

КРИВОРАЗБРАНИТЕ ВЕРОЯТНОСТИ ПРИ ТЕСТОВЕ ЗА НАЛИЧИЕ НА ЗАРАЗА

Доц. д-р Маргарита Ламбова, гл. ас. д-р Вania Стоянова

Икономически университет – Варна

Резюме. В представеното изследване са разгледани проблеми при възприятието и популярни заблуди при интерпретацията на информация, свързана със степента на сигурност/несигурност на тестове, чрез които се проверява наличието или отсъствието на зараза. Разкрива се теоретичната същност на вероятностите за фалшиво положителен и фалшиво отрицателен резултат, както и на степента на сигурност/несигурност на тестовите резултати и на тази основа са формулирани предположения относно възприятието на подобна информация и възможни заблуди при нейната интерпретация от страна на обществото. Направените предположения се проверяват на база информацията от непредставително анкетно проучване. Получените резултати разкриват затруднения на потребителите при осмисляне и тълкуване на предоставените данни, както и при откриване на логически връзки, характерни за подобен тип информация.

Ключови думи: условни вероятности; теоремата на Бейс; ниво на увереност; тест за наличие на инфекция

В условията на пандемия информацията, свързана със заразата, е вездесъща и се множат показателите за нейното характеризиране, измерители, представящи степен на ефективност на ваксини, както и степен на сигурност/несигурност на тестове, чрез които се проверява за наличие или отсъствие на зараза. Пандемията провокира и задейства огромна лавина от данни, която, от една страна, подпомогна разкриването на определени закономерности, но от друга – с „тежестта“ си парализира донякъде числовото „зрение“ на потребителите ѝ. Начинът за поднасяне на информацията в много случаи не предразполага към лесно осмисляне и води до числова слепота и замъглено мислене (термините са заимствани от Gigerenzer 2017), които не позволяват правилна интерпретация и водят до формиране на представи за несъществуващи закономерности. Съмнителното качество на част от данните е проблем, който тук няма да бъде засегнат, въпреки че отчасти той също се дължи на логически недоразумения – следствие от замъгленото мислене.

Проблеми при осмислянето и правилната интерпретация на информацията възникват най-често, когато тя е под формата на условни вероятности, относителни честоти или е свързана с относителни различия.

Изследването има за **цел** разкриването на проблеми при възприятието и популярни заблуди при интерпретацията на информация, свързана със степента на сигурност/несигурност на тестове, чрез които се проверява наличието или отсъствието на зараза.

Постигането на целта се осъществява чрез решаването на две основни задачи.

1. Разкриване на теоретичната същност на вероятностите за фалшиво положителен и фалшиво отрицателен резултат, както и на степента на сигурност/несигурност на тестовите резултати и на тази основа формулиране на предположения относно възприятието на подобна информация и възможни заблуди при нейната интерпретация от страна на обществото.
2. Проверка на направените предположения относно осмислянето на вероятностите и възможните заблуди при тяхната интерпретация от страна на потребителите на информацията на базата на непредставително анкетно проучване.

Следва да бъде подчертано, че предмет на изследването е възможната дезинформация на база числова слепота, а не същността, спецификата и ефективността на определени тестове, следователно не се цели възхвала или дискредитиране на такива, което е извън полето на нашата компетентност.

Резултатите от изследването биха могли да помогнат за създаването на по-добра представа за начина на възприемане на информация, зададена под формата на условни вероятности, от страна на обществото. Стремешът е да бъде даден импулс на търсенето на възможности за усъвършенстване на учебните програми по математика и статистика в средното и висшето образование с цел превъзможване на проблемите, свързани с възприятието и интерпретацията на подобна информация.

1. Сигурността на резултатите от тестовете за наличие на зараза – известна константна величина или почти неуловима променлива?

Под сигурност на резултатите от тестове за наличие на зараза се разбира вероятността, с която тестовият резултат съответства на действителното здравословно състояние на тестваното лице, т.е. вероятността лице с положителен тест да е действително заразено, или лице с отрицателен тест действително да не е заразено. Под несигурност на тестовите резултати съответно се разбира вероятността, с която лице с положителен тест в действителност е незаразено или лице с отрицателен тест в действителност е заразено. Както сигурността, така и несигурността на тестовите резултати са вероятности, свързани със съвкупността на всички тествани през определен период, която включва както заразени, така и незаразени лица. За тези вероятности обикновено липсва информация, тъй като те зависят не само от вероятностите за фалшиво поло-

жителен и фалшиво отрицателен тест, но и от други фактори, свързани със състоянието на тестваната съвкупност, които много бързо се променят и не могат лесно да бъдат околичествени.

Относително константни са само вероятностите за фалшиво положителен, съответно фалшиво отрицателен тест, които се отнасят за две различни съвкупности – вероятността за фалшиво положителен тест за съвкупността от незаразени лица, а вероятността за фалшиво отрицателен тест за съвкупността от заразени лица. Ако например вероятността за фалшиво положителен тест възлиза на 1%, това означава, че при 1% от съвкупността на всички незаразени лица тестът ще се окаже положителен. Ако вероятността за фалшиво отрицателен тест възлиза на 5%, това означава, че при 5% от лицата, формиращи съвкупността от заразени, тестът ще се окаже отрицателен. Тези вероятности обуславят несигурността на тестовия резултат, но не са тъждествени с нея. Сигурността, респективно несигурността на тестовите резултати освен това зависи от структурата на тестваната съвкупност, т.е. от относителния дял на заразените в нея, който, от своя страна, не е тъждествен с дела на заразените от цялото население в определен регион. Този относителен дял варира в зависимост от:

1. възприетата политика на тестване;
2. относителната честота на новозаразяване в цялата популация за определен период.

