

Глава четвърта

ИЗРАБОТВАНЕ НА ФИШ/КАРТА/ ВЪПРОСНИК ЗА ПРОУЧВАНЕ

I. Признаци/характеристики; поделения/разновидности

След като сме събрали и прегледали литературата по научен проблем (тема), който ни интересува и вече сме се ориентирали в по-важните факти, теории, мнения и противоречия. Знаем, Кой, Къде, Кога и Как е проучвал проблема, може да пристъпим към изработването на картата/въпросник за регистриране на данните в проучването, което ни предстои, т.е. трябва да опишем характеристиките/признаците, който ни интересуват, за да постигнем целта на проучването.

Съвкупността от признаци, които подлежат на наблюдение и регистриране, образува сърцевината на *програмата* на изследването. Например, ако провеждаме проучване, за да се изработят нормативи за физическо развитие на новородените деца тези характеристики/признаци ще бъдат: местораждане; пол, ръст, телесна маса, обиколка на главата, гестационна седмица на раждането на детето; възраст и образование на майката и т.н. Тези характеристики (разбира се и много други) трябва да опишем в специално изработен за проучването документ, който се нарича **карта** или **въпросник за проучване** на физическото развитие на новородените деца в гр. X (ако имаме по-големи амбиции – в Р България).

Въпросът за признаците е неразривно свързан с въпроса за *поделенията (разновидностите)* им. За да стане по-ясно какво разбираме под признак/характеристика и поделение на признака,

трябва да изтъкнем, че в картата/въпросника признаците получават формата на въпроси, а поделенията на признаците са възможните разновидности на отговорите.

Определянето на поделенията на признаците означава да се направи типология на възможните проявления на разновидностите им. Например признакът пол има две възможни разновидности, проявления – мъж, жена. Често пъти типологията се оказва не съвсем лека работа поради това, че изследователите нямат достатъчно опит и знания или пък защото възможните проявления са многообразни. Тук се проявява умението на изследователя, съобразно знанията му по проблема и неговата научната интуиция, да определи тези признаци, които са най-съществени и главни за конкретното проучване.

Винаги е възможно първоначално определените поделения да са недостатъчни. Задълбоченото разглеждане на получените резултати от пилотното проучване почти винаги налага допълнително изясняване в една или друга насока на някои признаци. Трябва да се внимава, тъй като увлечението да се определят много поделения може да се окаже вредно в смисъл, че прекалено раздробените данни са пречка за разкриване на закономерности, които се проявяват само при изследването на голям брой случаи.

Критерият на основата, на който ще се реши въпросът за поделенията (разновидностите) на признаците, е критерий, свързан със същината на разглеждания проучван проблем. Оказват влияние също така някои практически съображения, като например бюджетните възможности на изследователя.

II. Видове признаци/характеристики на изследваните лица/случай и скали за измерването им

В областта на медицината и здравеопазването проучванията са репрезентативни (представителни) и изучаваните събития (явления) имат вероятностен характер, поради което **характеристи-**

ките/признаците, чрез които се представят **се наричат случайни променливи величини**, а за краткост **променливи величини** или само **променливи**.

1. Качествени (описателни) и количествени променливи величини. Скали за измерване

Качествени променливи са тези, за които няма мерна единица, т.е. разновидностите им се описват с думи (термини), между които няма нищо общо. В даденият по-горе пример качествени променливи, които ще се регистрират по време на проучването са местораждане с разновидности например Плевен, Русе, София, Варна и т.н.; пол на новороденото дете с разновидности – момче и момиче. **За да бъдат измерени**, т.е., за да се представят количествено е необходимо **само брое**не и така се определя честотата им (повторяемостта им).

Тази форма на количествено изразяване се нарича номинално скалиране – **номинална скала на измерване**. Това е най-примитивната форма на количествено изразяване. За да се брои трябва само да се определят условията (критериите), при които лицата се разглеждат като еднакви, т.е. да се определи качествена им еднородност.

Когато разновидностите на качествена променлива са само две, както е полът или ако са повече от две, но се привеждат с алтернативата „да или не“, „има или няма“ до две, тогава номиналната скала се нарича **биноминална или алтернативна скала на измерване, респективно променлива**.

Друг вид скала за измерване на качествени променливи е **ординалната или рангова скала** –когато между разновидностите на качествена променлива съществува степенуване (с термините **по-малко, средно, по-голямо** или други подобни) или ранжиране по големина, т.е. разновидностите също се описват с думи (термини), но очевидно се различават по отношение степента на проявяване. В горния пример такава променлива е образованието на майката, на което разновидностите са подредени – без образование,

начално, основно, не завършено средно, завършено средно, специалист по..., бакалавър, магистър, доктор. Вижда се, че има разлика между отделните разновидности, но тази разлика не е еднаква между, които и да са последователни двойки. Например, разликата между начално и основно не е същата както между бакалавър и магистър.

