

Статистически подход за анализ и графично представяне на данни от анкета по проблеми на докторантското обучение

Атанас Николов, Виолета Иванова

The paper's aim is to present a method for processing, analysis and graphic representation of data, based on well-known statistical and algebraic approaches. The information is collected by an inquiry among phd students from different institutions in Bulgaria. The inquiry's purpose is to help doctoral education improvement. An analysis of the questions, which are interpreted according to their level of representing a problem and their consolidation level, is made. For graphical illustration of the data a comparison among the phd students from different education institutions is also made.

Key words: questionnaire data analysis, statistical approach, graphical representation

1. ВЪВЕДЕНИЕ

Качеството на което и да е обучение, зависи основно от правилния избор на учебна програма, систематизирано и последователно поднасяне (чрез надграждане) на материала от обучаващия, както и желание за възприемане от страна на обучаемия. Един начин за оценка на тези „необходимости”, водещи към качествено обучение, е провеждането на анкетно проучване към двете страни, участващи в учебния процес. Но за успешното интерпретиране на отговорите, трябва да са подбрани такива въпроси, от чиито отговори следват недвусмислени изводи. Затова създаването на качествена анкета е от първостепенно значение за оценка на проблемите в областта, към която е насочена анкетата, т.е. за постигане на целите на анкетата.

Използваният подход за статистическо и графично интерпретиране на данни, би могъл успешно да се приложи, както за създаване на качествени нови анкети, така и за усъвършенстване на вече създадени такива [1, 2]. Това може да стане чрез оценка на съдържащите се в тях въпроси от тесни специалисти в съответната област. Всъщност от една страна имаме анализ на това, какви въпроси да включва дадена анкета, а след това интерпретиране на отговорите, дадени от участниците в запитването.

2. КРАТКО ОПИСАНИЕ ЗА НАЧИНА НА СЪБИРАНЕ НА ДАННИ ОТ ПРОВЕДЕНОТО АНКЕТНО ЗАПИТВАНЕ

Анкетното запитване беше проведено измежду 29 докторанти от различни институции в България: ИИКТ-БАН, Русенски университет, Великотърновски университет, Пловдивски университет, Софийски университет, ТУ-София, ТУ-Варна и ТУ-Габрово, където броят на отговорилите докторанти е съответно {8, 5, 4, 4, 2, 2, 2, 2}. Броят на докторантите по институции е твърде малък от статистическа гледна точка, затова трябва да се вземат предвид само отговорите на всичките 29 докторанти. Въпреки това (от демонстративна гледна точка), за целите на графичното представяне е направено и едно сравнение между отговорите на докторантите от първите четири институции.

Анкетата е съставена от 11 въпроса, като всеки въпрос се състои от два, три или четири отговора, на които съответно отговарят равномерно разпределени стойности от 1 до 2, където 1 означава „не”, а 2 – „да”. За въпрос с три отговора, междинната стойност е 1.5, а за въпрос с четири отговора - съответно 1.33 и 1.67. Тези междинни стойности показват градацията на отговорите между „не” и „да”.

3. СТАТИСТИЧЕСКИ ПОДХОД ЗА ОБРАБОТКА И ИНТЕРПРЕТАЦИЯ НА СЪБРАНИТЕ ДАННИ

За извършване на обработка и след това анализ на събраните данни от анкетата, са използвани следните статистически измервания [1]:

- *Средна стойност* за даден въпрос, изчислена чрез отговорите на докторантите от институция I :

$$\mu_q^{(I)} = \frac{1}{|I|} \sum_{i=1}^{|I|} x_{q,i}^{(I)}, \quad (1)$$

където q е номер на въпрос, за който $1 \leq q \leq Q$, Q е броят на всички въпроси; $x_{q,i}^{(I)}$ е отговорът, даден от i -тия докторант на q -тия въпрос, $x_{q,i}^{(I)} \in \{1, 1.33, 1.5, 1.67, 2\}$, $avg\{x_q^{(I)}\} = const = 1.5$; $|I|$ е броят на докторантите от дадената институция I .