Политика на тестване, при която на тест се подлагат предимно лица със симптоми, за които се предполага, че са заразени, ще доведе до висок дял на заразени от съвкупността на тестваните. Когато обаче се тества повсеместно и отрицателният тест се възприема като „разрешително“ за определени действия и свободи, се очаква делът на заразените в съвкупността на тестваните да бъде приблизително равен на относителната честота на новозаразяване в цялото население за определен период, т.е. той ще бъде многократно по-малък отколкото при първата политика. Колкото е по-малък делът на заразените в съвкупността от тествани, толкова по-малка ще бъде сигурността на положителен резултат от теста, т.е. толкова по-малко вероятно ще бъде лице с положителен тест да е действително заразено. Ако приемем, че новозаразените през дадена седмица са 1% от населението (заболеваемост 1000 на 100 000), като са валидни горепосочените примерни вероятности за фалшиво положителен и фалшиво отрицателен резултат и се тества повсеместно, тогава положителният резултат от тест е по-малко сигурен, отколкото да се падне ези при хвърлянето на монета. Точната сигурност в подобна ситуация възлиза на 48,005%, рискът резултатът да не показва действителното състояние на 51,995%, като информационната стойност на резултата съответства на твърдението: или е заразен, или не е заразен, като двете възможности са почти еднакво вероятни. Тази сигурност е валидна при една много висока

относителна честота на новозаразяване за сравнение (според данни на платформата Statista¹⁾) Чехия, държавата с най-голяма заболяемост от COVID-19 в Европа в началото на 2021 г., към 11.03.2021 г. е със 768,1 новозаразени на 100 000 за последните седем дни, в България за същия период те са 211,1. За Чехия сигурността на положителен резултат при повсеместно тестване би възлизала на 42,375%, а за България – на 16,731%, ако са валидни горепосочените рискове за фалшиво положителен, съответно фалшиво отрицателен резултат. За същия период заболяемостта в Германия е 73,4, което означава, че при повсеместно тестване, каквато вече е политиката в Германия по това време, и горепосочените вероятности за фалшиво положителен и фалшиво отрицателен резултат, вероятността лице с положителен резултат от теста да е действително заразено, би възлизала на 6,52%. Както се вижда, вероятността, положителният резултат да показва действително съществуваща зараза, намалява драстично при по-ниски нива на заболяемост и повсеместно тестване, което означава, че свободите на огромното мнозинство от лица с положителен резултат биха били неоснователно ограничени, тъй като те в действителност не са заразени, в случай че „разрешителното“ за определени действия се дава на базата на само един направен тест. Естествено, всеки би могъл да повтори, потрети и т.н. теста, но до какво би довело всичко това? Гражданите непрекъснато да се тестват, цялото ежедневие да бъде подчинено на тестовите и техните несигурни резултати – излишни разходи, които никоя държава няма да поеме на 100%.

Възниква въпросът защо цялата лавина от информация, свързана с пандемията, съдържа много малко информация относно сигурността, респ. несигурността на тестовите резултати? Една от възможните причини е естеството на тези вероятности, които не са постоянни величини, а зависят в голяма степен от възприетата политика на тестване и от заболяемостта, представляваща брой заразени на 100 000 души от населението за определен период. Целенасоченото тестване на лица със симптоми води до много по-сигурни резултати, отколкото повсеместното тестване на население с ниска заболяемост, като изключително голямата вариация на степента на сигурност/несигурност се дължи на променливостта на структурата на съвкупността от тествани лица. Друга възможна причина се крие в широко разпространени заблуди, свързани с възприемането и интерпретацията на подобни вероятности.

Сигурността и несигурността на тестовия резултат представляват условни вероятности, за определянето на които е необходима априорна информация относно структурата на тестваната съвкупност, която по правило не е известна и се променя непрекъснато.

Под условна вероятност се разбира вероятност за настъпване на събитие А, при условие че е настъпило или ще настъпи събитие В (Tschirk 2014). Тя се изчислява по теоремата на Бейс по следния начин (Hartung 2009; Ross 2006):

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \text{ където:}$$

$P(A)$ – вероятност за настъпване на събитието А;

$P(B|A)$ – условна вероятност за настъпване на В, при положение че събитието А е настъпило;

$P(A|B)$ – условна вероятност за настъпване на А, след като събитието В е настъпило (апостериорна вероятност).

За вероятността, задаваща сигурността на тестовите резултати, теоремата може да бъде записана по следния начин (Gigerenzer 2017):

$$P(\text{болен} | + \text{тест}) = \frac{P(+ \text{тест} | \text{болен})P(\text{болен})}{P(\text{болен})P(+ \text{тест} | \text{болен}) + P(\text{здрав})P(+ \text{тест} | \text{здрав})}$$

където:

$P(\text{болен} | + \text{тест})$ е вероятността при положителен тест лицето да е действително заразено;

$P(+ \text{тест} | \text{болен})$ е вероятността тестовият резултат при заразено лице да е положителен;

$P(\text{болен})$ е относителната честота на заразяване в съвкупността от тествани лица;

$P(\text{здрав})$ е относителната честота на незаразените в съвкупността от тествани лица;

$P(+ \text{тест} | \text{здрав})$ е вероятността тестовият резултат при незаразено лице да е положителен.

Тази вероятност е възможно да бъде изчислена и на база абсолютни честоти на двумерното разпределение, получено чрез групировка на тестваните лица според действителното здравословно състояние и според резултата от теста (Dubben, Beck-Bornholdt 2010).

Резултат от теста	Действително състояние		Общо
	Заразен	Незаразен	
Положителен	a	b	a+b
Отрицателен	c	d	c+d
Общо	a+c	b+d	N=a+b+c+d

Ако с a е означен броят на тестваните лица с положителен тест, които действително са заразени, с b – броят на тестваните лица, които са с положителен резултат от теста, но не са заразени, тогава вероятността при положителен резултат от теста лицето да е действително заразено, възлиза на (Gigerenzer 2017):

$$P(\text{болен} \mid + \text{тест}) = \frac{a}{a + b}$$

Съответно вероятността при положителен тест лицето да не е заразено, ще бъде:

$$P(\text{здрав} \mid + \text{тест}) = \frac{b}{a + b}$$

Според студия на (Gigerenzer & Hoffrage 1995) относно възприятието на условни вероятности при медицински тестове, направена на базата на анкетно проучване, в което респонденти са били 48 лекари от болници в Мюнхен и Дюселдорф, информация, зададена чрез абсолютните честоти на двумерното разпределение, се осмисля и интерпретира много по-правилно, отколкото такава, зададена чрез относителни честоти (вероятности). Защищаваната от (Gigerenzer 2017) теза е, че човешкият разум все още не е в състояние интуитивно да осмисля вероятности, с които обществото е конфронтирано сравнително отскоро, докато без големи проблеми се справя със същата информация, но зададена чрез абсолютни числа. Това означава, че потребителите на информация би следвало да възприемат по-правилно информация, свързана със сигурността на резултатите от тест, когато рисковете за фалшиво положителен и фалшиво отрицателен тест, както и структурата на съвкупността от тествани лица са описани чрез абсолютни величини.

