ОЦЕНКА НА МЕТОДИ ЗА ДИАГНОСТИКА И ПРОГНОЗИРАНЕ – АНАЛИТИЧНИ ПРОЦЕДУРИ И ИНТЕРПРЕТАЦИЯ НА ДАННИТЕ

Стефан Матеев

1. Индекси за оценка на качество на диагностика

Диагностиката е важна дейност в клиничната и психологичната практика. При диагностика, специалистът е изправен пред задачата да прецени дали изследваното лице, пациент, или субект, страда от определено заболяване, или психично нарушение. Специалистът прилага определена *методика*, и въз основа на резултата от методиката преценява дали е налице заболяване или нарушение. Резултатите от методиката се изразяват в числен вид. Приемаме, че колкото е по-голямо числото, или *показателя*, толкова повече специалистът е склонен да даде положителен отговор, т.е. да реши, че е налице заболяване или нарушение. Това е обичаен начин за взимане на решение, например, при клинична оценка на резултати от кръвни проби, или при психологична оценка на резултати от специализиран въпросник. При прогнозирането, нещата седят по подобен начин. Прилага се определена методика и въз основа на получените данни се решава дали след известно време лицето ще развие дадено заболяване или разстройство. Разликата между диагностика и прогнозиране се състои единствено в реда на събитията; при диагностиката заболяването може би е вече налице, и тогава се прилага методиката, при прогнозирането първо се прилага методиката и по-късно се установява дали заболяването е настъпило. Разликата в реда на събитията не играе роля; диагностиката и прогнозирането се анализират с едни и същи методи.

Диагностиката се усложнява от факта, че взимането на решение въз основа на данните от методиката има вероятностен характер. Възможно е, измерваният показател да има висока стойност, но въпреки това, заболяване да отсъства. Възможно е също, заболяване обективно да е налице, но да се получат ниски стойности на показателя, които да водят до отрицателно решение, че заболяване или нарушение отсъства. Вероятностният характер на диагностиката поставя важен въпрос: доколко може да се вярва на прилаганата методика? Или, какво е нейното качество, до каква степен методиката може да се използва за различаване между здрави лица и лица с наличие на патология? Развитието на медицината и психологията води до създаване на нови методики за диагностика. Често те са значително по-сложни по-скъпи от старите и специалистите основателно се питат, дали си заслужава да ги използват. Настоящата статия е посветена на въпроса за оценката на дадена методика.

Оценката на диагностичните качества на дадена методика се извършва по следния начин. Подбират се две групи лица. Едната група се класифицира като нормална или контролна, т.е. по независими и обективни критерии се установява, че лицата в нея са здрави, поне по отношение на заболяването от интерес. Другата група се класифицира като анормална, като при лицата в нея заболяването е твърдо установено. Често тази група се нарича експериментална, което не е съвсем точен етикет. Изследването не включва целенасочено третиране на изследваните лица, което е характерно за експерименталните изследвания. Въпреки това, по-надолу ще използваме тези два етикета, контролна и експериментална група, тъй като те са обичайни в речника на изследователите.

Лицата от всяка група се подлагат на изследване с методиката, която подлежи на оценка. Както упоменахме по-горе, данните се получават в числен вид. Ако това са например, изследвания на кръвни проби, данните се получават директно във единици mmol/L. Но това не винаги е възможно. Типичен пример (Hanley & McNeil, 1982) представлява методиката, при която се прави рентгенова снимка, например на гръдния кош, и специалист (условно наречен експерт) преценява по снимката наличието или отсъствието на злокачествено образувание (лезия). Удобно е преценката да се прави с т.н. оценъчен метод (rating method). Експертът разполага с фиксиран брой категории на отговор относно наличието на лезия, като: „твърдо няма“, „по-скоро няма“, „може и да има, но може и да няма“, „по-скоро има“ и „твърдо има“. Това са пет категории, на които се приписват числата от 1 до 5, като 1 съответства на „твърдо няма“ и 5 – на „твърдо има“. Така, всяко изследвано лице получава числена оценка на състоянието си. Подобни категориални оценки са типични и за психологичната практика.

