

Уеб-базиран софтуер за обработка на анкети

Дияна Лечева, Красимир Харизанов, Павлина Йорданова¹

***An Web-based software for handling inquiries:** In this paper we present a new product that contains the most important statistical techniques for handling inquiries: absolute and relative frequencies, different percentages, descriptive statistics, confidence intervals, crosstabs, correlations. It is web based and appropriate for simple marketing, social investigations, political study.*

A demonstration with estimation of quality of education is presented.

Key words: Web-based software, handling inquiries, PHP, MySQL.

ВЪВЕДЕНИЕ

През последните години се отделя все по-голямо внимание на изучаването на общественото мнение особено в области като: социология, маркетинг или политика. Това прави използването на анкети все по-необходимо. Техният анализ е свързан с широк обхват от знания в наблюдаваната област, по статистика и информационни технологии. Това възпрепятства тяхната употреба от широкия потребител.

Защо предложеният продукт е актуален?

Можем да разделим до сега съществуващия софтуер за уеб-базирана обработка на анкети в три групи. Първата група продукти съдържат мощни статистически техники и основният техен недостатък е, че са скъпи. Такива са например: SPSS, SAS, Matlab, S+, създадени от IBM, SAS Institute Inc., MathWorks и TIBCO Software Inc. Към втората група влиза софтуер с отворен код, който изисква значителни познания по информатика преди да бъде използван. Виж например R от [7]. Той е разработен от широк кръг учени, но за негови създатели се смятат Robert Gentleman и Ross Ihaka. Третата група съдържа безплатният, но прекалено беден от към статистически техники уеб-базиран софтуер, който се използва предимно за електронен маркетинг и съдържа единствено статистически групировки.

С предложеният софтуер сме се опитали да отделим най-често използваните статистически техники. Интерфейсът е лесен за употреба, дава възможност за създаване на нови анкети и е подходящ за работа с различни видове въпроси. Обработката е реализирана на езика PHP, а базата данни се съхранява в MySQL. По-подробно описание на продукта може да бъде намерено в [6].

ИЗПОЛЗВАНИ СТАТИСТИЧЕСКИ ТЕХНИКИ

Анкетирането е една от основните техники за проучване на общественото мнение. Чрез него се събира обратна информация от обществото за изучаваната генерална съвкупност, процес или явление. От своя страна тази информация обикновено служи в последствие за тяхното управление. За това безпристрастното обработване на анкетните е важен и отговорен процес. Много полезни в това отношение са основните статистически методи. В софтуера, който предлагаме сме включили възможност за представяне на данни в статистически ред, получаване на описателни статистики, доверителни интервали, кръстосани честотни таблици, в абсолютни и относителни честоти, коефициенти за корелация, графично представяне на резултатите. Поради сравнително широката популярност на тези характеристики няма да ги разглеждаме подробно. Виж [2] и [3]. Ще се спрем само на нормирания χ^2 -коефициент на корелация на Пирсън и Чупров, разгледан в [1].

Да предположим, че наблюдаваните признаци X и Y имат възможни значения съответно X_1, \dots, X_k и Y_1, \dots, Y_s и че резултатите от групировката, след наблюдението са разположени както в Табл. 1.

¹ Статията е частично финансирана от НИП № РД-08-255/14.03.2013.

С f_{ij} е означен броят на статистическите единици, притежаващи i – тото значение на признака X и j – тото значение на признака Y . f_{11}^X е абсолютната честота в първата група на признака X . По аналогичен начин до f_{k1}^X, f_{11}^Y е абсолютната често-

Таблица 1.