Когато липсва информация за структурата на тестваната съвкупност и се обявяват единствено приблизителните рискове за фалшиво положителен и фалшиво отрицателен тест, съществува опасност потребителите на информация да възприемат тези вероятности директно като мерило за степента на сигурност/несигурност на резултатите от теста. Възможно е да се формира огледален образ на тези рискове, който сугерира висока сигурност за това, лице с положителен тест да е действително заразено, съответно лице с отрицателен тест да не е заразено. При подобно интуитивно обръщане на твърденията се разсъждава по следния начин: след като в 95% от случаите тестът разпознава заразата при заражено лице, тогава при положителен тест лицето със сигурност 95% е заразено, а след като тестът със сигурност 99% разпознава отсъствието на зараза при незаражено лице, тогава при отрицателен тест вероятността лицето да не е заразено, възлиза на 99%. Подобна интерпретация създава усещането за много висока сигурност на тестовите резултати,

оттам се генерира абсолютно доверие в тях, което при повсеместно тестване (включително безсимптомни) може да предизвика неоснователен дискомфорт и ограничаване на действията при незаразени лица с положителен тест. Алогизмът на такава интерпретация на сигурността на тестовите резултати не е съвсем явен, като на пръв поглед обърнатите твърдения изглеждат обосновани и необорими. Обръщането на твърденията относно вероятности за създаване придава привидна тежест на резултата. Несъстоятелността на изводи от този род (Dubben & Beck-Bornholdt 2010) онагледяват с помощта на следния пример: ако дадено същество е човек, то тогава с вероятност 50% то е мъж. Обратно твърдение – ако някой е мъж, то тогава той с вероятност 50% е човек. При този пример не е необходимо да се доказва, че второто твърдение не е вярно, алогизмът е явен. Ясно е, че останалите 50% също принадлежат към човешкия род.

На базата на изследванията на (Gigerenzer 2017) и (Dubben & Beck-Bornholdt 2010), включващи интуитивното възприятие на подобни вероятности, както и на направените от нас разсъждения, предполагаме, че потребителите на информация за резултатите от тестовете за наличие на зараза:

1. се затрудняват повече при осмисляне на информация, зададена с помощта на относителни честоти, отколкото на такава, зададена чрез абсолютни числа;
2. срещат проблеми при осмислянето на данни, свързани със сигурността на тестовите резултати;
3. при интерпретацията на сигурността на тестовите резултати не се съобразяват със структурата на тестваната съвкупност;
4. възприемат сигурността на тестовите резултати като обратна на рисковете за фалшиво положителен и фалшиво отрицателен тест, като игнорират информацията относно дела на заразените в тестваната съвкупност.

2. Възприятие на потребителите на информация за сигурността на тестовите резултати

За проверка на направените предположения относно интуитивното възприятие от страна на потребителите на информация за вероятности, свързани със сигурността на тестове за наличие на зараза, е осъществено непредставително проучване чрез анкета, в която са имали възможност да участват лица без ограничения, свързани с пол, образователна степен, професия или занимание, възраст. Проучването е проектирано и проведено онлайн чрез платформата Google Forms, като е активирана опцията за ограничаване на възможния брой отговори до един. Линк към анкетата е разпространен чрез социалната мрежа Facebook. Броят на лицата, участвали в проучването по метода на отзовалите се, е 102. Респондентите са информирани предвари-

телно, че се очакват освен точни и интуитивни отговори на зададените въпроси, като за по-висока степен на достоверност на резултатите са помолени да не търсят чужда помощ. Зададени са общо 9 въпроса, от които първите 5 включват казуси относно сигурността на тестовите резултати (таблица 1), а останалите четири са свързани с възрастовата група, пола, завършената степен на образование и професията. Петте въпроса, включващи казуси относно сигурността на тестови резултати, логически са свързани помежду си по следната схема.

1. При казуси 1, 2, 3 и 5 се предполага, че рискът за фалшиво положителен тест възлиза на 1%, а за фалшиво отрицателен – на 5%, казус 4 се различава по риска за фалшиво отрицателен тест, който е зададен в размер също на 1%.
2. В два от казусите (1 и 2) е дадена вероятността за положителен тест при заразено лице, съответно вероятността за отрицателен при незаразено лице, а в други два (3 и 4) – вероятността за отрицателен тест при заразено лице, съответно вероятността за положителен тест при незаразено лице.
3. В един казус (5) информацията относно риска за фалшиво положителен, съответно отрицателен тест е зададена чрез абсолютни числа.
4. Първи и втори казус се различават съществено по структурата на тестваната съвкупност – 10%, съответно 0,1% заразени.
5. Втори и трети казус са взаимно допълващи се при еднаква изходна ситуация: в казус 2 се търси вероятността, с която при положителен тест лицето действително е заразено, а в казус 3 – вероятността, с която при положителен тест лицето не е заразено, т.е. сумата от двата отговора възлиза на 100%.
6. Втори и пети въпрос включват идентични казуси, като информацията е поднесена по различен начин. Казус 2 съдържа информация под формата на вероятности, а казус 5 – под формата на абсолютни числа. Освен това в казус две са дадени вероятностите за положителен тест при заразено, съответно отрицателен тест при незаразено лице, а в казус две е включена информация относно риска за фалшиво положителен и фалшиво отрицателен тест.
7. Казус 4 не съдържа информация относно структурата на тестваната съвкупност, което означава, че не може да бъде установена сигурността/несигурността на тестовите резултати и тя е между 0 и 100%.