По този начин, всяко изследвано лице получава една числена оценка на състоянието си, определена, или по-правилно измерена, чрез методиката. Означаваме тези числа с X, като данните, получени от контролната група, означаваме с X_C, а данните от експерименталната група, с X_E. Данните от двете групи образуват две разпределения, които имат обща абсциса X. Здравият смисъл показва, че ако разпределенията се припокриват напълно, методиката не струва. Двете групи не могат да се разграничат от данните. Ако двете разпределения са напълно разделени едно от друго, методиката е перфектна, тя напълно разграничава контролната и експерименталната групи. Това са екстремални случаи, на практика разпределенията се припокриват в някаква степен.

Тук следва да предупредим читателя, да не бърза да пресметне средните стойности на разпределенията и да провери, дали разликата между тях е „значима“. Проверката на значимостта става, като се пресметне вероятността за т.н. грешка от първи род и се провери, дали тя е по-малка от 0.05 (понякога от 0.01). Този статистически метод може да даде отговор на някои въпроси, но в последните години той се отхвърля като начин за психологически или клиничен извод от данни. Изискванията, които се представят пред общността от изследователи е да описват данни с помощта на сила на ефекта (effect size) и доверителни интервали (confidence intervals). Изтъква се, че пресмятането на значимости, покрай други проблеми, които то създава, не носи необходимата информация за това, доколко два масива от данни са еднакви или не. По този въпрос препоръчваме подробните разглеждания на Cumming (2012, 2014).

Мярка за това, доколко две разпределения се припокриват, бе предложена още преди около тридесет години. Cohen (1988) предложи силата на ефекта, ES, да се пресмята като

$ES = |m_C – m_E|/s$,

където m_C и m_E са средните стойности на данните, получени от контролна и експериментална групи, а s е тяхното общо стандартно отклонение. Ако ES се получи нула, разпределенията се припокриват напълно. Когато разпределенията започнат да се разделят, ES прогресивно нараства. Стойности на ES около единица и повече говорят за съществени разлики между разпределенията. Този индекс може да се приложи успешно за случая на оценка на ефективността на методика, приложена на контролна и експериментална групи. Колкото е по-висока стойността на ES, толкова по-ефективна е методиката. При ES = 2 и повече, методиката е почти перфектна.

Индексът ES обаче, не е без проблеми. Преди всичко, той „работи“ добре при нормално разпределени данни. Ако разпределенията се отклоняват силно от нормалното, индексът става безсмислен. Второ, индексът следва да се прилага само в случаите, когато данните от методиката се получават в рамките на интервална скала или скала на отношения. Данните X, които са получени от приписване на категории, са в рангова скала, и операциите пресмятане на средна стойности стандартно отклонение са на практика съмнителни, а математически безсмислени (виж Матеев, 2011 за разглеждане на скалите на измерване). И накрая, разбирането на текстове, в които се дават сведения за ефективността на дадена методика в термини на ES, изискват поне минимални познания по статистика, които не може да се очакват от, например, общопрактикуващи лекари. Някои други статистически индекси, които не разглеждаме тук, изискват даже повече от базови познания по статистика.

В последните години започна да набира скорост един индекс за сила на ефекта, наричан вероятност за превъзходство (probability of superiority, P_{SUP}). Той се състои в следното. Нека контролната група е с обем n_C лица, а експерименталната – с обем n_E лица. Всяко от тези лица си има свое измерване от методиката, X. Нека по случаен начин изберем по едно лице от двете групи и сравним показателите им X_E и X_C. Тъй като по-високите стойности X означават принадлежност към експерименталната група, следва да се очаква, че $X_E > X_C$ за тези две лица. Но каква е вероятността случайно избрано лице от експерименталната група да има по-висока стойност от случайно избрано лице от контролната? На този въпрос може да се отговори, като разгледат всички възможни сравнения между лицата от двете групи, и се преброят случаите, в които се получава $X_E > X_C$. Броят на възможните сравнения е n_E*n_C[1], и тогава оценката на вероятността за превъзходство P_{SUP} е

