

Лекция – 7 (2ч)

Тема: Определение информативности и добротности теста Коэффициент надежности теста

План:

1. Информативность теста
2. Коэффициентом информативности
3. Добротность теста

1. Информативность теста — это степень точности, с какой он измеряет свойство (качество, способность, характеристику и т.п.), для оценки которого используется. Информативность нередко называют также валидностью (от англ. validity — обоснованность, действительность, законность). В разных случаях одни и те же тесты могут иметь разную информативность.

Вопрос об информативности теста распадается на два частных вопроса:

- 1) что измеряет данный тест?
- 2) как точно он измеряет?

Если тест используется для определения состояния спортсмена в момент обследования, то говорят о *диагностической* информативности теста. Если же на основе результатов тестирования хотят сделать вывод о возможных будущих показателях спортсмена — о *прогностической* информативности. Тест может быть диагностически информативен, а прогностически нет, и наоборот.

Степень информативности может характеризоваться количественно на основе опытных данных (так *называемая эмпирическая* информативность) и качественно — на основе содержательного анализа ситуации (*содержательная* или логическая информативность).

2. Коэффициентом информативности

Идея определения эмпирической информативности (греч. эмпирия — опыт) состоит в том, что результаты сравнивают с некоторым критерием. Для этого рассчитывают коэффициент корреляции между критерием и тестом (такой коэффициент называют коэффициентом информативности и обозначают r_{tk} , где: t — первая буква в слове “тест”, k — в слове “критерий”).

В качестве теста-критерия берется показатель, заведомо и бесспорно отражающий то качество системы, которое собираются измерять с помощью теста.

Чаще всего в спортивной метрологии критериями служат:

- 1) спортивный результат;
- 2) какая-либо количественная характеристика соревновательной деятельности (например, длина шага в беге, сила отталкивания в прыжках, процент точных передач в футболе и т.д.);

- 3) результаты другого теста, информативность которого доказана;
- 4) принадлежность к определенной группе. Например, можно сравнивать мастеров спорта и спортсменов низших разрядов; принадлежность к одной из этих групп является критерием;
- 5) так называемый составной критерий, например, сумма очков в многоборье.

Пример определения информативности одного и того же теста (скорость бега 30 м с ходу у мужчин) при разных критериях приведен в таблице 8. Данные получены на 62 спортсменах, показавших в прыжках в длину результаты от 600 до 772 см; результаты в многоборье брались на основании опроса.

Таблица 8

Информативность теста “бег 30 м с ходу” (n = 62)

Критерий	Мера критерия	Коэффициент информативности
Прыжок в длину с разбега	Результат прыжка (см)	0,658
Разбег в прыжках в длину	Скорость бега на последних 10 м (м/с)	0,918
Спортивные достижения в прыжках в длину	Разряд по легкой атлетике (от второго до мастера спорта)	0,715
Результат в троеборье: бег 100 м, прыжки в длину, бег 100 м с/б	Сумма очков	0,764

Вопрос о выборе критерия является, по существу, самым важным при определении реального значения и информативности тестов.

При практическом использовании показателей эмпирической информативности следует иметь ввиду, что они справедливы лишь по отношению к тем исследуемым и условиям, для которых они рассчитаны. Тест, информативный в группе начинающих, может оказаться совершенно не информативным в группе мастеров спорта.

Информативность теста неодинакова в разных по составу группах. В частности, в группах, более однородных по своему составу, тест обычно менее информативен.

Коэффициент информативности очень сильно зависит от надежности теста и критерия. Тест с низкой надежностью всегда мало информативен, поэтому не имеет смысла проверять малонадежные тесты на информативность. Недостаточная надежность критерия также приводит к снижению коэффициентов информативности. Однако в данном случае было бы не правильно пренебрегать тестом как мало информативным; ведь

верхней границей возможной корреляции теста является не ± 1 , а его индекс надежности. Поэтому надо сравнивать коэффициент информативности с этим индексом. Действительную информативность (с поправкой на надежность критерия) рассчитывают по формуле:

$$\hat{r}_{tk} = \frac{r_{tk}}{\sqrt{r_{kk}}},$$

где:

r_{tk} — коэффициент информативности,
 r_{kk} — надежность критерия.