Идентичните по съдържание втори и пети въпрос имат за цел проверката на предположението, според което потребителите на информация се затрудняват повече при осмислянето на сигурността на тестовите резултати, когато информацията е зададена с помощта на относителни честоти, отколкото ако тя е под формата на абсолютни числа.

Логическата връзка между въпросите дава възможност да се провери каква част от респондентите осмисля ситуациите, включени в отделните казуси и какъв е делът на тези, които не разбират логиката на зададените въпроси.

Комбинацията от отговорите на всички въпроси може да покаже дали при интерпретацията анкетираните се съобразяват със структурата на тестваната съвкупност.

Същата комбинация от отговори може да помогне за обосновка на направеното предположение, според което сигурността на тестовите резултати се възприема от потребителите на информация като обратна на рисковете за фалшиво положителен и фалшиво отрицателен тест.

Като допълнение към анализа, свързан с направените четири предположения, с помощта на събраната информация относно пола, вида и степента на образование на анкетираните лица е възможно да се проучи дали биха били логически обосновани предположения относно наличието на зависимост между възприятието за сигурност/несигурност на тестовите резултати и пола и/или образователната степен.

Тъй като проучването е непредставително, то не може да послужи за проверка на направените предположения с помощта на инструментите на теорията на статистическите заключения. Резултатите биха могли да бъдат в подкрепа или да противоречат на формулираните хипотези, но не могат да доведат до приемането или отхвърлянето им.

Таблица 1. Въпроси-казуси и резултати от проучването

Въпрос №	Отговор №	Въпроси/ отговори	Отговорили	
			Брой	%
1.		Вероятността за положителен PCR тест при заражено лице възлиза на 95%, а вероятността за отрицателен тест при незаражено лице – 99%. Каква според Вас е вероятността при положителен тест лицето действително да е заражено, ако относителният дял на заразените възлиза на 10% от всички тествани? (точен отговор: 91,35%)	102	100
	1.	До 10%	19	18,6
	2.	Над 10 до 90%	30	29,4
	3.	Над 90%	53	52,0
2.		Вероятността за отрицателен PCR тест при незаражено лице възлиза на 99%, а вероятността за положителен тест при заражено лице – на 95%. Каква според Вас е вероятността при положителен тест лицето действително да е заражено, ако относителният дял на заразените в съвкупността от тествани лица е 0,1%? (точен отговор: 8,68%)	102	100
	1.	До 10%	36	35,3
	2.	Над 10 до 90%	18	17,6
	3.	Над 90%	48	47,1

3.		Вероятността за отрицателен PCR тест при заражено лице възлиза на 5%, а вероятността за положителен тест при незаражено лице – на 1%. Каква според Вас е вероятността при положителен тест лицето да не е заражено, ако относителният дял на незаразените възлиза на 99,9% от всички тествани? (точен отговор: 91,32%)	102	100
	1.	До 10%	69	67,6
	2.	Над 10 до 90%	14	13,7
	3.	Над 90%	19	18,6
4.		Вероятността за отрицателен PCR тест при заражено лице възлиза на 1%, а вероятността за положителен тест при незаражено лице – също на 1%. Каква според Вас е вероятността при положителен тест лицето да е действително заражено? (Коректен отговор: между 0 и 100% в зависимост от структурата на тестваната съвкупност)	102	100
	1.	До 10%	13	12,7
	2.	Над 10 до 90%	11	10,8
	3.	Над 90%	59	57,8
	4.	Над 0 до 100%	13	12,7
	5.	0	6	5,9
5.		Тествани са 100 000 души, от които действително заразени са 100. PCR тест дава фалшиво отрицателен резултат при всеки двадесети заразен и фалшиво положителен резултат при 1 от 100 незаразени. Каква според Вас е вероятността при положителен тест лицето действително да е заражено? (точен отговор: 8,68%)	102	100
	1.	До 10%	27	26,5
	2.	Над 10 до 90%	36	35,3
	3.	Над 90%	39	38,2

Във връзка с първото предположение, според което потребителите на информация се затрудняват повече, когато тя е зададена с помощта на относителни честоти, отколкото когато се работи с абсолютни числа, използваме информацията за твържденията по логическо съдържание втори и пети въпрос. Прави впечатление, че делът на отговорилите вярно на втори въпрос, в който информацията е зададена чрез вероятности, е с 8,8 процентни пункта по-голям от този на отговорилите вярно на пети въпрос, съдържащ твърствен казус, зададен чрез абсолютни числа (таблица 1). Това означава, че резултатите от проучването влизат в противоречие с предположението, според което потребителите на информация по-лесно осмислят абсолютни числа в сравнение с информация, съдържаща относителни честоти и вероятности.

Във връзка с второто предположение, според което потребителите на информация срещат проблеми при осмислянето на данни, свързани със сигурността на тестовите резултати, подлагаме на анализ комбинациите от отговори на втори, трети и пети въпрос на отделните респонденти в анкетата.

Относителният дял на отговорилите вярно на втори въпрос е 35,3%, на трети – 18,6%, а на пети – 26,5%, което означава, че мнозинството не е дало верен отговор на нито един от трите въпроса. Ако отговорите на трите въпроса бъдат разгледани независимо един от друг, все пак може да се твърди, че делът на отговорилите правилно не е прекалено малък, но когато се наблюдават честотите на възможните комбинации от отговори чрез формиране на двумерни разпределения, се разкрива коренно различна картина, която може да послужи за подкрепа на второто предположение.

Чрез двумерните групировки може да се проследи каква част от анкетираните открива логическа връзка между взаимно допълващи се казуси и колко от тях са открили аналогията на втори и пети казус.

