

tom 2 /2012

# Rola informatyki w naukach ekonomicznych i społecznych Innowacje i implikacje interdyscyplinarne



**KAPITAŁ LUDZKI**  
NARODOWA STRATEGIA SPÓJNOŚCI



Wyższa Szkoła Handlowa  
im. Bolesława Markowskiego  
w Kielcach

**UNIA EUROPEJSKA**  
EUROPEJSKI  
FUNDUSZ SPOŁECZNY



# **Rola informatyki w naukach ekonomicznych i społecznych**

## **Innowacje i implikacje interdyscyplinarne**

redakcja  
ZBIGNIEW E. ZIELIŃSKI

**TOM 2**

Recenzja naukowa  
prof. zw. dr hab. Jan Turyna



Wydawnictwo  
Wyższej Szkoły Handlowej  
Kielce 2012

Publikacja wydrukowana została zgodnie z materiałem dostarczonym przez Autorów. Wydawca nie ponosi odpowiedzialności za treść, formę i styl artykułów.

#### **Komitet Naukowy**

prof. dr hab. Janusz Lewandowski  
prof. dr hab. Krzysztof Grysa  
dr hab. Wiesław Dziubdziela, prof. WSH

#### **Redaktor Naczelny**

prof. zw. dr hab. Tadeusz Grabiński

#### **Redaktor Recenzji**

prof. zw. dr hab. Jan Turyna

#### **Recenzenci**

prof. zw. dr hab. Tadeusz Grabiński  
prof. zw. dr hab. Witold Chmielarz  
prof. nadzw. dr hab. inż. Wacław Gierulski  
prof. dr hab. Witold Biały  
prof. dr hab. Wiesław Dziubdziela  
prof. dr hab. Ewa Grzegorzewska-Ramocka  
prof. dr hab. Grzegorz Kończak  
prof. dr hab. Włodzimierz Mosorow  
prof. dr hab. Mieczysław Muraszewicz  
prof. dr hab. Andrzej Radomski  
prof. dr hab. inż. Magdalena Wyrwicka  
doc. dr inż. Zbigniew Lis  
dr inż. Edward Wiszniowski  
dr inż. Radosław Wolniak  
dr Paulina Forma  
dr Małgorzata Paszkowska  
dr Lidia Pokrzycka  
dr Sławomir Wyciślak  
dr Dariusz Żak

#### **Redakcja**

dr Zbigniew E. Zieliński  
mgr Katarzyna Baziuk  
mgr Anna Błaszczuk  
mgr inż. Artur Janus  
mgr inż. Jarosław Kościelecki  
mgr Urszula Słowik  
mgr Piotr Sidor

#### **Wydawca publikacji**

Wyższa Szkoła Handlowa im. B. Markowskiego w Kielcach  
Projekt „PITWIN – Portal Innowacyjnego Transferu Wiedzy w Nauce”  
ul. Peryferyjna 15  
25-562 Kielce  
[www.pitwin.edu.pl](http://www.pitwin.edu.pl), [biuro@pitwin.edu.pl](mailto:biuro@pitwin.edu.pl)

© Copyright by Wyższa Szkoła Handlowa, Kielce 2012  
ISBN 978-83-89274-75-5  
Nakład 200 egz.

Publikacja została wydana w ramach realizacji projektu „PITWIN – Portal Innowacyjnego Transferu Wiedzy w Nauce”.

Publikacja jest współfinansowana przez Unię Europejską w ramach Europejskiego Funduszu Społecznego. Publikacja jest dystrybuowana bezpłatnie, dla osób które zarejestrują się na stronie internetowej projektu [www.pitwin.edu.pl](http://www.pitwin.edu.pl) (dostępna także w wersji elektronicznej).