$X \backslash Y$	Y_1	...	Y_s	Общо:
X_1	f_{11}	...	f_{1s}	$f_{1\cdot}^X$
...
X_k	f_{k1}	...	f_{ks}	$f_{k\cdot}^X$
Общо:	$f_{\cdot 1}^Y$...	$f_{\cdot s}^Y$	n

та в първата група на признака Y и т.н. до f_{1s}^Y . n е броят на всички наблюдавани единици. В сила са следните съотношения

$$n = \sum_{i=1}^k \sum_{j=1}^s f_{ij}, \quad f_{\cdot i}^X = \sum_{j=1}^s f_{ij}, \quad f_{\cdot j}^Y = \sum_{i=1}^k f_{ij}.$$

Ако признаците са независими, теоретичните (очакваните) честоти в групите биха били

$$\hat{f}_{ij} = \frac{f_{\cdot i}^X f_{\cdot j}^Y}{n} = n \hat{p}_{ij}, \quad (1)$$

където \hat{p}_{ij} е най-добрата оценка на $P(X = X_i, Y = Y_j)$. Да припомним, че ако признаците са независими

$$P(X = X_i, Y = Y_j) = P(X = X_i) \cdot P(Y = Y_j)$$

и най-добрите оценки на $P(X = X_i)$ и $P(Y = Y_j)$ са съответно $\frac{f_{\cdot i}^X}{n}$ и $\frac{f_{\cdot j}^Y}{n}$.

Коефициентът на средно-квадратична взаимосвързаност (на контингенция) на Пирсън има вида

$$\varphi^2 = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^s \frac{(f_{ij} - \hat{f}_{ij})^2}{\hat{f}_{ij}}.$$

Величините $r = \sqrt{\frac{\varphi^2}{1 + \varphi^2}}$ и $r = \sqrt{\frac{\varphi^2}{\sqrt{(k-1)(s-1)}}}$ се наричат съответно коефициенти на корелация на Пирсън и Чупров.

При анализа на коефициента на Пирсън трябва да се взема под внимание, че

$$r_{\text{Пирсън}} \leq \sqrt{\frac{\min(k-1, s-1)}{1 + \min(k-1, s-1)}}.$$

За да се сравнява с 0 и 1, вместо самия коефициент на Пирсън е по-удачно да се използва частното му с максималната му стойност.

В случая на два дихотомни признака, т.е. ако $k = 2$ и $s = 2$ корелационната таблица има вида на Табл. 2.

Таблица 2.

$X \backslash Y$	Y_1	Y_2	Общо:
X_1	a	b	$a+b$
X_2	c	d	$c+d$
Общо:	$a+c$	$b+d$	n

Коефициентът на средно-квадратична взаимосвързаност (на контингенция) на Пирсън приема вида

$$\varphi^2 = \frac{(ad - bc)^2}{(a+c)(b+d)(a+b)(c+d)}.$$

В този случай

$$\varphi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \leq \frac{1}{\sqrt{2}}$$

се нарича четириклетъчен коефициент на взаимосвързаност (контингенция) на Пирсън.

ОПИСАНИЕ НА ПРОДУКТА

Представената уеб-базирана система за обработка на анкети е дружелюбно настроена към потребителя, което улеснява работата с нея. Не са нужни предварителни специални компютърни умения и познания, а само използването на уеб-браузър и минимална компютърна грамотност, необходима за правилното ориентиране в интернет среда.

Задачите на приложението са програмна реализация за систематизиране и първична обработка на данни чрез функционалните езици PHP и MySQL.

Избран е MySQL защото, това е най-популярната система за управление на бази данни с отворен код. SQL е предназначен за създаване, обработка и четене на бази данни, които представляват пакети от свързана информация, съхранявана в таблици. MySQL поддържа интерфейси за програмиране под множество езици - C/C++, Eiffel, Java, Perl, Python, което дава много възможности за работа. Данните, съхранявани в MySQL могат да бъдат от всякакъв вид - от обикновени текстови списъци до сложна структурирана информация. MySQL организира данните в таблици и позволява бърз достъп до тях, поради което широко се използва при изграждането на уеб сайтове, особено такива с по-голям обем информация, изискваща бързо четене и обработка.

