УДК 519.25, 378.146

Шарапов М.М., к.ф.-м. н., доцент

Статистичне корегування результатів тестів

Київський національний університет імені Тараса Шевченка, 83000, м. Київ, пр-т. Глушкова 4д,
e-mail: sharapov@unicyb.kiev.ua.

M.M. Sharapov, Cadnd.Sci.(Phys.-Math.)

Statistical correction of test results

Taras Shevchenko National University of Kyiv, 83000, Kyiv, Glushkova st., 4d,
e-mail: sharapov@unicyb.kiev.ua.

В статті розглядається статистичне оцінювання кількості тих питань тестів, на які було свідомо дано правильні відповіді. Також пропонується новий підхід до тестування, за якого опитуваний може явно вказувати на своє незнання відповіді. Показано, що запропонований підхід не змінює кількісне значення статистичної оцінки, але дозволяє ввести до розгляду якісно нові показники оцінювання знань.

Ключові слова: оцінювання тестів, свідомі відповіді, статистичне корегування, нові показники якості знань.

Here we consider a statistical estimation of a number of consciously given correct answers during testing. If a testee doesn't know a correct answer in the test then he can try to guess the correct variant. The intuitive approach can be used to estimate the number of non-guessed correct answers while statistical approach can be used too. Furthermore, the statistical approach gives a higher estimated grade. A new approach to testing is considered as well. Under new conditions a student has opportunity to mark the question as unbeknown without trying to guess the correct answer. It is shown that new method doesn't change the estimation of consciously given correct answers under statistical approach but makes it possible to introduce new ratings for knowledge estimation. Moreover, the introduced sub-indicator of knowledge strength makes it possible to build new grades, estimation and indicators using the described technique.

Key Words: tests assessment, consciously given correct answers, statistical correction, new ratings for knowledge estimation.

Статтю представив д. т. н., професор Гаращенко Ф.Г.

The main questions we are going to deal with are "What is statistical correction of test results (SCTR)?" and "Why/when do we need it?". Well, the SCTR is a method (algorithm) to make test results (TR) closer to a real skill level (SL) of a student (testee) because sometimes TR and SL can differ seriously. So we do need SCTR to decrease this difference. Suppose the test consists of 40 questions with 4 variants of answer each while student gave 28 correct answers and 12 incorrect ones. It is logically to suppose that he guessed some answers when he didn't know the right answer. So, the real result should be 24 instead of 28.

The aspects that should be taken onto account:

(A) The purpose of test.
(B) The kind of the test.

(A) If the purpose is just a ranking then we do not need a SCTR because it will not change the result (ranks of students). But if the purpose is to obtain the SL over some standard scale then the SCTR is needed.

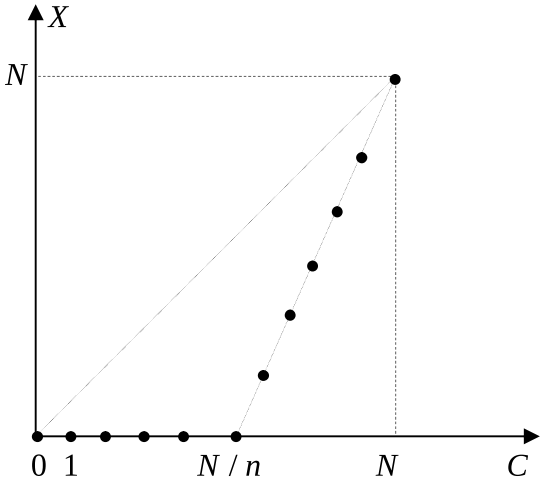
(B) There are many different kinds of tests and we will start with the simplest ones when test consists questions under three conditions:

- **(i)** All questions have the same difficulty level.
- **(ii)** All questions have the same number of variants of answer.
- **(iii)** All variants of answers have the same (equal) probability to be guessed.

Under conditions (i)-(iii) it is easy to generalize the result of mentioned example: if test consists of N questions with $n > 1$ variants of answer each while student gave C correct answers and I incorrect ones ($I + C = N$) then the corrected result should be $X = C - Y$ where Y is a number of guessed answers. Well, $X \le C$ means that the corrected result supposed to be less or equal to C.

We state that X could be obtained (estimated) as follows

$$X_0=\begin{cases}\dfrac{Cn-N}{n-1}, & \dfrac{N}{n}<C\le N\\ 0, & C\le\dfrac{N}{n}\end{cases}. \quad (1)$$



In troth, the average number of guessed answers should be $Y=\dfrac{N-X}{n}$ (each n-th of all unknown ones) and $X=C-Y=C-\dfrac{N}{n}+\dfrac{X}{n}$ yield (1). By the way, the case $X=0$ $\left(C\le\dfrac{N}{n}\right)$ means that unfortunately student couldn't even guess the sufficient answers.