Таблица 2. Разпределение на анкетираните според отговорите на втори и трети въпрос

Отговори на въпрос 2	Отговори на въпрос 3			Общо
	До 10%	Над 10 до 90%	Над 90%	
До 10%	19	5	12	36
Над 10 до 90%	12	5	1	18
Над 90%	37	4	7	48
Общо	68	14	20	102

От таблица 2 се вижда, че 12 участници, т.е. 11,7% от анкетираните, са отговорили вярно едновременно на втори и трети въпрос и може да се предполага, че са открили логическата връзка между тях. Сред отговорилите вярно на тази двойка въпроси има 3 участници със средно образование, 2 са със степен „бакалавър“, 6 са със степен „магистър“ и 1 участник е с научна степен „доктор“. Структурата на далите верни отговори на тази двойка въпроси според образователната им степен е следната:

Таблица 3. Структура според степента на образование на отговорилите вярно на втори и трети въпрос

(%)

Образователна степен	Относителен дял от отговорилите вярно	Относителен дял от всички анкетирувани
Средно	25	23,5
Бакалавър	16,7	24,5
Магистър	50	44,1
Доктор	8,3	7,8
Общо	100	100

Структурата на вярно отговорилите участници според тяхната образователна степен почти съвпада със структурата на всички участници по признака „степен на образование“. Този резултат не позволява да се направи предположение за влияние на образователната степен върху възможността за откриване на логическата връзка между наблюдаваната двойка въпроси. Интерес би представлявала и областта на образование. Логично би било да се предположи, че правилно възприемане на въпросите ще се формира сред лекари и сред лица с образование в областта на точните науки, за които е характерно по-усилено изучаване на математика и статистика, като математици, статистици, инженери и преподаватели в тези области. За съжаление, съвкупността на участниците в анкетата е много разнородна според областта на образование и това не позволява да се направи извод по отношение на направеното предположение. Анализът в тази посока допълнително се затруднява поради това, че отговорите на въпроса, свързан с професионалната област, са със свободен текст и не всички отговори са изчерпателни, което не позволява да се извърши необходимата групировка. Сред вярно отговорилите на въпроси 2 и 3 има 11 жени и 1 мъж. В процентно изражение това са 91,7% жени и 8,3% мъже. Структурата на всички анкетирувани по пол е следната: 80,4 % жени и 19,6% мъже. На базата на сравняването на двете структури би могло да се направи предположение, че сред анкетираните откриването на логическата връзка между двата въпроса се е удало по-добре на жените.

Въпрос 5 и въпрос 2 съдържат еднакви по логика казуси, но докато във въпрос 2 информацията е предоставена чрез относителни дялове, във въпрос 5 е чрез абсолютни честоти. Логическата връзка между въпроси 3 и 5 е същата както при въпроси 3 и 2.

Таблица 4. Разпределение на анкетираните според отговорите на пети и трети въпрос

(брой)

Отговори на въпрос 5	Отговори на въпрос 3			Общо
	До 10%	Над 10 до 90%	Над 90%	
До 10%	18	2	7	27
Над 10 до 90%	15	11	10	36
Над 90%	35	1	3	39
Общо	68	14	20	102

Само 7 участници, т.е. 6,9% от анкетираните, са отговорили вярно едновременно на пети и трети въпрос (таблица 4). Откриването на логическата връзка между тази двойка въпроси може би затруднява допълнително участниците поради различния начин на предоставяне на информацията (чрез относителни дялове и абсолютни честоти). Структурата на вярно отговорилите на тази двойка въпроси според образователната степен е следната:

Таблица 5. Структура според степента на образование на отговорилите вярно на пети и трети въпрос

(%)

Образователна степен	Относителен дял от отговорилите вярно	Относителен дял от всички анкетирани
Средно	14,3	23,5
Бакалавър	14,3	24,5
Магистър	57,1	44,1
Доктор	14,3	7,8
Общо	100	100

Сравняването на двете структури разкрива, че в по-голяма степен откриването на логическата връзка между двата въпроса е характерно за участниците с по-високо образование. Като цяло, различието в начина на представяне на информацията в двата въпроса е затруднило участниците и това води до по-малко верни отговори на тази двойка въпроси. Но сред вярно отговорилите имат по-висок относителен дял тези с по-високо образование. От вярно отговорилите 43% са лекари, статистици и инженери, което в някаква степен е потвърждение на формулираното по-рано очакване за влияние на професионалната област върху способността за коректно възприемане на предоставената информация и за откриване на логическа връзка между въпросите. Разпределението на вярно отговорилите по пол е 85,7% жени и 14,3% мъже, което е почти аналогично на разпределението на всички анкетирани по пол и това не

позволява да се направи извод за влиянието на този признак върху способността за откриване на логическата връзка при тази двойка въпроси.

Таблица 6. Разпределение на анкетираните според отговорите на втори и пети въпрос

(брой)

Отговори на въпрос 2	Отговори на въпрос 5			Общо
	До 10%	Над 10 до 90%	Над 90%	
До 10%	18	12	6	36
Над 10 до 90%	4	12	2	18
Над 90%	5	12	31	48
Общо	27	36	39	102

18 участници, т.е. 17,6% от анкетираните, са отговорили вярно на двата тъждествени въпроса (таблица 6). Общо са 61, т.е. 59,8% тези, които са дали еднакви отговори (независимо верни или неверни) на двата въпроса и за които може да се предполага, че са открили тъждеството на казусите.

Разпределението на вярно отговорилите по пол е 94,4% жени и 5,6% мъже, което дава основание да се предположи, че жените по-лесно са възприели двата въпроса като тъждествени. Разпределението на вярно отговорилите според образователната степен е следното:

Таблица 7. Структура според степента на образование на отговорилите вярно на втори и пети въпрос

(%)

Образователна степен	Относителен дял от отговорилите вярно	Относителен дял от всички анкетирани
Средно	33,3	23,5
Бакалавър	11,1	24,5
Магистър	50	44,1
Доктор	5,6	7,8
Общо	100	100

Най-голямо различие между двете структури се наблюдава по отношение на групите на анкетираните със средно образование и със степен „бакалавър“, което не позволява да се направи предположение за влияние на образователната степен върху способността на възприемане на двата въпроса като тъждествени. По отношение на различния начин на задаване на информацията – чрез относителни дялове и чрез абсолютни честоти може да се предположи, че казусът с информация чрез относителни величини е бил по-разбираем

за анкетираните, защото на този въпрос верните отговори са 35,3%, докато при казуса с информация чрез абсолютни честоти верните отговори са 26,5%. Възможно е, разбира се, да има и известно „налучкване“ на верния отговор, което няма как да бъде анализирано.