$$P_{SUP} = n(X_E > X_C)/n_E*n_C \qquad (1)$$

където $n(X_E > X_C)$ е броят на случаите, при които се е оказало, че $X_E > X_C$. Оказва се, че вероятността за превъзходство е много удобен индекс за описание на ефективността на методиката. Ако $P_{SUB} = 0.5$, то методиката определено не струва. Половината от лицата в едната група имат по-високи показатели от лицата от другата група, и половината имат по-ниски показатели. Ако знаем колко е показателят, не можем да установим от коя група е лицето. Ако вероятността е единица, тогава сме сигурни, че всички лица от контролната група имат по-ниски показатели от тези на експерименталната група. В

[1] Нека експерименталната група се състои от лицата A, B, и C, а контролната от лицата D, E, F и G. Лицето A се сравнява с всеки от D, E, F и G. Лицата B и C поотделно също се сравняват с всяко от четирите лица от контролната група. Така броят на всички сравнения става 3*4 = 12, или n_E*n_C.

реалните измервания вероятността не е нито 0, нито 1. Тя може да служи за индекс за качеството на методиката, и то да бъде сравнявано с качеството на други методики (Матеев, 2014a).

Основните предимства на вероятността за превъзходство са следните. Първо, формата на разпределенията на данните от двете групи не играе роля. Индексът може да се използва и при разпределения които силно се отклоняват от нормалното. Второ, индексът може да се използва и при данни, които са получени в рангова скала, каквито са категориалните оценки. При пресмятането му се извършват само операциите броене и сравняване, които са смислени и математически допустими в рангова скала. Пресмятанията не изискват броят на лицата в двете групи да е еднакъв. И не на последно място, разбирането на индекса не изисква даже и базови познания по статистика. Всеки знае (или поне си мисли, че знае) какво е вероятност, и няма да си задава трудни за отговаряне въпроси, когато чете описание на качествата на една методика.

Любознателният читател вероятно се е досетил за недостатъка на определянето на P_{SUB} според равенство (1). При прегледа на всичките сравнения между X_E и X_C може да се окаже, че в част от тях се получава $X_E = X_C$, т.е. две лица, едното от контролната и другото от експерименталната група имат еднакви стойности на показателя от методиката. В такива случаи се казва, че между данните от двете групи има *връзки* (ties).

Връзки между двете групи е възможно да няма, ако методиката дава измервания в непрекъсната скала. Например, ако методиката се основава на определяне на кръвно налягане с уред, който измерва с точност от стотни от mmHg, най-вероятно няма да има измервания в едната група, които да съвпадат с измервания от другата група.

Подобен е и случаят, при който времена на отговор се измерват с точност до 1 ms, малко вероятно е такива числа да се повтарят. Но ако измерването става с точност от 5 mmHg (което е обичайна практика), много е възможна появата на повтарящи се измервания, особено ако групите са с по-голям брой лица. Следователно, появата на връзки може да се определя от точността на методиката за измерване, и от броя на лицата в групите. Връзки почти сигурно ще има, ако измерването се осъществява по категориална, например 5-бална скала. Ако групите са от по около 20-30 лица, със сигурност ще има лица от двете групи, чиито оценки по скалата ще са еднакви. Накрая, дори и скалата на измерване да е почти непрекъсната, при обобщаването на данните те често се обединяват в класове (bins) с цел по-нагледното им представяне с хистограма на разпределение. Изследователят може да се изкуши да припише категории на класовете и да проведе по-нататъшния анализ с категориите, вместо със суровите данни. Отново появата на връзки е неизбежна. Възниква въпросът: така или иначе, връзки в данните може да има, какво да правим с тях?