Коэффициент информативности может определяться как коэффициент корреляции между результатом теста и теста-критерия (например, ранговый коэффициент корреляции):

$$r_{x,y}^S = 1 - \frac{6 \cdot \sum (dx_i - dy_i)^2}{n \cdot (n^2 - 1)}$$

Пример:

Эмпирическим путем определить информативность теста становой динамометрии (**X**) в оценке силовых возможностей исследуемых, используя в качестве теста-критерия (**Y**) максимальное количество подтягиваний на перекладине. При этом данные выборки таковы:

x_i , кг ~ 150; 140; 135; 100; 110; 115; 140.
 y_i , кол-во раз ~ 9; 6; 6; 3; 3; 4; 8.

Решение:

1. Занесем результаты тестирования в рабочую таблицу и сделаем все необходимые расчеты:

x_i	dx	y_i	dy	$dx - dy$	$(dx - dy)^2$
150	1	9	1	0	0
140	2,5	6	3,5	-1	1
135	4	6	3,5	0,5	0,25
100	7	3	6,5	0,5	0,25
110	6	3	6,5	-0,5	0,25
115	5	4	5	0	0
140	2,5	8	2	0,5	0,25
				$\sum(dx - dy)=0$	$\sum(dx - dy)^2=2$

2. Подставим полученные данные в формулу расчета рангового коэффициента корреляции:

$$r_{x,y}^S = 1 - \frac{6 \cdot \sum (dx_i - dy_i)^2}{n \cdot (n^2 - 1)}$$

$$r_{x,y}^S = 1 - \frac{6 \cdot 2}{7 \cdot (49 - 1)} = 1 - \frac{12}{336} \approx 1 - 0,03 \approx 0,97.$$

3. Определим число степеней свободы по формуле:

$$k = n.$$

Тогда при $k = 7$ и $\beta = 99\%$ $r_{\text{табл.}} = 0,94$.

Вывод: т.к. $r_{x,y}^S = 0,97 > r_{\text{табл.}} = 0,94$ при $k = 7$, то с уверенностью в 99% можно говорить о том, что в данной группе исследуемых тест становой динамометрии информативен при оценке уровня развития силовых возможностей.

Рассматривая данный частный пример, необходимо отметить, что с информативностью и надежностью теста тесно связано понятие *о его различительной возможности* — том минимальном различии между исследуемыми, которое диагностируется с помощью теста. Различительная возможность теста зависит от:

1. Межиндивидуальной вариации результатов, которая во многих случаях (например, межклассовая вариация) может быть повышена за счет увеличения трудности теста.
2. Надежности (т.е. соотношения меж- и внутрииндивидуальной вариации) теста и критерия.

Нет фиксированной величины информативности теста, после которой можно считать тест пригодным. Здесь многое зависит от конкретной ситуации: желаемой точности прогноза, необходимости получить хотя бы какие-то дополнительные сведения о спортсмене и т.п. Практически для диагностики используются тесты, информативность которых не меньше 0,3. Для прогноза, как правило, нужна более высокая информативность — не менее 0,6.

Информативность батареи тестов, естественно выше, чем информативность одного теста. Нередко бывает так, что информативность одного отдельно взятого теста слишком низка, чтобы им пользоваться. Информативность же батареи тестов, в которую он входит, может быть вполне достаточна.

3. Добротность теста

В практической деятельности принято выделять такое понятие как **добротность** теста, которая складывается из его информативности и надежности, определенных с помощью корреляционного анализа (рис.15).

