8,78$ м и $\bar{x}_2 = 0,66$ м

Существует ли значимая статистическая разность между I-м и 2-м измерениями и можно ли результат в толкании ядра использовать в качестве теста? Уровень значимости взять равным 0,05.

I4. Два теста примерно одинаковой трудности и сложности были предложены 64 студентам ИФК для оценок технической подготовленности. Результаты тестирования следующие (табл. I8):

Таблица I8

Тест T_I	Тест T_2		
	не выполнено	выполнено	всего
Выполнено	I0	26	36
Не выполнено	22	6	28
Всего	32	32	64

Проанализировать полученные результаты тестирования, выбрать метод определения эквивалентности этих тестов, рассчитать коэффициент надежности и оценить его значимость.

I5. Измерялся результат о прыжка в высоту с места в стандартных условиях у I0 испытуемых. Первые три попытки выполнены в течение одного дня с интервалом I5 мин, остальные две попытки выполнялись в течение недели с интервалом два дня. Результаты тестирования следующие (табл. I9):

Таблица I9

Испытуемые / Испытания	I	2	3	4	5	6	7	8	9	I0	
	Результаты (в см)										
I-е	60	55	62	64	63	64	65	55	56	54	(X_I)
2-е	6I	54	62	64	62	65	6I	54	56	5I	(X_2)
3-е	60	53	63	62	62	66	63	53	57	52	(X_3)
4-е	6I	60	6I	66	63	65	66	56	57	53	($У_I$)
5-е	60	62	63	64	63	65	66	56	57	54	($У_2$)

Проанализировать полученные результаты тестирования, выбрать соответствующий математический метод определения надежности данного теста, рассчитать коэффициенты надежности, сравнить и оценить их статистическую значимость.

I6. Измерялась сила удара по мячу у II испытуемых специализации волейбол в двух попытках и стандартных условиях. Результаты тестирования следующие (табл. 20):

Таблица 20

Испытуемые / Испытания	I	2	3	4	5	6	7	8	9	I0	II	
	Результаты (в кг)											
I-е	59	60	58	55	60	62	60	49	6I	53	49	(X)
2-е	6I	65	59	60	58	64	6I	47	6I	54	48	(У)

Проанализировать полученные результаты тестирования, выбрать соответствующий математический метод определения надежности данного теста, рассчитать коэффициент надежности и оценить его значимость.

I7. Два теста предположительно одинаковой трудности и сложности были предложены 64 студентам ИФК для оценки физической подготовленности. Результаты тестирования следующие (табл. 2I):

Таблица 2I

$Тест_I$	$Тест_2$: не выполнено	выполнено	всего
Выполнено	8	24	32
Не выполнено	22	I0	32
Всего	30	34	64

Проанализировать полученные результаты тестирования, выбрать соответствующий математический метод для определения эквивалентности этих тестов и оценить значимость коэффициента эквивалентности.

18. Группа студентов ИФК проходила испытания по тесту - гладкий бег на дистанции 30 м с низкого старта. Испытания проводились 4 раза с интервалом в два дня. Результаты тестирования 8 испытуемых следующие (табл. 22):

Таблица 22

Испытуемые / Испытания	Результаты (в с)								
	I	2	3	4	5	6	7	8	
I-е	4,5	4,5	4,6	4,6	4,5	4,7	4,6	4,7	(X_I)
2-е	4,5	4,6	4,6	4,5	4,6	4,6	4,6	4,6	(X_2)
3-е	4,6	4,5	4,7	4,5	4,7	4,5	4,5	4,6	(X_3)
4-е	4,7	4,6	4,6	4,7	4,6	4,6	4,6	4,6	(X_4)

Проанализировать полученные результаты тестирования, выбрать соответствующий математический метод определения надежности данного теста, рассчитать коэффициент надежности и оценить его значимость.

19. В группе из 2I студента проведены испытания на поднятие тяжести (гири весом I0 кг). I-е испытание проводилось после легкой разминки, 2-е после выполнения комплекса специальных упражнений, 3-е после электростимуляции. Каждое испытание проводилось через неделю, в строго стандартных условиях (табл.23).

Таблица 23

Испытания / Испытуемые	Результаты (кол.раз)		
	I-е	2-е	3-е
I	49	50	47
2	38	42	43
3	45	47	50
4	3I	33	36
5	47	48	49
6	36	42	43
7	3I	40	42
8	34	38	38
9	36	33	38

Продолжение

Испытания / Испытуемые	Результаты (кол.раз)		
	I-е	2-е	3-е
I0	22	23	24
II	4I	43	50
I2	29	25	30
I3	50	53	57
I4	5I	49	54
I6	43	48	48
I6	40	4I	42
I7	3I	23	29
I8	60	68	75
I9	4I	39	39
20	38	32	33
2I	43	43	43

Необходимо определить: в каком испытании этот тест (упражнение) имеет наибольшую стабильность? Проверить достоверность этой стабильности. Как ответить на вопрос - влияет ли фактор стимуляции на результат в поднятии тяжести?