- *Стандартно отклонение* за q -тия въпрос около средната стойност $\mu_q^{(I)}$ за дадена институция I :

$$\sigma_q^{(I)} = \sqrt{\frac{1}{|I|} \sum_{i=1}^{|I|} (x_{q,i}^{(I)} - \mu_q^{(I)})^2}. \quad (2)$$

- *Обобщена средна стойност* за дадена институция I :

$$\mu^{(I)} = \frac{1}{Q} \sum_{q=1}^Q \mu_q^{(I)}. \quad (3)$$

- *Осреднено стандартно отклонение* за дадена институция I :

$$\sigma^{(I)} = \sqrt{\frac{1}{Q} \sum_{q=1}^Q (\sigma_q^{(I)})^2}. \quad (4)$$

Чрез тези оценки е извършен анализ относно отговорите на докторантите *от всички институции*, както и на тези *от първите четири институции*. По тази причина за статистическите изчисления ще използваме само първите пет от деветте елемента на множеството I :

$$I = \{All, ИИСТ, RU, VTU, PU \mid SU, TU(S), TU(V), TU(G)\}, \quad (5)$$

където броят на докторантите от съответната институция е:

$$|All| = 29, \quad |ИИСТ| = 8, \quad |RU| = 5, \quad |VTU| = 4, \quad |PU| = 4 \quad (6)$$

Въз основа на горните статистически измервания, можем да сравним получените стойности за всеки въпрос и да направим определено заключение, например кои са *проблематичните въпроси* и къде има *силна консолидация* на докторантите относно наличието или липсата на проблем.

Можем да въведем следните дефиниции:

- *Проблематичността* на даден въпрос означава наличието или липсата на проблем. Тя е реципрочна на съответната средна стойност $\mu_q^{(I)}$ на отговорите за даден въпрос q , виж (1).

• *Консолидацията* около даден въпрос означава консенсус между докторантите около наличието или липсата на проблем за дадения въпрос. Консолидацията е реципрочна на съответното стандартно отклонение $\sigma_q^{(I)}$ около средната стойност $\mu_q^{(I)}$ за даден въпрос, виж (2).

По този начин, може да дефинираме също:

• *Силно проблематичен въпрос* е този, за който $\mu_q^{(I)} \leq \text{avg}\{x_q^{(I)}\} = 1.5$. Посоката на това неравенство се определя от начина, по който са зададени въпросите, т.е. дали отговор с по-ниска стойност означава проблем или не. В нашия случай, това е изпълнено за всички въпроси, с изключение на последния въпрос, за който отговор „да“ се интерпретира като оценка 1, т.е. като проблем за докторантурата. Проблематичните въпроси са маркирани с ‘!’ в Табл. 1 и 2.

• *Силна консолидация* около даден въпрос, означава че $\sigma_q^{(I)} < \sigma^{(I)}$, като тези въпроси са маркирани с ‘§’ в Табл. 1 и 2.

Следните заключения могат да се направят, разглеждайки степента на проблематичност и/или консолидацията около въпросите (виж. Табл. 1):

(!) Силно проблематичните въпроси според докторантите от всички институции ($\text{avg}\{x_q^{(I)}\} = 1.5$) са: “Работите ли и на друго място” ($\mu_{11}^{(All)} = 1.31$), “Удовлетворява ли ви месечното възнаграждение” ($\mu_{10}^{(All)} = 1.41$), “Имате ли взаимни интереси и подкрепа от други млади учени в научната ви област” ($\mu_9^{(All)} = 1.43$).

(§) Четирите най-силно консолидирани въпроси според докторантите от всички институции ($\sigma^{(All)} = 0.33$) са: “Желаете ли всяка година да се провежда докторантско училище по проблемите свързани с дисертацията ви” ($\sigma_5^{(All)} = 0.18$), “Приятна ли е работната ви атмосфера по отношение на хората с които споделяте обща стая” ($\sigma_8^{(All)} = 0.18$), “Добра ли е комуникацията ви с научния ръководител” ($\sigma_6^{(All)} = 0.24$), “След завършването на докторантското си обучение мислите ли да се занимавате с научна дейност” ($\sigma_1^{(All)} = 0.24$).

Таблица 1. Обобщение на отговорите на докторантите от всички институции

No.	Въпрос, (брой възможни отговори)	$\mu_q^{(All)}$	$\sigma_q^{(All)}$	
1	След завършването на докторантското си обучение мислите ли да се занимавате с научна дейност? (3)	1.81	0.24	§
2	Доволни ли сте от обучението си до момента? (3)	1.74	0.34	
3	Полезни ли са за вашата област избраните задължителни докторантски курсове? (3)	1.67	0.30	§
4	Желаете ли да имате повече курсове подпомагащи обучението? (2)	1.72	0.45	
5	Желаете ли всяка година да се провежда “докторантско училище” по проблемите свързани с дисертацията ви? (2)	1.97	0.18	§
6	Добра ли е комуникацията ви с научния ръководител? (4)	1.83	0.24	§
7	Имате ли необходимото техническото оборудване за провеждане на нужните ви експерименти? (4)	1.70	0.30	§
8	Приятна ли е работната ви атмосфера по отношение на хората с които споделяте обща стая? (4)	1.74	0.18	§
9	Имате ли взаимни интереси и подкрепа от други млади учени в научната ви област? (3)	1.43	!	0.31 §
10	Удовлетворява ли ви месечното възнаграждение? (3)	1.41	!	0.44
11	Работите ли и на друго място? (2)	1.31	!	0.46
Осреднени стойности: $\mu^{(All)}$, $\sigma^{(All)}$		1.67	0.33	