Променливи, които имат метрична единица за представяне (г, кг, см и др.) са количествени. В горния пример за физическото развитие на новородените такива са телесна маса (разновидностите се описват с измереното тегло в грамове), ръст (разновидностите се описват с измереният ръст в см.) и т.н. Когато измерването е с целочислена мерна единица променливата се нарича **прекъсната (дискретна)**, а когато се измерва и с подразделенията на мерната единица (с дробно число) променливата е **непрекъсната (недискретна или продължителна)**.

Скалата за измерване на количествените променливи може да бъде **интервална скала** (когато липсва истинска нула, например температурата – нула градуса означава определена степен на студено, а не липса на променливата температура) и **пропорционална скала** (когато има истинска нула, т.е. при нула самата променлива липсва). Тази скала дава възможност да се сравняват съотношения и да се правят изводи като например, два пъти по-голям или др.

Погрешно е измеримите променливи да се превръщат в неизмерими. Подобно превръщане може да служи само за разграничаване и подробно изучаване на определена група като например, лицата със стойности на измеримата променлива над средното ниво плюс или минус две стандартни отклонения.

2. Зависими и независими променливи

При изучаване на връзки или зависимости променливите се делят на **зависими (резултативни) и не зависими (фактори, причини)**. Отговор на въпроса коя променлива е зависима и коя независима трябва да се търси в логиката, здравият разум и теорията на

проучваната област, а не в статистическите методи. Трудностите при определянето им идват от факта, че не винаги е възможно да се определи времевата връзка между тях и едновременното им съществуване, поради което често се налага изключване влиянието на фактори свързани с провеждането на проучването.

Забележка: За подробности виж Социална медицина и Медицинска статистика.

III. Изработване на карта/въпросник за научно проучване

Работата по програмата на изследването завършва със съставяне на картата/въпросник. Там се отразява цялата съвкупност от признаци, за които изследователят е установил, че се нуждае от информация. Въпросникът в своя първоначален вид в същност представлява в по-голяма или по-малка степен един списък на необходимата информация по проучвания проблем.

Правилото, каквато е структурата на литературният обзор, такава да е и структурата на резултатите от проучването важи в пълна сила и за структурата и съдържанието на картата/въпросника за регистриране на данните (информацията) от проучването.

В зависимост от целите и хипотезите на проучването трябва да се **определят областите**, които ще се изучават **и променливите величини (характеристиките, признаците)**, които ще се включат.

Всички разновидности на качествените променливи в картата за проучване се описват и номерират с арабски числа, но често се налага да се остави един номер за разновидност, която не е била предвидена първоначално. На количествените променливи се остава само място за записване на измерената стойност.

В картата за проучването областите се номерират с римски числа, а променливите – с арабски числа (независимо, че и разновидности на качествените променливи също се номерират с арабски числа). Номерирането на променливите е в непрекъснат последователен ред, т.е независимо, че се преминава в нова област

първата променлива няма да има номер едно, а ще има следващият номер от предходната област. Така лесно се възприема и помни съдържанието както на областите, така и на цялата карта за проучване, което значително улеснява и работата по обработката на информацията от проучването.

Не е задължително, но е желателно **да се изработят подробни инструкции за начина на попълване на данните** и специфични особености за някои променливи, например, може ли да има повече от една разновидност на качествените променливи. Ако за някои променливи има възможност, а за други – няма, добре е да се отбележи като указание към съответният номер на променливата, за която това е възможно.

Картата за проучването трябва да се обсъди с по-широк кръг специалисти или **експерти**, за да се установи дали е пълна (обхваща цялостно съдържанието на проучвания проблем), дали въпросите са ясни и терминологията адекватна.

Разработената карта за проучване трябва да се тества чрез провеждане на **пилотно проучване** върху малък брой лица/случаи. Така ще се определи дали картата е разработена подходящо за събирането на качествена информация и какви трудности се срещат. На базата на информацията от пилотното проучване може да се установят и необходимите ресурси за провеждане на проучването и да се планират дейностите във времето.

Обикновено след пилотното проучване настъпват промени в картата за проучване – някои променливи отпадат, други се включват, трети се преработват. Така се оформя **окончателен вариант на картата за проучването**.