20. В стрельбе на круглом стенде необходимо оценить надежность показателя теста "вскидка". В группе из I2 мастеров спорта проводились 34 испытания с интервалом в два дня. Данные следующие (табл. 24):

Таблица 24

Испытуемые / Испытания	Результаты (в с)											
	I	2	3	4	5	6	7	8	9	I0	II	I2
I-е	0,34	0,30	0,28	0,28	0,30	0,26	0,28	0,28	0,28	0,20	0,I8	0,24
2-е	0,20	0,22	0,20	0,I8	0,26	0,28	0,26	0,22	0,22	0,20	0,22	0,32
3-е	0,28	0,26	0,28	0,22	0,28	0,28	0,22	0,26	0,30	0,24	0,26	0,24

Проанализировать результаты тестирования, выбрать метод вычисления коэффициента надежности, рассчитать его и оценить достоверность.

2I. У I0 детей (8-9 лет), имеющих Ш юношеский спортивный разряд по фигурному катанию, необходимо оценить надежность теста "максимальное число оборотов вправо", выполняемого в трех попытках с интервалом в I неделю. Результаты тестирования представлены в табл. 25.

Таблица 25

Испытуемые / Испытания	Результаты (в град.)									
	I	2	3	4	5	6	7	8	9	I0
I-е	540	560	500	450	460	370	450	495	520	500
2-е	490	520	5I0	490	470	390	420	5I0	490	5I0
3-е	530	540	5I0	470	470	380	470	420	480	530

Проанализировать результаты тестирования, выбрать метод вычисления коэффициента надежности, рассчитать его и оценить достоверность.

22. В тренировке тяжелоатлетов используется показатель - тест "рывок предполагаемый". Испытания этого теста проведены в группе мастеров спорта (I3 человек) три раза с интервалом в I неделю. Результаты даны в табл. 26.

Таблица 26

Испытуемые / Испытания	Результаты (в кг)												
	I	2	3	4	5	6	7	8	9	I0	II	I2	I3
I-е	I35	95	I42,5	I05	I47,5	I07,5	I47,5	I20	92	95	I30	92,5	I05
2-е	I35,5	75	I44	I05	I47	I08	I47	I20,5	93	94	I30	93	I02
3-е	I34,5	95	I44	I04	I47,5	I08	I48	I20	92,5	94	I30,5	92	I03

Проанализировать результаты тестирования, выбрать метод вычисления коэффициента надежности, рассчитать его и оценить достоверность.

Вопросы для самопроверки

I. Основной смысл определения надежности теста.

2. Достаточно ли надежны оценки на вступительных экзаменах в институт?

3. Основные разновидности надежности.

4. Основные методы определения надежности.

5. Каким образом можно определить число попыток или испытуемых для получения достоверного коэффициента надежности?

6. Может ли коэффициент надежности быть недостоверным?

7. Как определить согласованность оценок судей на соревнованиях по фигурному катанию?

8. Придумайте схему педагогического эксперимента, где можно было бы проверить надежность теста и влияние на нее следующих факторов:

а) колебаний в состоянии испытуемых в разные дня недели;

б) тестирования в утренние и вечерние часы;

в) присутствие посторонних наблюдателей.

9. Что физически обозначает "индекс надежности"?

10. Какая связь между управлением тренировочным процессом и теорией тестов?

Заключение

При исследовании возможностей научного подхода к объективизации педагогического контроля в спорте необходимо учитывать такие стороны, как выбор наиболее эффективных форм и методов, оперативность тактических приемов, диагностика и оценка навыков и умений, статистический анализ результатов контроля, выводы и принятие решений в управлении тренировочным процессом.

В настоящее время ни у кого не вызывает сомнений то, что возрастающие требования к научной обоснованности и достоверности данных диктуют необходимость замены общих нечетких определений качества учебно-тренировочного процесса однозначными количественными характеристиками. Важную роль призвана сыграть в этом теория спортивных тестов, точнее использование математического аппарата в исследовании качества управления тренировочным процессом.

Литература

Бернштейн М.С. К методике составления и проверки тестов. - "Вопросы психологии", 1968, № 1.

Бондаревский Е.Я. Надежность тестов, используемых для характеристики моторики человека.-"Теория и практика физической культуры", 1970, № 5.

Зациорский В.М. Основы спортивной метрологии. М., ФиС, 1979.

Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. М., "Прогресс", 1976.

Стивенс С. Экспериментальная психология. ч. II, М., 1963.

Суппенс П., Зиннесс Дж. Основы теории измерений. - В кн.: Психологические измерения, М., 1967.

Содержание

ЗАДАЧИ ПО СПОРТИВНОЙ МЕТРОЛОГИИ.
НАДЕЖНОСТЬ ТЕСТОВ

Методические разработки для
институтов физической культуры

Редактор И.Дубнова. Корректор А.Домбровская.

Объем I 48 уч.-изд.л. Тираж 1000 экз. Зак. 973/1345

Издание Редакционно-издательского отдела ГЦОЛИФКа.

Типография В/о "Союзспортобеспечение",
Москва, Мичуринский проспект, 40.