Таблица 8. Разпределение на анкетираните с еднакви отговори на втори и пети въпрос според отговорите на трети въпрос

(брой)

Отговори на въпрос 3	Отговори на въпроси 2 и 5		
	До 10%/до 10%	Над 10 до 90%/над 10 до 90%	Над 90%/над 90%
До 10%	11	6	28
Над 10 до 90%	1	5	1
Над 90%	6	1	2
Общо	18	12	31

Анкетираните, дали еднакъв отговор на втори и пети въпрос, за които има основание да се предполага, че са осмислили твърдеството на двата въпроса, са общо 51, т.е. 50% от всички (таблица 8). Част от тях не открива логическата връзка с трети въпрос. Интересно е, че тези, които са дали еднакъв неверен отговор (вероятност над 90%) на твърдествените въпроси, преобладаващо (28 от 31) са осъзнали, че трети въпрос е противоположен на втори и пети и отговорът е разликата до 100%. Изключително малко са тези, открили логическата връзка между втори, трети и пети въпрос и съответно дали верен отговор едновременно на трите въпроса. От 18 отговорили вярно на втори и пети въпрос само 6 са осмислили логическата връзка между трите въпроса и са дали верен отговор и на трети въпрос. Тези 6 души представляват едва 5,9% от всички анкетирани. За общо 34 респонденти (33,3% от анкетираните) може да се предполага, независимо дали отговорите им са верни или неверни, че са открили логическата връзка между втори, трети и пети въпрос, като отговорът на трети въпрос представлява 100% минус отговора на втори, респ. пети въпрос.

Във връзка с третото предположение, според което потребителите на информация не се съобразяват с дела на заразените в тестваната съвкупност при преценка на сигурността на тестовите резултати, е възможно да бъдат привлечени като доказателство комбинациите от възможни отговори и съответните честоти на двумерните групировки според отговорите на въпрос 1 и въпрос 2, респективно въпрос 1 и въпрос 5. Тъй като казусът във въпрос 1 се различава от тези във въпросите 2 и 5 единствено по структурата на тестваната съвкупност, еднаквите отговори биха означавали, че респондентите не вземат под внимание голямата разлика в дела на заразените в тестваната съвкупност.

Таблица 9. Разпределение на анкетираните според отговорите на първи и втори въпрос

(брой)

Отговори на въпрос 1	Отговори на въпрос 2			Общо
	До 10%	Над 10 до 90%	Над 90%	
До 10%	16	2	1	19
Над 10 до 90%	12	12	6	30
Над 90%	8	4	41	53
Общо	36	18	48	102

Двумерната групировка по отговорите на въпрос 1 и въпрос 2 показва, че 69 от анкетираните (67,6%) са дали еднакъв отговор на двата въпроса, т.е. са преценили, че няма разлика в изходната ситуация (таблица 9). Само 8 души (7,8%) са дали верен отговор едновременно на двата въпроса. Всички отговорили вярно на тези два въпроса са жени. Структурата на вярно отговорилите според образованието е поместена в таблица 10.

Таблица 10. Структура според степента на образование на отговорилите вярно на първи и втори въпрос

(%)

Образователна степен	Относителен дял от отговорилите вярно	Относителен дял от всички анкетирани
Средно	12,5	23,5
Бакалавър	25	24,5
Магистър	50	44,1
Доктор	12,5	7,8
Общо	100	100

Сравняването на двете структури позволява да се направи предположение, че лицата с висше образование по-добре са се ориентирали и са отчели влиянието на относителния дял на заразените в структурата на изследваната съвкупност. Сред далите верен отговор преобладават лицата, които са посочили, че според професионалната област са статистици.

Таблица 11. Разпределение на анкетираните според отговорите на първи и пети въпрос

(брой)

Отговори на въпрос 1	Отговори на въпрос 5			Общо
	До 10%	Над 10 до 90%	Над 90%	
До 10%	11	6	2	19
Над 10 до 90%	11	16	3	30
Над 90%	5	14	34	53
Общо	27	36	39	102

Двумерната групировка по отговорите на въпрос 1 и въпрос 5 показва, че 61 от анкетираните (59%) са отговорили по един и същ начин на двата въпроса, а само 5 анкетирани (4,9%) са дали верен отговор едновременно на двата въпроса (таблица 11). Въпроси 2 и 5 са твърждествени, но информацията в тях се задава по различен начин – чрез относителни дялове и чрез абсолютни честоти. Верните отговори на тази двойка въпроси (1 и 5), както и при предходната (1 и 2), може да се тълкува като отчитане ролята на относителния дял на заразените в структурата на тестваните лица. Тъй като при тази двойка въпроси верните отговори са по-малко, то едно възможно обяснение би могло да бъде това, че възприятието и осмислянето на въпросите се затруднява, когато информацията е представена по различен начин. От вярно отговорилите на тази двойка въпроси 80% са жени и 20% са мъже, като това разпределение съответства на разпределението по пол на всички анкетирани. Структурата на вярно отговорилите според образователната степен е поместена в таблица 12.

Таблица 12. Структура според степента на образование на отговорилите вярно на първи и пети въпрос

(%)

Образователна степен	Относителен дял от отговорилите вярно	Относителен дял от всички анкетирани
Средно	20	23,5
Бакалавър	20	24,5
Магистър	40	44,1
Доктор	20	7,8
Общо	100	100

Сравняването на двете структури показва, че в случаите, когато информацията в казусите се задава по различен начин – чрез относителни дялове и чрез абсолютни честоти, възприемането и разкриването на влиянието на относителния дял на заразените в съвкупността на тестваните лица затруднява анкетираните. Най-голямо различие между двете структури се наблюдава при групата на анкетираните със степен „доктор“, което е предпоставка за определяне на образованието като фактор за правилното възприятие и осмисляне на казуса.