IV. Качество на измервателния инструментариум – надеждност и валидност

Едва ли има изследовател, който да не желае получените от изследването резултати да бъдат верни. Това субективно желание обаче често не намира обективен израз поради неспазването на

научните изисквания в хода на подготовка, провеждане и обработка на данните от изследването. Например: неточна или двусмислена дефиниция на обекта на изследването; некачествен или неадекватен за изучаваната съвкупност списък на признаците; неспазване на преднамереността на подбор; заменяне на попадналите в извадката случаи с други, по-удобни за изследване случаи. Всичко това води до *грешки, наречени систематични*.

Наред със систематичните грешки съществуват и *случайни грешки*, които, както е известно, са породени от обстоятелството, че се изследват само част от случаите, а не всички случаи на изучаваното явление.

Общата грешка е равна на сумата от двата вида грешки. Статистическата теория е разработила методи за количествена оценка предимно на грешките, произтичащи от репрезентативния характер на изследването. Останалите грешки, включващи се в категорията на систематичните, се поддават много трудно на измерване и в този смисъл са по-опасни, поради което неразделна част от методиката на регистрацията е разработването и на план за проверка качеството на попълването на въпросника.

Проблемът е: как да се провери дали онова, което е регистрирано, отразено при попълването на въпросника е вярно, дали точно отразява действителността.

Осигуряването на достоверна информация е задача, която се решава при съставянето на методиката на регистрацията. Тя зависи от това, дали е определен най-подходящият метод, с който ще се достигне до търсената информация, от подготовката на апарата, осъществяващ изследването, от правилното отношение на хората, които по един или друг начин са свързани с изследването и др.

Задачата на контрола на качеството е да се набележат мерки, с помощта на които да се провери в някаква степен качеството на реализацията на регистрираната информация, за да се получи увереност относно нейното правилно реализиране. В литературата се препоръчват различни средства за постигане на тази цел – сравняване на регистрираните данни с данни от други източници; срав-

няване на данните от един въпрос на въпросника с тези от други въпроси на същия въпросник, които имат някаква връзка помежду си; формулиране на контролни въпроси или на въпроси, изискващи по същество един и същ отговор, но с различна формулировка (например: „Възраст в навършени години“ и „дата на раждане“); сравняване на нанесените данни с първоизточниците им; определяне на надеждност и валидност.

За да се разбере материята относно надеждността и валидността като основни характеристики на качеството на измервателния инструментариум ще ги разгледаме с примери от прилагането на тестове за успеваемост на студенти.

1. Анализ и измерване на надеждността на тест за успеваемост на студенти

Надеждността (стабилността) показва точността на измерване и постоянството на резултатите при повторно измерване, т.е дали се получават еднакви резултати. Ако резултатите са идентични, то надеждността е висока. Например, за да се провери надеждността на тест за успеваемост на студенти са необходими резултатите от проведен тест.

А. Тест-ре-тест в една група. Теста се повтаря в една и съща извадка (група) в интервал от време. Времето зависи от целта и може да се движи между няколко минути до няколко години (обикновено от 4 до 6 седмици).

Алгоритъм за определяне надеждността на тест за оценяване на студенти

Ако е приложен методът на повторението – **тест-ре-тест**, може да се приложи следната последователност:

Стъпка 1. Първия вариант на съставения тест се дава на група студенти за отговаряне като се спазват изискванията по организацията и провеждането на теста.

Стъпка 2. След известен период (всеки колектив решава колко дълъг да е периодът) същият първи вариант се дава на същата група студенти за отговаряне.



Стъпка 3. Съставя се протокол за резултатите (бал, брой точки) от първото и второто тестиране. Срещу името на всеки студент в две колони се отбелязват точките от първото и второто тестиране.

Стъпка 4. От данните в двете колони на протокола се изчислява коефициент на корелация (r) между резултатите от първото и второто тестиране – използват се различни методики (формули). Изчислява се и коефициента на надеждност (r_{ii}) – най-често по формулата на Спирман-Браун – $r_{ii} = \frac{2r}{1+r}$

Стъпка 5. Определя се стандартна (репрезентативна) грешка (SE) на измерването (на надеждността) и на доверителния му интервал.

$$SE = S\sqrt{1-r_{ii}}, \text{ където}$$

S – стандартно отклонение на точките (тестовия бал), получени от всички тествани студенти

r_{ii} – коефициент на надеждност на теста като цяло

Чрез стандартната грешка се определя доверителният интервал, в който с определена вероятност за грешка ще попаднат резултатите от евентуални следващи измервания със същия измерителен инструмент, т.е. тестът.