Постъпваме по следния прост начин. Половината от броя $n(X_E = X_C)$ добавяме към броя $n(X_E > X_C)$, а другата половина към останалия брой случаи, в които имаме $n(X_E < X_C)$. Формално това може да се извърши по следния начин. Преглеждаме всичките n_E*n_C двойки числа, и определяме за всяка двойка величина S по следния начин:

$S = 1$ ако $X_E > X_C$ (2a)

$S = 0.5$ ако $X_E = X_C$ (2b)

$S = 0$ ако $X_E < X_C$ (2c)

след което сумираме всичките стойности на S и ги разделяме на общия брой на двойките за сравнение n_E*n_C, или

$$A = [n(X_E > X_C) + 0.5*n(X_E = X_C)]/n_E*n_C \tag{3}$$

Ако изразим равенство (3) чрез относителни дялове, получаваме

$$A = P(X_E > X_C) + 0.5*P(X_E = X_C) \tag{4}$$

Изразите (3) и (4) обозначаваме със символа А както е въведен от Vargha & Delaney (1998), Delaney & Vargha (2002). Индексът А не е точно вероятността за превъзходство P_{SUP}, както е дефинирана с равенство (1). Когато показателите на две от лицата са еднакви, не може да се говори за превъзходство. Само ако отсъстват връзки, тогава равенства (1) и (3), съответно и (4), са еквивалентни. Независимо от това, както показват Vargha & Delaney (1998), индексът А има свойства, подобни на тези на P_{SUP}. При $n(X_E > X_C) = n(X_E < X_C)$ имаме $A = 0.5$. Ако всички сравнения показват, че $X_E > X_C$, тогава $A = 1$. Vargha & Delaney (1998) наричат А индекс за стохастично превъзходство на променливата X_E над X_E; те избират символа А от първата буква на “above”. В крайна сметка, името на индекса и буквата не са толкова важни, по-надолу ще видим, че същият индекс има и друго име. При анализа на данните следва да има винаги пред вид, какво точно се изчислява, и от какви данни. Както ще видим, наличието или отсъствието на връзки може да играе съществена роля при анализа и интерпретацията на данните.

2. Пресмятане на индекса за стохастично превъзходство чрез сумиране на рангове

Пресмятането на индекса А може да стане *точно* чрез преброяване на изходите от сравненията, представени от равенства (2a), (2b) и (2c) и пресмятане на равенство (3). Преброяването на изходите на ръка е изтощителна процедура, особено когато броят на лицата в групите е голям. При наличието на съвременните мощни статистически програми, това вече не е особен проблем. Тук обаче, се стараем да обясним смисъла на изчислителните методи, поне доколкото е възможно. Смятаме за порочна практиката, данните да се подадат на програмата, да се натисне Enter, и да се приеме изходния резултат, без да се има представа какво точно извършва програмата. В този раздел представяме отделните етапи на пресмятане на индекса A с помощта на числен пример. Етапите включват подреждане (ранжиране) на данните и аритметични операции, които лесно се осъществяват с програмата Excel.

В Таблица 1 са представени данни от измервания на показател, получен с прилагането на 5-бална скала. Данните от контролната група, (n_C = 10), са кодирани с 0, е тези от експерименталната – с 1 (n_E = 8). Кодовете са нанесени във втория ред на таблицата. В третия ред са нанесени суровите данни. Вижда се, че има връзки между данните от двете групи, например лицето E от експерименталната група има същия показател 4, както и лицето K от контролната група. Има и още връзки. Следва да се отбележи, че повтарянето на един и същи показател в рамките на една и съща група не се класифицира като връзка. Връзка имаме само когато повторението е между групите, т.е. когато $X_E = X_C$ за дадена двойка лица.

Таблица 1. Лицата от A до H принадлежат към експерименталната група (код 1), а лицата от I до R – към контролната (код 0). Данните от 5-балната скала са нанесени в реда „показател“, като високите стойности на показателя показват увереност в положително решение. В четвъртия ред са нанесени средните рангове на всички показатели.