! – маркирани са *проблематичните въпроси*, където $\mu_q^{(I)} \leq \text{avg}\{x_q^{(I)}\}$, виж (1), $I = All$, виж (5)

§ – маркирани са въпросите със *силна консолидация*, $\sigma_q^{(I)} < \sigma^{(I)}$, виж (2) и (4), $I = All$, виж (5)

Могат да се направят най-различни изводи и чрез разглеждане на самите стойности на степента на проблематичност и консолидираност около въпросите, например:

♦ За четирите *най-силно консолидирани* въпроси се вижда, че съответните им стойности на $\mu_q^{(All)}$ са също най-високите, което означава консолидираност около най-непроблематичните въпроси.

♦ За въпроса „Имате ли взаимни интереси и подкрепа от други млади учени в научната ви област“, ($\mu_9^{(All)} = 1.43$), въпреки че това не е най-проблематичният въпрос спрямо другите два ($\mu_{11}^{(All)} = 1.31$) и ($\mu_{10}^{(All)} = 1.41$), той има силна степен на консолидация и затова се счита, че е всъщност *най-проблематичният въпрос*.

Относно стойностите в Таблица 2, също могат да се направят аналогични заключения, но понеже броят на отговорилите докторанти от съответните институции е твърде малък от статистическа гледна точка, направените изводи не може да се разглеждат като обобщаващи за съответната група докторанти. Въпреки това, тази извадка е представена тук, във връзка с демонстрация на графичното изобразяване на статистически данни в q -мерното пространство.

Таблица 2. Сравнение на отговорите на докторантите от определени институции

Question No.	ИСТ			RU			VTU			PU		
	$\mu_q^{(ИСТ)}$	$\sigma_q^{(ИСТ)}$	§	$\mu_q^{(RU)}$	$\sigma_q^{(RU)}$	§	$\mu_q^{(VTU)}$	$\sigma_q^{(VTU)}$	§	$\mu_q^{(PU)}$	$\sigma_q^{(PU)}$	§
1	1.75	0.25	§	1.70	0.24	§	1.75	0.25		2.00	0.00	§
2	1.56	0.39		1.90	0.20	§	1.75	0.25		1.88	0.22	§
3	1.63	0.41		1.70	0.24	§	1.63	0.22	§	1.75	0.25	
4	1.63	0.48		1.60	0.49		2.00	0.00	§	2.00	0.00	§
5	2.00	0.00	§	2.00	0.00	§	2.00	0.00	§	2.00	0.00	§
6	1.75	0.14	§	1.93	0.13	§	1.92	0.14	§	1.92	0.14	§
7	1.58	0.32	§	1.87	0.27	§	1.83	0.17	§	2.00	0.00	§
8	1.67	0.17	§	1.80	0.16	§	1.83	0.17	§	1.75	0.28	
9	1.31	! 0.35		1.60	0.20	§	1.50	! 0.35		1.50	! 0.00	§
10	1.56	0.46		1.30	! 0.40		1.25	! 0.25		1.50	! 0.50	
11	1.25	! 0.43		1.60	0.49		1.25	! 0.43		1.25	! 0.43	
$\mu^{(I)}, \sigma^{(I)}$	1.61	0.34		1.73	0.29		1.70	0.24		1.78	0.24	

! – маркирани са *проблематичните въпроси*, където $\mu_q^{(I)} \leq \text{avg}\{x_q^{(I)}\}$, виж (1), $I \neq All$, виж (5)

§ – маркирани са въпросите със *силна консолидация*, $\sigma_q^{(I)} < \sigma^{(I)}$, виж (2) и (4), $I \neq All$, виж (5)

4. ГРАФИЧНО ПРЕДСТАВЯНЕ НА РЕЗУЛТАТИТЕ

За общата графична интерпретация на докторантите от четирите избрани институции (виж Табл. 2) може да се използва 11-мерно евклидово пространство E^{11} , в което отговорите на всеки докторант $phd_i^{(I)}$ за всички въпроси се разглеждат като точка (вектор):

$$phd_i^{(I)} \equiv (x_{1,i}^{(I)}, x_{2,i}^{(I)}, \dots, x_{11,i}^{(I)}), \quad (7)$$

където $x_{q,i}^{(I)} = \{1, 1.33, 1.5, 1.67, 2\}$ е отговорът, даден от i -тия докторант, $1 \leq i \leq |I|$, на q -тия въпрос $1 \leq q \leq 11$, виж коментарите към (1), както (5) и (6).