Колкото SE има по-малка стойност, толкова доверителният интервал ще има по-тесни граници, т.е. получените оценки от теста ще съвпадат с действителните. **Например** ако получените точки от теста на даден студент са „ X “, то при 95% сигурност, реалната оценка на студента попада в интервала $r_{ii} \pm 2SE$. Ако доверителните интервали на двама студенти се застъпват, т.е. имат обща част върху числовата ос, може да се приеме (с вероятност за грешка 95%) че постиженията им не се различават съществено, а разликата се дължи на случайни фактори.

Стъпка 6. Анализ на резултатите – преценява се доколко са адекватни.

Надеждността се измерва чрез:



а. Корелационни коефициенти (r) определени между получения бал (точки) на студентите от две групи

Сложната статистическа методика (формули) на изчисляване на корелационните коефициенти не е необходимо да представяме. За това има специалисти и компютри, достатъчно е само да се разполага с такива, а също и със софтуер (EXCEL, STATGRAFICS и др.).

С първата методика чрез корелационни коефициенти определени между получения бал (точки) от студентите се определя надеждността на тест-ре-тест

Известни са достатъчно техники (формули) за изчисляване на надеждността, изискващи предварително изчислен корелационен коефициент. Не е необходимо да се познават и всички тези техники, тъй като те имат еднаква познавателна стойност и идентични резултати. Най-често използваната техника за **определяне надеждността на теста е формулата на Спирман-Браун:**

$$r_{ii} = \frac{2r}{1+r}, \text{ където}$$

r – коефициент на корелация от тестовия бал (между резултатите от тест-ре-тест, между две еквивалентни форми на един тест). Според тази формула бързо може да се прецени (изчисли), че корелационния коефициент трябва да има стойност по-голяма от 0,6, за да е достатъчно надежден теста. Не бива надеждността да е по-малка от 0,9 (понякога може да се използва и надеждност 0,8).

б. чрез χ^2 на Пирсон – двукратно в една извадка

Надеждността на тест за успеваемост на студенти – изпробван двукратно в една извадка (група) преди и след обучението – се определя чрез χ^2 на Пирсон по формулата:

$$\chi^2 = \frac{\left((ad - bc) - \frac{n}{2} \right)^2 n}{(a+b)(c+d)(a+c)(b+d)}, \text{ където (за символите вж.}$$

табл. № 5)

След като се провери значимостта на изчисления χ^2 се прави интерпретация, дали тестът е надежден и в какви граници е надеждността му.

Например: Проведен е критериален тест двукратно в група от 50 студенти. Получените резултати са представени в табл. № 12.

Табл. 5. Разпределение на студентите според резултатите от тест преди и след обучението им

Ре-тест/тест	Постигнали целта	Не постигнали целта	Общо
Постигнали целта	32 (a)	3 (b)	35
Не постигнали целта	2 (c)	13 (d)	15
Общо	34 (a+c)	16 (b+d)	50

$$\chi^2 = \frac{\left((32 \cdot 13 - 2 \cdot 3) - \frac{50}{2} \right)^2 \cdot 50}{35 \cdot 15 \cdot 34 \cdot 16} = 24,25$$

χ^2 – таблична стойност при $k = 1$ и $\alpha = 0,05$ е 6,64.

$24,25 > 6,64$, следователно има статистическа значима корелация (връзка) между резултатите от двете тествания.

От данните (само за таблица с четири полета) се изчислява коефициент на корелация (Φ) по формулата:

$$\Phi = \sqrt{\frac{\chi^2}{n}} = \sqrt{\frac{24,51}{50}} = 0,70$$

Надеждността (стабилността) на теста е 0,70

в. чрез χ^2 на Пирсон и коефициент „Капа“ - два паралелни теста

Надеждността на два паралелни варианта **А** и **Б** на тест за успеваемост на студенти се определя по два начина:

- чрез χ^2 – по гореописаната методика
- чрез коефициент „Капа“ на Д.Коуен (J.Cohen)

$$K = \frac{(P_0 - P_c)}{1 - P_c}, \text{ където}$$

P_0 – относителен дял на студентите, за които и при двете тествания е взето едно и също решение, т.е. показател за фактичестката надеждност на теста.

P_c – относителен дял на студентите, за които теоретично е възможно и при двете тествания да се вземе едно и също решение – (f_i), т.е. показател за теоретичната надеждност на теста.

$1 - P_c$ – относителен дял на студентите, за които теоретично е възможно и при двете тествания да се вземе различно решение.

Например: Резултатите от тестване на една и съща група от 50 студенти с две различни форми (А и Б) на един критериално ориентиран тест са представени в таблица 6.