Now let's try to use common probabilistic approach to obtain the desired estimation for X. Let $P(C\,|\,X)$ be a probability to give C correct answers while X of them have been guessed ($C\ge X$). Then

$$P(C\,|\,X)=P\{\text{guessed}\,Y-C-X\text{ of }N-X\}=$$
$$=C_{N-X}^{Y}\,p^{Y}(1-p)^{N-X-Y}$$

is nothing but binomial probability where $p=\dfrac{1}{n}$ is a probability of success and $C_a^b=\dfrac{a!}{b!(a-b)!}$ is a binomial coefficient. In order to obtain the argmaximum (X) for $P(C\,|\,X)$ we'll deal with the inequality

$$\frac{P(C\,|\,X+1)}{P(C\,|\,X)}=\frac{N-C}{(N-X)(1-p)}\ge 1 \qquad (2)$$

because for the corresponding values of X $P(C\,|\,X+1)\ge P(C\,|\,X)$. Inequality (2) has a plane solution $X\le\dfrac{Cn-N}{n-1}$ so the desired argmaximum (see the numerator of LHS at (2)) is $X+1=\dfrac{Cn-N}{n-1}+1$. So, we get the estimation

$$X_0'=X_0+1=\begin{cases}\dfrac{Cn-N}{n-1}+1, & \dfrac{N}{n}<C\le N\\ 0, & C\le\dfrac{N}{n}\end{cases} \qquad (3)$$

and it differs from (1). Why? What's the problem? Well, let's try to substitute X_0' into (2) and we'll see that

$$\frac{P(C\,|\,X_0'+1)}{P(C\,|\,X_0')}=\frac{N-C}{(N-X_0')(1-p)}=1.$$

This equality means that both $X_0=\dfrac{Cn-N}{n-1}$ and $X_0'=X_0+1=\dfrac{Cn-N}{n-1}+1$ give the same maximum for $P(C\,|\,X)$. SCTR decreases amount of earned points so it'll be honestly to deal with X_0' instead of X_0. For example, if $N=12$, $n=4$, $C=6$ then $X_0=4$ but we'll propose $X_0'=5$ instead (we're going to steal only one point instead of two).

The next interesting idea is to add to every question one more variant of answer – "I don't know" (IDK). Well, the testee can be honest (choosing IDK every time he doesn't know the correct answer) or can try to guess the correct answer anyway.

If the testee is honest then we have N questions with $n+1$ variants of answer each (n real variants and IDK one), the probability to guess is $p=\dfrac{1}{n}$ (when the honest testee really thinks that he knows the correct answer while he doesn't), C is a number of given correct answers, D is a number of IDK answers, X is a number of questions when testee really knew the correct answer, Y – the number of guessed correct answers (when the honest testee really thought that he knew the correct answer while he didn't). $C=X+Y$. Well, this Y is the same one we used at the beginning, so the SCTR will be

$$X_1=\begin{cases}\dfrac{Cn-(N-D)}{n-1}+1, & \dfrac{N-D}{n}<C\le N-D\\ 0, & 0\le C\le\dfrac{N-D}{n}\end{cases}. \quad (4)$$

The number C in (4) differs from C in (3) because now $C=X+Y$ and $Y=\dfrac{N-D-X}{n}$. But the numerical values of X_0' and X_1 are the same because of

$$C=X+Y=X+\frac{N-D-X}{n}$$

that is $Cn+D=N-X$ doesn't depend upon D; SCTR doesn't depend upon chosen schema or testee's honesty because (4) turns into (3) when $D=0$). E.g. $n=4$, testee knew 40 questions, mistakenly assumed he knew 40 questions and didn't know 20 questions $(N=100)$. Then in case of (3) we have $C=40+\dfrac{60}{4}=55$, $X_0'=\dfrac{55\cdot4-100}{3}+1=41$. And in case of (4) $C=40+\dfrac{40}{4}=50$ and $X_1=\dfrac{50\cdot4-(100-20)}{3}+1=41$. So, **if the testee is dishonest**, will he get the same statistical result? In statistical sense the answer is positive but unfortunately sometimes he will get an even better result just guessing correct answers.

In order to remedy the situation we introduce **a new sub-indicator of knowledge strength** (SIKS) in case of IDK-test

$$k=\frac{X_1-1}{N-D}=$$

$$=\begin{cases}\dfrac{Cn-(N-D)}{(n-1)(N-D)}, & \dfrac{N-D}{n}<C\le N-D\\ 0, & 0\le C\le\dfrac{N-D}{n}\end{cases} \quad (5)$$

It is time to explain why we reduced the numerator. Wasn't it logical to deal with $\dfrac{X_1}{N-D}$? Well, it was. But we're not going to obtain a new SCTR, we're going to introduce SIKS with comfy properties.