Таблица 13. Разпределение на анкетираните, отговорили правилно на втори и пети въпрос, според отговорите на първи въпрос

(брой)

Отговори на въпрос 1	Отговорили правилно на въпроси 2 и 5
До 10%	9
Над 10 до 90%	6
Над 90%	3
Общо	18

Разпределението на отговорилите правилно едновременно на втори и пети въпрос според отговорите на първи въпрос показва, че от общо 18 души само 3 са се справили едновременно с трите въпроса (таблица 13). И трите верни отговора са дадени от жени със степен „магистър“ и „доктор“.

Получените резултати са в подкрепа на предположението, според което потребителите на информация не вземат под внимание структурата на тестваната съвкупност, когато преценяват сигурността на тестовите резултати.

В подкрепа на четвъртото предположение, според което сигурността на тестовите резултати се възприема от потребителите на информация като обратна на рисковете за фалшиво положителен и за фалшиво отрицателен тест, игнорирайки информацията относно дела на заразените в тестваната съвкупност, могат да послужат резултатите от петмерната групировка според възможните отговори на петте въпроса.

Една комбинация от отговори (първи въпрос – до 90%; втори въпрос – до 90%; трети въпрос – до 10%; четвърти въпрос – до 90%; пети въпрос – до 90%) може да се тълкува като осмисляне на логиката на казусите, но поставяне на сигурността на тестовия резултат в пълна зависимост с рисковете за фалшиво положителен и за фалшиво отрицателен тест, като делът на заразените в тестваната съвкупност не играе никаква роля. 26 от респондентите са отговорили еднотипно по тази схема, която предполага, че са осмислили ситуациите и логическата връзка между тях, но въпреки всичко не са успели да преценят правилно границите, в които се движи търсената вероятност.

Единственият верен отговор е на първи въпрос (над 90%), при който заради големия дял (10%) заразени в тестваната съвкупност вероятността при положителен тест лицето действително да е заразено, е висока и попада в границите, в които се движат вероятностите за положителен тест при заразено лице и за отрицателен тест при незаразено лице.

Тези участници в анкетата дават еднакъв неверен отговор на втори и пети въпрос (над 90%), като това може да означава, че са установили тъждеството на двата казуса. Отговорът на трети въпрос е вероятност до 10%, което означава, че респондентите са разбрали, че изходните ситуации в трите въпроса са еднакви, като търсената вероятност допълва до 100% отговора на втори, респ. пети въпрос.

Отговорът на четвърти въпрос, при който казусът не позволява изчисляването на точна вероятност заради липсващата информация за структурата на тестваната съвкупност, съответства на отговорите на първи, втори и пети въпрос, т.е. вероятност над 90%.

За тези 26 участници в анкетата може да се твърди, че използват като мерило за сигурността/несигурността на тестовия резултат вероятностите за фалшиво положителен и фалшиво отрицателен резултат, отнасящи се съответно за съвкупността на незаразените и за съвкупността на заразените лица, без да се съобра-

заяват със структурата на тестваните лица. Неверните и еднотипни отговори са огледален образ на рисковете за фалшиво положителен и фалшиво отрицателен резултат, като това съответства на направеното предположение, според което сигурността на тестовите резултати се възприема от потребителите на информация като обратна на тези рискове, като се игнорира информацията относно дела на заразените в тестваната съвкупност. В подкрепа на направеното предположение е и обстоятелството, че тези 26 анкетирувани, отговорили по един и същ начин, представляват абсолютната честота на една от общо 405 възможни комбинации от отговори на петте въпроса. Ясно е, че при 102 респонденти повечето от комбинациите на петмерната групировка са с честота, равна на 0, като тази комбинация е модална с внушителен превес на броя на отговорилите в сравнение с всички останали. В противовес комбинацията от петте верни отговора е с честота 0. Нито един респондент не е отговорил правилно на всички въпроси от 1 до 5.

Структурата на тези 26 участници според образователната степен е поместена в таблица 14.

Таблица 14. Структура на анкетираните, формиращи честотата на модалната комбинация от отговори на петмерната групировка според степента на образование

Образователна степен	(%)	
	Относителен дял от 26-те анкетирувани	Относителен дял от всички анкетирувани
Средно	7,7	23,5
Бакалавър	19,2	24,5
Магистър	57,7	44,1
Доктор	15,4	7,8
Общо	100	100

Това разпределение показва, че отговорилите с тази комбинация от отговори са предимно с по-висока степен на образование. Независимо че не са дали верни отговори, те може би са успели да открият логическата връзка между въпросите – взаимно допълващи се и аналогични. В същото време, не отчитат влиянието на относителния дял на заразените в структурата на съвкупността. Тяхното разпределение по пол е 46,1% мъже и 53,9% жени. Тъй като тази структура значително се различава от структурата по пол на всички анкетирувани (19,6% мъже и 80,4% жени), то може да се предположи, че откриването на логическата връзка между казусите се е удало в по-голяма степен на мъжете, след като 60% от анкетираните мъже са отговорили точно по този начин. Сред тези 26 участници 35% са посочили като своя професия „статистик“, „инженер“ и „лекар“; 27% са се определили като „икономисти“. Това разпределение показва, че участници с образование, включващо по-усилено обучение по статистика и математика, в по-голяма степен

се ориентират в представените ситуации и успяват да открият логическите връзки между тях, независимо че не успяват да отговорят вярно.

На базата на осъществения анализ на резултатите от проведеното непредставително анкетно проучване могат да бъдат направени следните изводи по отношение на формулираните предварително предположения и очаквания.