Табл. 6. Разпределение на студентите според резултатите от две форми на един тест

А / Б	Постигнали целта	Не постигнали целта	Общо
Постигнали целта	36 (31,2)	4 (8,8)	40
Не постигнали целта	3 (7,8)	7 (2,2)	10
Общо	39	11	50

$$P_0 = \frac{36+7}{50} = 0,86$$

Теоретично очакваните честоти (в таблицата са поставени в скоби до фактичестките) се определят чрез просто тройно правило.

$$f_i = \frac{40 \cdot 39}{50} = 31,2$$

При четирикратна таблица щом имаме теоретична честота изчислена за една от клетките, за останалите клетки теоретичните честоти се определят чрез допълване до общия сбор на реда или колоната (тъй като общите сборове на фактичестките и теоретич-

ните честоти са еднакви). Отбелязани са в скоби до фактическите честоти в таблицата.

$$P_c = \frac{31,2 + 2,2}{50} = 0,668$$

$$K = \frac{0,86 - 0,668}{1 - 0,668} = 0,58$$

Показателят Капа може да се модифицира за случаи, когато на установено несъответствие се дава различно тегло. Например:

- несъответствие между експертите
- несъответствие между два варианта (А и Б) на теста

2. Анализ и видове валидност на теста за успеваемост на студенти

Валидността на един тест показва степента, в която той измерва това за което е създаден (знания, умения). Валидността се определя по отношение на дадена цел или дадена извадка (група). Поради това валидността на теста показва съответствието му с предварително определен външен критерий. Степента на съответствието може да бъде много висока, много ниска, но не може да липсва (има я целта, има и данни от извадка).

● Видове валидност на теста

Видовете валидност на теста са формулирани от Американската психологическа асоциация „Стандарт за образователни и психологически тестове“ (APA – Standards for Educational and Psychological Tests).

А. Критериална валидност (criterion- validity)

Б. Съдържателна валидност (content validity)

В. Коструктна валидност (construct validity)

Отделните видове валидност не са изолирани помежду си. Те взаимно се обуславят една от друга.

А. Критериална валидност означава степен на съответствие на теста на предварително дефиниран критерий. Например нормативните изисквания на учебните програми при тестове за успеваемост на студенти. Когато тези изисквания са обективни, тогава валидността на теста като показател е по-голяма.

а. Видове критериална валидност

- **Външна критериална валидност** – в основата на теста е учебната програма, т.е. резултатите от теста корелират с учебната програма
- **Вътрешна критериална валидност** – резултатите от теста се сравняват с резултатите от друг тест, който измерва знания. Коефициента на надеждност показва степента на съвпадение на резултатите при това сравнение.
- **Валидност на предсказването** – показва в каква степен изпълнението на теста е показател за бъдещи постижения. Тази валидност е много важна за тестовете за прием на студенти по медицина.
- **Конкурентна валидност** – тази валидност е близка до вътрешната валидност, тъй като се прави сравнение с норма (например брой места за обучение във ВУЗ).
- **Валидност на решението (decision validity)** – тази валидност е частен случай на конструктната валидност. Тя насочва авторите на теста към някои съображения относно причините за получените резултати от теста (дава теоретично обяснение).

Критериалната валидност се определя чрез т.н. валидност на решението – показва доколко са правомерни измерванията (например, оценките на едно тестирано лице с тест за успеваемост).

При валидизиране на критериалния тест трябва да се провери и анализира скалата, чрез която се оценяват постиженията на тестираните (за съжаление, това се прави много рядко). Емпиричното определяне на валидността на решението се извършва по следния начин:

Тестът се изпробва в извадка (група), като резултатите се оценяват количествено и се изготвя протокол. На основата на протокола се образуват две групи – на силни и слаби студенти. Определя се относителния дял на студентите, които според експертите при тестирането са оценени погрешно според възприетия стандарт за Среден 3.

б. Методи за определяне на критериалната валидност

Методите за определяне на стандарта за успешност (в английската и американската литература се обозначава като cut-off) са различни (**експертни, емпирични и комбинирани**). Какъв метод ще бъде използван, зависи от важността на решението, времето отпуснато за разработване на теста, наличните ресурси – хора, средства и др., професионалните качества на експертите и др.

Основното различие между методите за определяне на стандартите е в обекта на експертната оценка – в единия случай на експертна оценка се подлагат въпросите и задачите от теста и отделните компоненти на учебното съдържание, а в другия – изследваните студенти и тяхното ниво на професионализъм в съответна област (последните се изследват с т.н. комбинирани методи).