лице	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
код	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
показател	5	5	4	4	4	1	3	2	1	1	4	1	2	2	3	4	3	2
среден ранг	17.5	17.5	14	14	14	2.5	10	6.5	2.5	2.5	14	2.5	6.5	6.5	10	14	10	6.5

Първият етап е подреждането, или ранжирането на тези данни, като ранжирането се извършва общо за двете групи, т.е. за данните с код 1 и код 0 заедно. В Excel то се осъществява с командата =rank.avg(number;ref;order). За order слагаме 1, тогава ниски рангове се приписват на ниски стойности на показателя. Тази команда приписва т.н. средни рангове на еднаквите стойности на показателя. Това е важно, защото по този начин се запазва общата сума на ранговете, която трябва да е равна на сбора на числата от 1 до n_E+n_C. Ранговете са нанесени в третия ред на Таблица 1. За примера този сбор е $1 + 2 + 3+ \ldots\ldots 18 = 171$.

Следващият етап е пресмятането на сбора на ранговете на показателите, получени от двете групи, но поотделно. За примера те са R1 = 96 за числата с код 1, и R0 = 75 за числата с код 0. Сборът на ранговете с код 1 е по-голям, които показва , че при ранжирането, тези данни като цяло са по-напред от данните с код 0. Това е добър знак в полза на методиката, която анализираме. Ако двата сбора са приблизително равни, това означава, че групите не могат да се разграничат.

Следващата стъпка е да пресметнем броя на сравненията $n(X_E > X_C)$, като добавим към тях половината от $n(X_E = X_C)$, т.е. сбора $n(X_E < X_C) + 0.5*n(X_E = X_C)$.

Това става по следния начин:

$$n(X_E > X_C) + 0.5*n(X_E = X_C)] = n_C*n_E + n_C*(n_C+1)/2 – R0 = U1 \qquad (5)$$

За примера

$U_1 = 10*8 + 10*(10 + 1)/2 -75 = 80 + 55 – 75 = 60$

$$n(X_E < X_C) + 0.5*n(X_E = X_C)] = n_C*n_E + n_E*(n_E+1)/2 – R1 = U2 \qquad (6)$$

За примера

$U_2 = 10*8 + 8*(8 + 1)/2 – 96 = 80 + 36 – 96 = 20$

Стойностите U_1 и U_2 са статистики, които се извличат от непараметричния статистически тест на Ман-Уитни за значимост на разлика между две извадки, които не са нормално разпределени (виж например Калинов, 2013). Тук показваме смисъла на тези статистики – те ни информират за броя и вида на сравненията между показателите, които се получават от двете групи. Сборът $U_1 + U_2 = 60 + 20 = 80$ е равен на общия брой на всички сравнения $n_E*n_C = 8*10 = 80$, както и трябва да бъде. Последният етап от изчисленията е просто да се пресметне относителният дял A от равенство (3). За разглеждания пример той е равен на по-голямото число от U_1 и U_2, разделено на общия брой на сравненията $n_E*n_C = 80$. За примера това е

$$A = U1/n_C*n_E = 60/80 = 0.75 \quad (7)$$

Така пресметнатият индекс описва качествата на методиката на измерване. Индексът не зависи от вида на разпределенията на данните в двете групи, също така той може да се използва и при методика, в която показателят се получава в рангова скала.

Според съвременните изисквания, освен индекс за силата на ефекта, описанието на едно изследване трябва да включва и доверителен интервал. За да се построи доверителен интервал, трябва предварително да се пресметне стандартната грешка SE на оценката на параметъра, в случая на индекса A. Тя се пресмята по следния начин

$$SE = \{[A*(1 - A) + (n_E - 1)*(Q_1 – A^2) + (n_C – 1)*(Q_2 – A^2)]/n_E*n_C\}^{0.5} \quad (8)$$

където Q_1 и Q_2 могат да се оценят с добро приближение (Hanley & McNeil, 1982) като

$$Q_1 = A/(2 – A) \quad (9)$$

и

$$Q_2 = 2*A^2/(1 + A) \quad (10)$$

За примера Q1 = 0.600 и Q2 = 0.642 и тогава SE = 0.121

95%ният доверителен интервал ДИ се пресмята по „класическия“ начин, именно като ДИ = A ± 1.96*SE.