Другият основен език, на който е написан кодът на предложението софтуер, е PHP. Това е скриптов език с отворен код, който е проектиран за уеб програмиране и е широко използван за създаване на сървърни приложения и динамично уеб-съдържание.

Уеб приложението разполага с административно ниво на достъп, осигуряващо лесно управление на анкетните въпроси с възможност да се определи мерността на оценяване и броя на възможните отговори.

Администраторът определя, наименованието на анкетата и броя на въпросите, включени в нея. Те се въвеждат в последователен ред до тяхното изчерпване. При попълване на конкретен въпрос за първи път, се определя броят на възможните отговори и начина, по който ще се посочи отговора (радио бутон, чек бокс, списък и др.). Предложеният продукт съдържа инструменти, позволяващи на администратора след въвеждане на всички въпроси да ги редактира и да управлява тяхното съдържание. Всеки от отговорите на анкетирания съответства на една емпирична стойност на наблюдавания признак. При случаен избор на анкетирания, на техните отговори се гледа като на реализации на случайни величини. Това дава възможност при обработката им да се използва апаратът на Теорията на вероятностите и Математическата статистика. Потребителят след оторизация избира анкетата, която ще попълва. За запазване на анонимността той не въвежда лични данни.

В софтуерът, който представяме на отговорите всеки един анкетирания съответства един списък. Различните отговори на анкетирания са компоненти на този списък, които могат да бъдат от различен тип, в зависимост от вида на въпроса.

След въвеждането на анкетите с диалогов прозорец могат да се направят едномерни и двумерни групировки, в брой, в процент от реда, в процент от колоната или в процент от общото. При реализирането на групировките те са съобразени с видовете въпроси. Виж [4]. Резултатите от попълнените анкети администратора,

може да визуализира в различни варианти, с визуализиране на текст, с визуализиране на кода на отговора и т.н.

Според броя на възможните едновременни отговори от един анкетиран въпросите са разделени на:

✓ **Едновариантен въпрос:** Това е въпрос, при който анкетираният може да посочи най-много един отговор. Техен подвид са **дихотомните въпроси**. Това са въпроси, с два възможни отговора. Такъв например е въпросът:

Какъв е вашият пол?

✓ **Многовариантен въпрос:** Това е въпрос, при който анкетираният може да посочи повече от един отговор. Такъв например е въпросът:

Какви езици владеете?

При обработването на анкетите многовариантните въпроси се свеждат към няколко дихотомни. Като например:

Владееете ли български?

Владееете ли английски?

Владееете ли руски? и т.н.

Друг начин за обработването на многовариантните въпроси е като се сведат към един едновариантен въпрос, като на всяка комбинация от въпроси се съпостави съответен шифър.

✓ **Въпрос със свободен отговор (отворени въпроси):** Това са въпроси, при които анкетираният, по негова преценка използва измислено от него, словестно описание на отговора си. При обработката тези въпроси се свеждат към един от горните два вида.

Друг съществен момент, с който е съобразена обработката на отговорите на въпросите е вида на признака, който се описва със съответния въпрос.

✓ **При неметриран** признак обикновено класовите представители на групата са кратки, описателни характеристики на съответната подсъвкупност.

✓ **При метриран признак с малко възможни значения** отново за всяко различно значение на признака имаме отделна подгрупа.

✓ **При метриран признак с много възможни значения на признака** се оформят интервали. При групировката всяка статистическа единица се отнася към съответния интервал. За определянето на ширините и краищата на интервалите сме използвали правилата, описани в [2].

В представеният софтуер е направен диалогов прозорец за определяне на дескриптивни статистики заедно с техните доверителни интервали. Той е удачно да се използва след извършване на групировка. Формулите, които сме използвали за пресмятане на дескриптивните статистики и доверителните интервали са дадени например в [3]. След посочване на конкретна клетка от статистическия ред, кръстосаната таблица или списъкът с дескриптивни статистики, в диалогов прозорец се появява доверителният интервал на съответната характеристика. Рискът за грешка при построените доверителни интервали е съобразен с очакваните области, в които ще се използва предложеният софтуер и е 0,01 или 0,05. Той, заедно с границите на доверителните интервали, дава представа за точността на изследването.