● Метод на контрастните групи

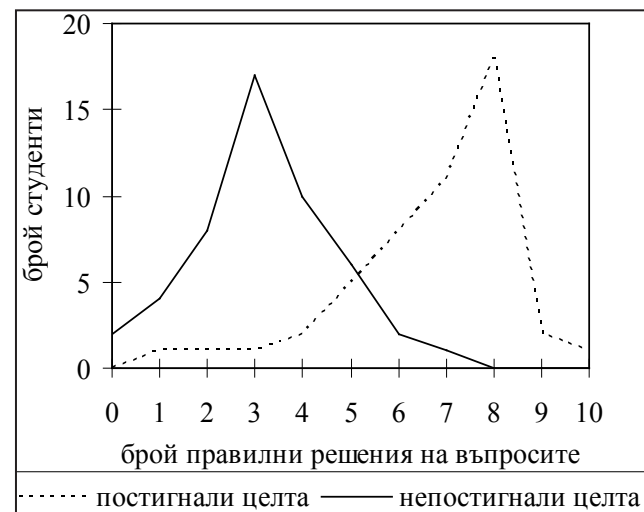
Чрез експертна оценка се определят две контрастни групи – „силна“ и „слаба“. В силната група са студентите, за които експертите смятат, че са постигнали дадената учебна цел. Условие за прилагане на този метод е обемът на извадката да не е по-малък от 30.

След провеждане на теста резултатите на двете групи се представят на една координатна система. За стандарт се приема пресечната точка на двете фигури.

Например: Един критериален тест от 10 въпроса е проведен в две групи по 50 студенти. Резултатите са представени в табл. № 7.

Табл. 7. Резултати от критериален тест на две контрастни групи

Бр. верни решения	„силна“ група	„слаба“ група
0	0	2
1	1	4
2	1	8
3	1	17
4	2	10
5	5	6
6	8	2
7	11	1
8	18	0
9	2	0
10	1	0



Фиг. 1. Разпределение на резултатите от проведен тест в две групи „силни“ и „слаби“

На фигура № 1 се вижда, че пресечната точка на двете графики е между 5 и 6 верни отговора. Ако се приеме за стандарт 5, то студентите от силната група, които имат 5 точки (те са 1+1+1+1+2=5) ще се категоризират на основата на този стандарт като „не постигнали целта“. От слабата група, тези които имат над 4 точки (те са 6+2+1=9), ще бъдат оценени като „постигнали целта“.

Следователно при така определения стандарт общо 14 от 100 студенти ще бъдат оценени погрешно, т.е. 14%. Това означава, че валидността на решението (отговорите) е $0,86\% = 1 - 0,14\%$. По същия начин могат да се оценят грешните решения (отговори) при други стандарти.

Например:

Стандарт – брой верни отговори	% на грешните отг. на „силни“ „слаби“	Валидност
4	3 19	$1 - 0,21 = 0,79$
6	10 3	$1 - 0,13 = 0,87$

Ако стандарт е 6 (от тест с 10 въпроса), т.е. при 6 и повече верни отговора е постигната целта с валидност 0,87. Вместо „силна“ и „слаба“ група трябва да се използват термините обучени и необучени.

● Метод на средната група

При метода на средната група тестът се провежда само с тези студенти, за които експертите не могат да преценят дали са постигнали или не дадената цел. След като се проведе тестът с тях, резултатите се обработват и за стандарт се взема медианата на разпределението на бала им. Следват отново процедурите за определяне на валидността при този стандарт.

Стойността, която е получена по един от горните методи (или чрез използване на статистическия дискриминантен анализ) най-добре разграничава две предварително дефинирани състояния (напр. постигнали и непостигнали учебната цел; болни и здрави), се нарича **дискриминантна стойност**. Ако за постигната учебна цел е необходим резултат по-висок от **дискриминантната стойност**, то тези над тази стойност се означават като „положителни“, а тези под нея – като „отрицателни“. Обратното е при състояния, свързани с намаляване или по-ниски резултати – тези, които са под дискриминантната стойност се означават като „положителни“, а тези

над нея като „отрицателни“. Така чрез стандартната стойност (наричана още разграничителна, cut-off point, дискриминантна, прагова, референтна граница) количествената информация се превръща в качествена отговор – „да“, респ. **положителен (+) резултат** или „не“, респ. **отрицателен (-) резултат**. При това разграничаване има опасност от две интерпретационни грешки: изследователят да приеме, че положителният резултат винаги означава постигната учебна цел (или наличие на заболяване) и отрицателният – непостигната учебна цел (или липса на заболяване). В действителност поради наличие на зона на припокриване на стойностите резултатите могат да бъдат четири разновидности:

- **Истински положителни (ИП или TP om true positive)** – брой положителни резултати, установени при студенти постигнали целта (или брой лица с наличие на дадено заболяване);
- **Истински отрицателни (ИО или TN om true negative)** – брой отрицателни резултати, установени при студенти непостигнали целта (или брой здрави лица).
- **Фалшиво положителни (ФП или FP om false positive)** – брой положителни резултати у студенти не постигнали целта (или у лица с отсъствие на даденото заболяване, т.е. при здрави лица).
- **Фалшиво отрицателни (ФО или FN om false negative)** – брой на отрицателните резултати у студенти постигнали целта (или у болни с наличие на даденото заболяване).