Коректно е изследователят да съобщи, че е установил индекс за качеството на диагностичната методика A = 0.75, като за примера границите на 95%-ния доверителен интервал на A са [0.99, 0.51]. Стойността на A не е лоша, но доверителният интервал е огромен, което говори за много ниска точност на определяне на A. При малкия брой лица в този хипотетичен пример друго не може да се очаква. Ограниченията на този начин на пресмятане на доверителен интервал са известни – той не следва да се използва при малък брой на лицата n в някоя от групите и при ниски стойности на (1 – A). За съжаление, все още не са ни известни разработени точни и достъпни методи за пресмятане на доверителния интервал на A без тези ограничения.

3. Пресмятане на индекса за стохастично превъзходство A чрез площ под кривата на работната характеристика (ROC–крива)

При прилагането на методиката за диагностика, на специалиста му се налага да вземе решение за положителен или за отрицателен отговор. За тази цел, той трябва да приеме критерий, който е една възможна стойност на показателя. Ако изхода от изследването е довел до показател, по-висок от критерия, решението е положително, т.е. има заболяване или разстройство. Ако полученият показател е по-нисък от критерия, решението е отрицателно, т.е. лицето е нормално. Вероятностният характер на взимането на решение води до това, че не винаги лицата от експерименталната група дават показатели, които са над критерия. Също така, лица от контролната група е възможно да дадат резултати, които да надхвърлят критерия, и така погрешно да бъдат обявени за носители на заболяване.

Вярно положително решение се нарича събитието, при което показателят на лицето е по-висок от критерия и то е от експерименталната група. Относителният дял на тези случаи се нарича сензитивност.

Грешно положително решение се нарича събитието при което показателят на лицето е по-висок от критерия, но то е от контролната група. Относителният дял на тези случаи се нарича (1-специфичност).

Тези два относителни дяла определят координатите на точка на графика, по абсцисата на която се нанася (1-специфичност), а по ординатата – сензитивност. При промяна на критерия, тези координати също се менят и изписват т.н. крива на работната характеристика (receiver-operating characteristic), или ROC-крива.

Подробното разглеждане на ROC-кривата е извън целите на настоящия текст. За по-достъпно запознаване с въпроса препоръчваме Матеев (2011). Това, което е важно за практикуващия специалист е следното. *Площта под ROC-кривата има смисъл на относителен дял, и строго е доказано, че той е равен точно на индекса A, дефиниран от равенство (3)* (който също е относителен дял).

Площта под кривата на работната характеристика, или ROC-кривата, може да се пресметне с помощта на широко разпространения статистически пакет SPSS. Освен площта, програмата пресмята и стандартната грешка SE чрез равенство (8), както и доверителния интервал на площта. Представянето на резултати от оценка на методика рутинно се представят с площ под ROC-кривата (Zweig & Campbel, 1993). В този текст изясняваме, че въпросната площ е равна точно на сбора на два относителния дяла, този на случаите $X_E > X_C$ и този на половината от случаите $X_E = X_C$. Тук спестяваме подробното описание на операциите по пресмятане на площта под ROC-кривата с SPSS, те са подробно описани в тюториала на програмата.