След разглеждането на кръстосаната таблица, в продукта, който представяме има възможност и за измерване на зависимостите между наблюдаваните признаци. За целта е необходимо, например в горния ляв ъгъл на съответната корелационна таблица, с десен бутон да изберем от падащо меню съответният корелационен коефициент. Имплементирани са

✓ **корелационният коефициент на Браве,**

✓ **корелационният коефициент на Спирмън,**

✓ **корелационните коефициенти на Пирсън и Чупров.**

За тяхното описание и тълкуване виж например [2] или [5].

ПРИМЕР ОТ ОЦЕНКА НА КАЧЕСТВОТО НА ОБРАЗОВАНИЕТО

Представената уеб-базирана система за обработка на анкети може да служи за анализ на качеството на обучение и преподаване. В зависимост от целите на изследването съставяме въпросите в анкетата. Първият файл от приложението е start.php. Той служи за вход и създаване на въпроси, които са общи за всички анкети както и препратки към задаване на другите въпроси, към възможности за тяхното попълване и обработка. Например чрез него в текстова кутия въвеждаме курса. От списъчно поле избираме специалността. Изборът на пол става чрез радио бутон. Преминаваме към създаването на въпросите от анкетата. След активирането на файла от списъчно поле избираме тази анкета, която искаме да попълним. Накрая се натиска бутон „Влез“ и започва попълването на анкетата. Клиентът се препраща към файл quiz.php, където се намира самата анкета за попълване. След приключване на попълването ѝ се натиска бутон „Отговори“ и информацията се прехвърля в базата данни. Накрая във файл analiz.php се извършват статистическите групировки по различни признаци. Той осъществява връзката на потребителя с тази базата данни. За всеки въпрос има зададени бутони, които дават възможност да бъде определен какъв процент от отговорите е в съответната група, брой на отговорилите и процент от отговорилите. При групировка по два едновариантни въпроса, по два многовариантни и едновариантен с многовариантен въпрос, може да се определи процент от анкетираните, процент от отговорите, процент от сумата в реда, процент от сумата в колоната, процент от общото. Накрая се определят дескриптивните статистики и се прави анализ на получените резултати.

ЛИТЕРАТУРА

- [1] А. Афифи, С. Айзен, Статистически анализ. Подход с използване на ЕВМ., Мир, М., 1982.
- [2] Хр. Вълчев, П. Йорданова, Статистика. УИ „Епископ Константин Преславски“, Шумен, 2004.
- [3] К. Гатев, „Въведение в статистиката“, София, 1995 г.
- [4] В. Гоев, „Статистическа обработка и анализ на информацията от социологически, маркетингови и политически изследвания със SPSS“, София, 1996 г.
- [5] Б. Димитров, Н. Янев, Вероятности и статистика, УИ „Климент Охридски“, София, 1990 г.
- [6] Д. Лечева, Дипломна работа: Уеб-базиран софтуер за първична обработка на анкети, ШУ „Епископ Константин Преславски“, Шумен, 2013.
- [7] <http://www.r-project.org/>

За контакти:

Гл. ас. д-р Павлина Йорданова, Катедра „Икономика и моделиране“, Шуменски университет „Епископ Константин Преславски“, тел.: 054-830 495/ 138, e-mail: pavlina_kj@abv.bg

ас. Красимир Харизанов, Катедра „Методика на обучението по математика и информатика“, Шуменски университет „Епископ Константин Преславски“ тел.: 054-830 495/ 136, e-mail: kr.harizanov@abv.bg

Дияна Любенова Лечева, студентка в Шуменски университет „Епископ Константин Преславски“, специалност „Икономическа информатика“ и втора специалност „Икономика“, тел.: 0883303221 e-mail: didi.lecheva@abv.bg

Докладът е рецензиран.