В практиката на научните проучвания реално не съществува разпределение на резултатите, при което кривите на изследваните групи (виж **фиг. 3.1**) са напълно разделени, т.е. всички положителни резултати са истински положителни и всички отрицателни резултати са истински отрицателни. Винаги има **зона на припокриване между истински положителните и истински отрицателните резултати**. Това означава, че сред студентите непостигнали целта (или сред здравите лица) ще има лица с фалшиво положителни

(ФП), а сред постигналите целта (или сред болните) ще има такива с фалшиво отрицателни (ФО) резултати. Това зависи от много обективни и субективни фактори (присъщото вариране на процедурите за измерване на данните и измервателните инструменти, както и различни систематични грешки), които трудно може да се контролират от изследователите.

Степента на припокриване е мярка за диагностичната стойност на даден тест, т.е. за възможността му да разграничава постигналите целта студенти от непостигналите целта или възможността на даден тест да разграничава наличие или отсъствие на дадено заболяване.

Съотношението между ИП, ИО, ФП и ФО резултати позволява изчисляването на индекси, които дават възможност за оценка на диагностичната надеждност чрез определени критерии – чувствителност, специфичност, ефективност, положителни и отрицателни предсказващи стойности.

Последователността на работата при определяне на индексите на критериите за чувствителност, специфичност, ефективност, положителни и отрицателни предсказващи стойности обхваща няколко етапа:

- **уточняване на двете групи изследвани лица** – обучаваани и необучаваани студенти; здрави и лица с конкретно заболяване; пациенти с доказан миокарден инфаркт с един метод и пациенти с доказан миокарден инфаркт с друг метод; референтна група и болни. Решаващо значение за получаване на значимо заключение има ясното дефиниране и определяне на изследваните групи.
- **представяне на резултатите от проучването в четирискратна таблица** – брой на ИП, ИО, ФП и ФО – резултатите от проучването се верифицират като се приема една или няколко стойности за стандарт (може и чрез обработка на оценките на експерти).
- **изчисляване на критериите за чувствителност, специфичност, ефективност, положителни и отрицателни предсказващи стойности по следните формули:**

$$1. \text{ Чувствителност (\%)} = \frac{\text{ИП}}{\text{ИП} + \text{ФО}} \cdot 100.$$

Това е вероятността (в %) за положителен резултат у студент постигнал целта или у болен, т.е. способността на даден тест да определи студентите постигнали целта или пациентите с определено заболяване (положителните лица). При сериозни, но лечими заболявания е желателна възможно най-висока чувствителност на диагностичните тестове, близка до 100%, за да не се пропуснат такива заболявания и да се избегне хронифицирането им.

$$2. \text{ Специфичност (\%)} = \frac{\text{ИО}}{\text{ИО} + \text{ФП}} \cdot 100$$

Това е вероятността (в %) за отрицателен резултат у студент непостигнал целта или у здраво лице, т.е. способността на даден тест да определи студентите не постигнали целта или лицата без определеното заболяване (отрицателните, негативните). Необходима е възможно най-висока специфичност на даден диагностичен тест в медицината, когато дадено заболяване се диагностицира трудно и е трудно лечимо или нелечимо засега.

$$3. \text{ Диагностична ефективност или точност (\%)} = \frac{\text{ИП} + \text{ИО}}{\text{ИП} + \text{ИО} + \text{ФП} + \text{ФО}} \cdot 100.$$

Това е вероятността (в %) за правилно класиране чрез теста, т.е. способността на даден тест да определи студентите постигнали и не постигнали целта или лицата със и без заболяване (положителните и отрицателните). Възможно най-висока ефективност (висока чувствителност и висока специфичност) се изисква при тежки заболявания, чието лечение е съпроводено със сериозни странични прояви, тъй като както ЛП, така и ЛО резултати имат еднакво тежки последствия за пациента.

$$4. \text{ Положителна предсказваща стойност (\%)} = \frac{\text{ИП}}{\text{ИП} + \text{ФП}} \cdot 100.$$

Показва каква е частта (в %) на постигналите целта или на болните всред положителните резултати.

$$5. \text{Отрицателна предсказваща стойност (\%)} = \frac{ИО}{ИО + ФО} \cdot 100.$$

Показва каква е частта (в %) на непостигналите целта или на здравите лица сред тези с отрицателен резултат.