Целта на тази интерпретация е да се използва Евклидовата норма $\|v\|$ (или дължина) на даден вектор v , $v \in E^{11}$:

$$\|v\| = \|v_1, v_2, \dots, v_{11}\| = \sqrt{\sum_{q=1}^{11} v_q^2}.$$

Имайки предвид (8) и Питагоровата теорема, разстоянието $d(A, B)$ между два произволни вектора A и B в \mathbf{E}^{11} се изчислява чрез:

$$d(A, B) = \|A - B\| = \sqrt{\sum_{q=1}^{11} (A_q - B_q)^2}, \quad (9)$$

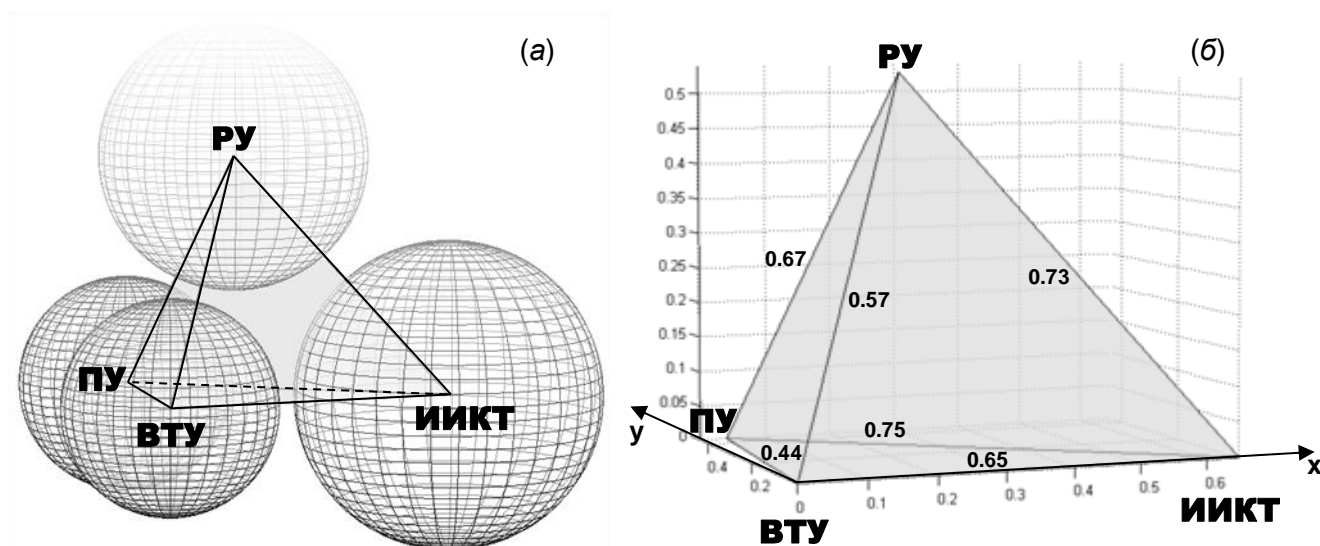
където $A \equiv (A_1, A_2, \dots, A_{11}) \in \mathbf{E}^{11}$, $B \equiv (B_1, B_2, \dots, B_{11}) \in \mathbf{E}^{11}$.

По аналогичен начин, средните стойности от отговорите на всички докторанти от институция I за даден въпрос, могат да се разглеждат като координати на точка (вектор) в \mathbf{E}^{11} , която представлява център на съответната институция:

$$(I) \equiv (\mu_1^{(I)}, \mu_2^{(I)}, \dots, \mu_{11}^{(I)}), \quad I = \{ИКТ, RU, VTU, PU\}, \quad (10)$$

където $\mu_q^{(I)}$ са средните стойности за даден въпрос, виж (1).

По такъв начин, чрез (9), са изчислени съответните разстояния между центровете на четирите институции, които се разглеждат като тетраедър с върхове (VTU, ИИКТ, PU, PU). Тяхното разположение в $\mathbf{E}^{11} \rightarrow \mathbf{E}^3$, е илюстрирано на Фиг. 1.