При пресмятания с SPSS изследователят може да обърне внимание, че при представяне на площта под ROC-кривата, програмата дава предупреждение, че „данните съдържат връзки (ties), поради което стойността на площта може би е *изместена* (biased)“. „Изместване“ на дадена статистика означава, че тя не отразява коректно (занижава) популационния параметър, който се оценява. В случая, популационната стойност на площта под ROC-кривата е тази, която хипотетично би се получила, ако се изследват безкраен брой лица (или цялата популация) и ако методиката е такава, че *няма връзки в данните.* Двете групи, контролна и експериментална, могат да се разглеждат като извадки от популациите. Ако няма връзки между показателите, получени от двете групи, тогава оценката на площта под ROC-кривата представлява неизместена оценка на площта под популационната ROC-крива, и SPSS не дава предупреждение за изместване. Но ако има връзки, то оценката е изместена (занижена), като изместването е толкова по-голямо, колкото е по-голям броя на връзките в данните. За съжаление, не ни е известен достъпен начин за коригиране на изместването. Следва да предупредим читателя, че изместването ще е налице, дори и ако площта пад ROC-кривата се пресмята по начина, описан в Раздел 2. Сметките, описани в този раздел са подобни на тези, които се извършват с SPSS. Това, което може да посъветваме изследователя, е по възможност да избягва методики, при които показателят се извлича по скала с малък брой категории. Тогава процентът на връзки в данните ще е висок, и съответно ще се получи занижена стойност на индекса A, както и на площта под ROC-кривата. Zweig & Campbel (1993) разглеждат пример, в който неизместената (без връзки в данните) оценка на площта под ROC-кривата е 0.747, но при разделяне на данните в 8 класа (което води до поява на връзки), площта спада до 0.718.

Изместване и точност на определяне на A или на площта под ROC-кривата са различни неща. Точността се определя от размера на доверителния интервал, който от своя страна, зависи от големината на стандартната грешка. От равенство (8) може да се види, че стандартната грешка намалява с нарастването на A и с увеличаването на обема на двете извадки. Броят на връзките между показателите влияе върху определянето на стандартната грешка, доколкото A се занижава поради изместването. Това влияние не е особено силно и може да се пренебрегне.

Пресмятането на индекса A или на площта под ROC кривата е полезно и препоръчително не само в случая на диагностика, клинична или психологична. То може да се приложи и когато изследователят трябва да опише и интерпретира разлика между данни от две различни групи лица, когато разпределенията им се отклоняват от нормалното и когато данните са получени в рангова скала. Както споменахме по-горе, индексът ES не е подходящ за такива данни. Препоръчваме на читателя да използва програмата SPSS като изчислителен инструмент. Пресмятанията в Раздел 2 са повече с цел да обясним какво точно се пресмята със SPSS и по този начин да улесним интерпретацията на данните.

Литература

Калинов, К. (2013). *Статистически методи в поведенческите и социалните науки*. Нов български университет, София (3-то издание)

Матеев, С. (2011). *Начала на психофизиката*. Нов български университет, София

Матеев, С. (2014). Същественост и значимост на резултати от научни изследвания в психологията. *Научен електронен архив на НБУ*, http://eprints.nbu.bg/2277/

Матеев, С. (2014а). Вероятност за превъзходство“ – лесно разбираем индекс за сила на ефекта при описание на данни. В: *Сборник научни доклади от VII-я Национален* *конгрес по психология* (ред. Джонев, С., Димитров, П. & Матеева, Н.), Продуцентски център ЛМ ЕООД, София, стр. 337-346

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Lawrence Erlbaum Ass., Hillsdale, New Jersey

Cumming, G. (2012). *Understanding the new statistics. Effect sizes, confidence intervals, and meta-analysis.* New York, NY: Routledge.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25,* 7-29

Delaney, H.D. & Vargha, A. (2002). Comparing several robust tests of stochastic equality with ordinally scaled variables and small to moderate sized samples. *Psychological Methods,* **7**, 485-503

Hanley, J.A. & McNeil, B.J. (1982). The meaning and use of the area under the receiver operating characteristic (ROC) curve. *Radiology*, **143**, 29-36

Vargha A, & Delaney, H.D. (1998). Kruskal-Wallis test and stochastic homogeneity. *Journal of Educational and Behavioral Statistics*, **23**, 179-192

Zweig, M.H. & Campbel, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, **39**, 561-577