Трябва да се отбележи, че в медицинската практика според клинични съображения е допустимо изместването на стандарта, което може да доведе до следните последици:

- чувствителността намалява, а специфичността нараства (и обратно)
- намалява делът на ФП, но нараства делът на ФО (и обратно)
- рязко нараства положителната прогностична стойност
- рязко нараства диагностичната ефективност (точност)

Чувствителността на теста трябва да се увеличава за сметка на по-ниска специфичност, когато определянето на положителен резултат има решаващи последици за живота на болния, т.е. при тежко, но лечимо заболяване. Обратно, при диагностициране на заболяване с неблагоприятна прогноза и болни в неизлечимо състояние е необходима максимално висока специфичност, т.е. висока вероятност за отрицателен резултат у лице без заболяване. Най-честият клиничен проблем е да бъде включено в диагностичната схема едно състояние, което само се подозира, но липсват достатъчно данни, за да бъде отхвърлено. Потвърждаването на една лекарска хипотеза се подпомага от изследвания с висока положителна предсказваща стойност, докато отхвърляне на хипотезата се подкрепя от висока отрицателна предсказваща стойност.

Б. Съдържателна валидност на тестовете за оценяване

Съдържателната валидност е показател за това, доколко включеното в теста учебно съдържание е адекватно усвоено според изискванията на учебната програма. За целта **експертите преценяват** дали учебното съдържание предвидено за усвояване е отразено в теста, т.е. дали теста измерва знанията, уменията и навиците, които се изискват от учебната програма. С други думи казано дали извадката от въпроси и задачи в теста е представителна за учебното съдържание.

Трябва се отбележи, че в историята на нормативните тестове анализът на съдържателната валидност на теста винаги е извършван с помощта на експерти, а коефициента на валидност се пресмята като корелационен коефициент между оценките на експертите. В зависимост от данните се изчисляват корелационни коефициенти при количествени данни, бисериални или точково-бисериални коефициенти на корелация.

- Ако данните са нормално разпределени и имат количествен характер се използва корелационен коефициент на Пирсън-Браве или друг за измерими данни (интервална скала).
 - Ако данните са от номинална скала в алтернативна форма се използва коефициент на корелация на Пирсън –
- $$r = \sqrt{\frac{\chi^2}{n}}$$
- Ако данните са от номинална скала с повече от две разновидности се използва коефициент на контингенция –
- $$r = \sqrt{\frac{\chi^2}{n + \chi^2}}$$
- Ако данните са от ординална скала се използва корелационен коефициент на Спирман:

$$\rho = 1 - 6 \sum \frac{D^2}{N(N^2 - 1)}, \text{ където}$$

D – разлика в ранговите места

N – обем на извадката

В. Конструктивна валидност (construct validity)

Трябва да се прави разлика между конструктивна и конструктивна валидност. **Конструктивна валидност** – това са теоретичните основания и съображения, с помощта на които по-точно се обясняват конкретните постижения, например на студентите. Това е показател за това, доколко изпълнението на теста (постиженията) могат да се обяснят с такива „конструкти“ като интелигентност, съобразителност, памет и т.н., т.е. психическото образуване на личността. Конструктите са предимно предположения, които не винаги може да се докажат емпирично. Това валидизиране на теста изисква продължително събиране на данни от различни източници от една страна, и теоретичното им осмисляне, от друга.

За определяне на конструктивната валидност се използват следните **методи**:

- **Факториален анализ** – за връзки и зависимости между отделни субтестове проведени в една група. Така се определят основните фактори за постиженията.
- **Вътрешна съгласуваност** – между резултатите по отношение на отделни въпроси или субтестове. Така се установява делът на определящия фактор.
- **Корелация с други тестове** в същата област изготвяни преди години. Така се обясняват по-големите постижения в момента на децата в сравнение с техните родители.
- **Корелация с тестове, предназначени за други характеристики**

Тези методи са взаимно свързани когато се обясняват конкретните резултати от теста.

Проверката на конструктивната валидност изисква:

- Определяне на конструктите, които биха могли да обяснят резултатите от теста
- Формулиране на хипотези относно определените конструкти

- Събиране на данни за проверка на хипотезите
- Проверка на хипотезите чрез анализ на данните за силните и слабите студенти, изчисляване на корелационни коефициенти между двете групи.

Конструктивната валидност е по-широко понятие от критериалната и съдържателна валидност. Тя насочва вниманието на авторите на теста към причините за различните резултати и теоретичното им обяснение.