1. Резултатите от проучването влизат в противоречие с предположението, според което информация, представена чрез абсолютни числа, се възприема по-лесно от участниците и води до правилно осмисляне на казуса.
2. Дадените от респондентите отговори показват, че откриването на логическа връзка между два казуса е по-лесно в случаите, когато информацията в тях се представя по един и същи начин – чрез относителни честоти.
3. Осмислянето на връзката между казусите и даването на верни отговори се е удало в по-голяма степен на лицата с по-високо образование.
4. Твърде голямата вариация по признака „професионална област на участниците“ не позволява да се формулира хипотеза по отношение на влиянието на този признак върху отговорите, но все пак беше направено едно предположение за по-лесно осмисляне на казусите от лицата, чието обучение е включвало по-задълбочена подготовка по „статистика“ и „математика“.
5. Резултатите от анкетното проучване са в подкрепа на хипотезата, че при осмисляне на казусите участниците не отчитат влиянието на структурата на съвкупността.
6. На базата на дадените верни отговори може да се твърди, че отчитането на влиянието на структурата на съвкупността се е удало само на жени.
7. Дадените отговори подкрепят хипотезата, че лицата срещат затруднения при осмисляне на данни, свързани със сигурността на резултата.
8. Основателна е хипотезата, че участниците възприемат сигурността на тестовите резултати като обратна на рисковете за фалшиво положителен и фалшиво отрицателен тест. Според резултатите от проучването това е валидно за участници с различни степени на образование и различна професионална област.
9. Влиянието на признака „пол“ е трудно да бъде измерено, но при петмерната групировка на дадените отговори се откроява комбинация с голяма абсолютна честота, която е характерна за повече от половината участници мъже. Това позволява да се направи предположение, че осмислянето на логическата връзка между казусите се е удало в по-голяма степен на мъжете, отколкото на жените, но точно тези мъже приемат сигурността на тестовите резултати като обратна на рисковете за фалшиво положителен и фалшиво отрицателен резултат.

Заклучение

Изправени пред непозната заплаха, хората търсят допълнителна информация, която да им даде увереност и надежда. Начинът, по който се поднася информацията, е от голямо значение, особено когато се отнася за непозната/непривична област. По-голямата част от потребителите на информация нямат необходимата подготовка, за да осмислят и възприемат правилно данните, отнасящи се до сигурността на резултатите от тестове за наличие на зараза. Невъзможността за правилна интерпретация на информацията, особено тази, която е свързана със сигурността на резултата, води до числова слепота и замъглено мислене. Това, от своя страна, е предпоставка както за поява на силно преувеличено усещане за сигурност, така и за отрицание на получаваната официална информация.

При поднасяне на информация за използваните тестове за наличие на зараза и тяхната сигурност би следвало да се има предвид, че по-голямата част от потребителите й (независимо от образованието си) не са в състояние да осмислят представените данни, както и да отчетат влиянието на структурата на съвкупността от тествани лица. Би трябвало да се има предвид, че различният начин на поднасяне на данните – чрез абсолютни величини и чрез относителни дялове, също затруднява потребителите при осмисляне на тези данни и интерпретацията им. Резултатите от проучването обаче, влизат в противоречие с хипотезата за по-лесно възприемане на абсолютните честоти. Възприетият подход за представяне сигурността на използваните тестове може да се счита за „познат“ в някаква степен само за лица с по-добра подготовка по статистика и математика или за лица с медицинско образование.

Като извод се оформя мнението, че е необходимо да се имат предвид направените заключения при съставяне на учебните програми по математика и статистика както за ученици, така и за студенти с цел ограничаване на възможността за формиране на числова слепота при конфронтация с подобна информация.

БЕЛЕЖКИ

1. <https://de.statista.com/statistik/daten/studie/1180169/umfrage/laender-mit-den-meisten-coronainfektionen-in-der-letzten-woche-in-europa/>

REFERENCES

- Dubben, H. H., Beck-Bernholdt, H. P., 2010. *Der Hund, der Eier legt. Erkennen von Fehlinformation durch Querdenken*. Reinbeck bei Hamburg: Rowohlt Taschenbuch Verlag.
- Gigerenzer, G., 2017. *Das Einmaleins der Skepsis. Über den richtigen Umgang mit Zahlen und Risiken*. München: Piper Verlag.

- Gigerenzer, G., Hoffrage, U., 1995. How to Improve Bayesian Reasoning without Instruction: Frequency Formats. *Psychological Review*, 4(102), 684 – 704.
- Hartung, J., 2009. *Statistik. Lehr- und Handbuch der angewandten Statistik*. München: Oldenbourg Verlag.
- Ross, Sh. M., 2006. *Statistik für Ingenieure und Naturwissenschaftler*. München: Spektrum Akademischer Verlag.
- Tschirk, W., 2014. *Statistik: Klassisch oder Bayes. Zwei wege im Vergleich*. Berlin: Springer Verlag.

CROWDED PROBABILITIES IN TESTS FOR THE PRESENCE OF INFECTION

Abstract. In the presentation, problems have been addressed in perception and popular delusions in the interpretation of information related to the level of confidence / uncertainty of tests by which the presence or absence of contamination is verified. The theoretical essence of the probabilities of a false positive and false negative result, as well as the level of confidence / uncertainty of the test results and on this basis there are assumptions about the perception of such information and possible misconceptions in its interpretation by society. Assumptions made are checked on the information based on a non-reporting survey. The results obtained reveal user difficulties in rationalizing and interpreting the data provided as well as when detecting logical links characteristic of such information.

Keywords: conditional probabilities; Bayes theorem; level of confidence; test for the presence of infection

✉ **Dr. Margarita Lambova, Assoc. Prof.**

ORCID iD: 0000-0003-1865-8962

University of Economics – Varna

77, Knyaz Boris I Blvd.

9002 Varna, Bulgaria

E-mail: lambowa@ue-varna.bg

✉ **Dr. Vanya Stoyanova, Assist. Prof.**

ORCID iD: 0000-0001-6304-7824

University of Economics – Varna

77, Knyaz Boris I Blvd.

9002 Varna, Bulgaria

E-mail: vstoyanova@ue-varna.bg