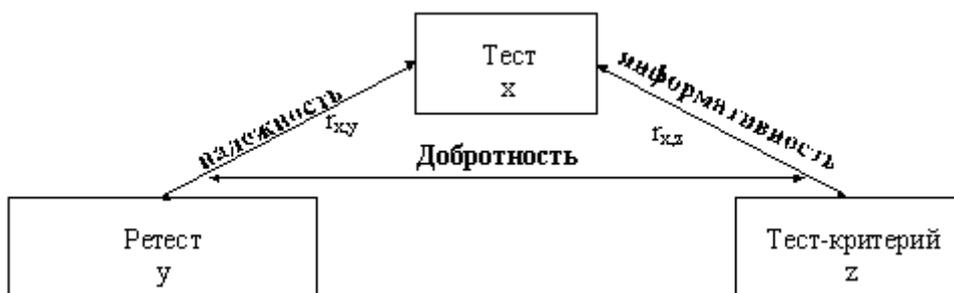


Рис. 15. Схема определения добротности тестов.

При определении добротности теста необходимо учесть следующие положения:

1. Если тест информативен ($\beta = 99\%$) и надёжен ($\beta = 95\%$), то он считается добротным, при этом берется меньшая величина уверенности ($\beta = 95\%$)
2. Если же тест не отвечает хотя бы одному из этих свойств (т.е. ненадежен, или неинформативен), то он считается недобротным.

Пример:

Для оценки уровня развития скоростных возможностей группы исследуемых был использован показатель времени опоры (м/с) в 10-секундном беге на месте на контактной платформе с высоким подниманием бедра (теппинг-тест). Необходимо доказать добротность данного теста (X), сравнив его результаты с результатами ретеста (Y) и теста-критерия (Z), которым служит время бега на 100 м, если данные выборки таковы:

x_i , мс ~ 157; 126; 156; 126; 138; 125; 121; 130.
 y_i , мс ~ 154; 129; 160; 128; 132; 121; 124; 132.
 z_i , мс ~ 15,2; 14,1; 15,6; 14,3; 14,9; 14,2; 14,4; 15,0.

Решение:

1. Занесем результаты тестирования в рабочую таблицу и выполним необходимые расчеты:

x_i	dx	y_i	dy	$dx-dy$	$(dx-dy)^2$	z_i	dz	$dx-dz$	$(dx-dz)^2$
157	8	154	7	1	1	15,2	7	1	1
126	3,5	129	4	-0,5	0,25	14,1	1	2,5	6,25
156	7	160	8	-1	1	15,6	8	-1	1
126	3,5	128	3	0,5	0,25	14,3	3	0,5	0,25
138	6	132	5,5	0,5	0,25	14,9	5	1	1
125	2	121	1	1	1	14,2	2	0	0
121	1	124	2	-1	1	14,4	4	-3	9
130	5	132	5,5	-0,5	0,25	15,0	6	-1	1
				$\Sigma=0$	$\Sigma=5$			$\Sigma=0$	$\Sigma=19,5$

2. Определим надежность данного теста с помощью расчета рангового коэффициента корреляции между результатами теста и ретеста:

$$r_{x,y}^S = 1 - \frac{6 \cdot 5}{8 \cdot (64 - 1)} = 1 - \frac{30}{504} \approx 1 - 0,06 \approx 0,94$$

Тогда при $k = 8$ и $b = 99\%$ табличное значение рангового коэффициента корреляции равно 0,88.

3. Определим информативность теппинг-теста с помощью расчета рангового коэффициента корреляции между данными теста и теста-критерия:

$$r_{x,z}^S = 1 - \frac{6 \cdot 19,5}{8 \cdot (64 - 1)} = 1 - \frac{117}{504} \approx 1 - 0,23 \approx 0,77$$

Значение рангового коэффициента корреляции из таблицы при $k = 8$ и $(\beta = 95\%)$ равно 0,72.

Вывод: 1) т.к. $r_{x,y}^S = 0,94 > r_{\text{табл.}} = 0,88$, можно говорить о том, что в данной группе исследуемых теппинг-тест надежен в оценке скоростных возможностей;

2) т.к. $r_{x,z}^S = 0,77 > r_{\text{табл.}} = 0,72$, то с уверенностью $(\beta = 95\%)$ можно говорить об информативности данного теста;

3) т.к. тест надежен и информативен, то с уверенностью $(\beta = 95\%)$ можно говорить о том, что он добротен.