Фигура 1. Графична илюстрация на разположението на институциите в пространството (а) и съответните разстояния между тях (б).

Необходимо е да бъдат направени следните обяснения във връзка с интерпретирането на Фиг. 1, [1]:

- Областта на интерес е ограничена до 11-мерен хиперкуб, чиито 11 основни ръба съвпадат с координатните оси на \mathbf{E}^{11} , като всяка ос съответства на определен въпрос от анкетата.

- Всяка институция I се разглежда като клас (или облак) от точки, (т.е. попълвания на анкетата от съответната група докторанти), чиито брой за съответния клас е равен на $|I|$, виж (6).

- Всеки клас може да се разглежда като 11-мерен елипсоид, с център (I) , според (10) и (1), и с размери по неговите 11 главни оси, пропорционални на съответните стандартни отклонения $\sigma_q^{(I)}$, $q = 1 \div 11$, съгласно (2).

- За да бъде възможно графичното построение, елипсоидът на всеки клас $I = \{ИКТ, RU, VTU, PU\}$ е апроксимиран чрез 11-мерна хиперсфера със същия

център (I) и радиус, равен на осреднено стандартно отклонение $\sigma^{(I)}$ за съответната институция, виж (4).

• С други думи, може да разглеждаме илюстрацията от Фиг. 1, като „поглъщане” на 11-мерното пространство в тримерното пространство ($E^{11} \rightarrow E^3$), по такъв начин, че шестте 11-мерни разстояния между четирите върха (ИИКТ, ВТУ, ПУ, РУ) *остават непроменени* в тримерното пространство E^3 .

Понеже не е коректно да се правят конкретни изводи на базата на недостатъчно статистическа информация, както е в случая, ще разгледаме по принцип как трябва да се интерпретира *дължината на разстоянието* между сравняваните институции, както и *големината на сферите* около тях, (Фиг. 1):

1) По-голямо разстояние между два върха означава, че разликата между осреднените отговори ($\mu_q^{(I)}$) на един и същ въпрос (q), дадени от докторантите от сравняваните институции (I), е по-голямо. Следователно, ако са дадени близки осреднени отговори на по-голям брой въпроси, то разстоянието между центровете на институциите ще бъде по-малко.

2) Когато консолидацията на докторантите (от една институция) около всеки един въпрос е по-голяма (т.е. по-малко $\sigma_q^{(I)}$), тогава и средната консолидация около всички въпроси ще бъде по-голяма (т.е. по-малко $\sigma^{(I)}$). Следователно радиусът на съответната сфера около дадената институция, ще бъде по-малък.

5. ЗАКЛЮЧЕНИЕ

В настоящата статия са приложени количествени и качествени подходи за оценка, анализ и (идея за) графично представяне на резултатите от проведената анкета, свързана с някои проблеми на докторантското обучение в България. Направени са обобщаващи изводи, относно отговорите на докторантите от всички институции чрез прилагане на определена класификация към въпросите, а именно – степен на проблематичност и степен на консолидация между докторантите около даден въпрос. В заключение може да се каже, че статията е успешно адаптирана към методите за статистически анализ и класификация на данни, предложени в [1].

БЛАГОДАРНОСТИ

Тази разработка е подкрепена финансово от Проект No: BG 051 PO 001-3.3.04/13, “Подкрепа на творческото развитие на докторанти, пост-докторанти и млади учени в областта на компютърните науки”, финансиран от Европейския социален фонд, Оперативна програма “Развитие на човешките ресурси” 2007-2013.

Статията е разработена под ръководството на доц. д-р Д. Димов от ИИКТ-БАН.

ЛИТЕРАТУРА

[1] Dimov D., N. Dokev, P. Asenova, and S. Ivanov. On Generalized Evaluation of PhD Competences in Computing. In: Proceedings of 3-d e-Learning Conf. Sept 7-8, 2006, Coimbra, Portugal, pp. 2.3.1-8.

[2] ETN-DEC Final Report, 2007, ETN-DEC final meeting in Istanbul, August 27, 2007, Avangard Print Ltd, Rousse, BG, 2007.

ЗА КОНТАКТИ:

Атанас Филипков Николов, Секция “Обработка на сигнали и разпознаване на образи”, ИИКТ, БАН, тел.: (02) 979 2925, e-mail: a.nikolov@iinf.bas.bg

Виолета Иванова Иванова, Секция “Обработка на сигнали и разпознаване на образи”, ИИКТ, БАН, тел.: (02) 979 2925, e-mail: lety78_2000@abv.bg