First, $k\in[0;1]$ monotonically while

$$C\in\left[\frac{N-D}{n};N-D\right].$$

Second, SIKS k depends upon D – the greater D, the greater k. Third, no matter what kind of user we have – a cheater or just an unsure one – SIKS k will describe the strength of his knowledge. Under conditions of mentioned example in case of (3) we have

$$k=\frac{X_0}{N}=\frac{40}{100}=0.4$$

while in case (4)

$$k=\frac{X_1-1}{N-D}=\frac{41-1}{100-20}=0.5.$$

In practice it will be useful and interesting to include in IDK-test several questions without corrects answers (dead question) and analyze student's choice. For example it is interesting to investigate the correlation between SIKS k and answer to the dead question.

One more interesting question is how SIKS k depends on the level of honesty. Obviously, the testee can behave honesty with some probability $p\in[0;1]$. $p=0$ means the dishonest behaviour (trying to guess correct answer by all means), $p=1$ means the honest behaviour (choosing IDK in appropriate cases). Let $\sim$ be a part of those episodes when student erroneously supposes that he knows the correct answer, in other words $N=X+\sim(N-X)+(1-\sim)(N-X)$ where student really knows X questions and really doesn't know $(1-\sim)(N-X)$ ones. Then he will guess approximately

$$Y=\frac{\sim(N-X)}{n}+\frac{(1-\sim)(N-X)(1-p)}{n}=$$
$$=\frac{(N-X)\left(\sim+(1-\sim)(1-p)\right)}{n}$$

correct answers. If he knows the formula (4) he will try to maximize the value of $Cn+D=C(p)n+D(p)\to\max$ but he'll not be able to because

$$C(p)n+D(p)=(X+Y)n+(1-\sim)(N-X)p=$$

$$=Xn+(N-X)\left(\sim+(1-\sim)(1-p)\right)+$$

$$+\left(1-\sim\right)\left(N-X\right)p=X(n-1)+N$$

doesn't depend upon p. Well, this effect we saw earlier – the honesty doesn't affect the estimated grade. On the other hand, the honesty affects the SIKS k. Really, according to independency X_1 of p we obtain

$$k=k(p)=\frac{X_1-1}{N-D(p)}=\frac{X_1-1}{N-(1-\sim)(N\text{ - }X)p}.$$

This means that honesty is a good strategy to increase the SIKS k though if $\sim\to 1$ the honesty loses its effect: if one cannot distinguish white from black no matter how honest he is. And on the contrary, if $\sim\to 0$ (and may be in addition $X\to 0$ in sense of first line at (4)) then honesty affects SIKS k strongly. E.g. $N=40$, $n=4$, $\sim=0$, $X=2$. Then

$$D=D(p)=(1-\sim)(N-X)p=38p,$$

$$C=C(p)=X+Y(p)=$$

$$=X+\frac{(1-\sim)(N-X)(1-p)}{n}=2+\frac{38(1-p)}{4},$$

$$X_1=\frac{Cn-(N-D)}{n}+1=$$

$$=\frac{8+338(1-p)-40+38p}{3}+1=3$$

(here is our bonus grade – compare $X_1=3$ *vs* $X=2$).

$$k=k(p)=\frac{X_1-1}{N-D(p)}=\frac{2}{40-38p}\in\left[\frac{1}{20};1\right].$$

Well, the effect is rather great: $k_{\min}=k(p)=\dfrac{1}{20}$ and $k_{\max}=k(1)=1$ (the biggest possible value for SIKS k).

Well, we have introduced so-called point estimation while interval estimation can be examined too. The standard confidence interval seems to have a doubtful practical value but it will be usefully to obtain the estimate that could be guaranteed (like "not less than" or "not more than") with some given probability. This can be done both in cases with IDK variant and without it. Some kinds of confidence intervals can be built for SIKS too. The described technique splits up quantitative and qualitative grades but makes it possible to obtain new grades and ratings as functions of introduced estimates (3), (4) and (5).

Список використаних джерел

1. *Булах І. Є.* Створюємо якісний тест: навч. посіб./ І. Є. Булах, М. Р. Мруга. – К.: Майстер-клас, 2006 – 160 с.
2. *Білоусова Л. І.* Потенціал комп'ютерного тестування / Л. І. Білоусова // Вісник ТІМО. – 2008. – № **10**. – С. 40 – 44.

References

1. BULAH I.Ye. (2006) *Creating high quality tests*/ Kiev: Maister-klas.
2. BILOUSOVA L.I. (2008) *Computer-based testing potential*. Visnyk TIMO, **10**.

Надійшла до редколегії 30.05.2015