Spis treści .....	3
-------------------	---

## VI. ANALIZY IŁOŚCIOWE W NAUKACH EKONOMICZNYCH I SPOŁECZNYCH

1. <b>Aleksandra Baszczyńska</b> - Estymacja funkcji gęstości z pakietem MATLAB .....	7
2. <b>Tomasz Bąk</b> - Udział lasów w procesie redukcji CO <sub>2</sub> – aspekty ekonomiczne .....	17
3. <b>Agnieszka Buś-Bidas</b> - Gospodarowanie mieniem gminnym i jego znaczenie dla funkcjonowania przedsiębiorstw .....	22
4. <b>Barbara Buzowska</b> - Monitoring percepcji klienta jako narzędzie doskonalenia systemu zarządzania jednostką samorządową.....	30
5. <b>Sebastian Chmielewski</b> - Komputerowa symulacja dynamicznych procesów ekonomicznych .....	35
6. <b>Magdalena Chmielińska</b> - Wykorzystanie narzędzi informatycznych w konstrukcji planów odbiorczych .....	47
7. <b>Paulina Duma</b> - Kryptografia. Rola informatyki w tworzeniu zabezpieczeń teleinformatycznych .....	61
8. <b>Wiesław Dziubdziela</b> - Matematyczne modele ryzyka systemowego po kryzysie finansowym w latach 2007-2008.....	72
9. <b>Anna Fiema, Grzegorz Kończak</b> - Wykorzystanie rozkładów wartości ekstremalnych w analizie zagrożeń meteorologicznych.....	80
10. <b>Łukasz Goczek</b> - Skutki regulacji dla inwestycji w badania i rozwój na poziomie firm –wnioski z analizy regresji logistycznej.....	90
11. <b>Justyna Karkoszka</b> - Rozwój sfery badawczo-rozwojowej oraz innowacji w krajach Europy Środkowo-Wschodniej należących do Unii Europejskiej – finansowanie B+R w latach 2004-2010.....	98
12. <b>Adam Kiersztyn</b> - MarkovEdu – koncepcja programu wspierającego nauczanie łańcuchów Markowa .....	110
13. <b>Karolina Klimańska</b> - Poziom życia a warunki mieszkaniowe w Polsce w układzie wojewódzkim w latach 2002 i 2008.....	120
14. <b>Grzegorz Kończak</b> - Symulacja komputerowa – Monte Carlo czy quasi-Monte Carlo?.....	151
15. <b>Przemysław Kowalik</b> - O zasadach konstruowania taryf w pasażerskim transporcie kolejowym .....	160
16. <b>Przemysław Kowalik</b> - Wykorzystanie arkuszy kalkulacyjnych do wyboru zmiennych objaśniających przy pomocy metody wskaźników pojemności informacyjnej (metody Hellwiga).....	168
17. <b>Małgorzata Krzciuk, Piotr Ziuziański</b> - O teście niezależności trzech zmiennych na pewnym przykładzie empirycznym .....	179
18. <b>Marta Małecka</b> - Wykorzystanie nowych funkcji programu Mathematica 8 w ocenie ryzyka rynkowego .....	192
19. <b>Michał Mierzwa</b> - O złej specyfikacji modelu nadpopulacji na przykładzie badania ankietowego studentów UE Katowice .....	204
20. <b>Artur Mikulec</b> - Analiza skupień z wykorzystaniem programu komputerowego ClustanGraphics .....	214
21. <b>Adam Pająk</b> - Analiza rankingów uczelni wyższych w Polsce z lat 2002-2012.....	225
22. <b>Dorota Pekasiewicz</b> - Zastosowanie metod symulacyjnych do badania własności estymatorów otrzymanych metodą kwantyli .....	236
23. <b>Dorota Pekasiewicz, Anna Szymańska</b> - Zastosowanie testów zgodności do badania rozkładów wielkości szkód w ubezpieczeniach komunikacyjnych .....	245
24. <b>Witold Rzymowki, Agnieszka Surowiec</b> - Method of parameters estimation of pseudologistic model .....	256
25. <b>Hanna Soroka-Potrzebna</b> - Sytuacja ekonomiczno-finansowa polskich przedsiębiorstw na przestrzeni lat 2000-2010 .....	267
26. <b>Jacek Stelmach</b> - O weryfikacji hipotez z wykorzystaniem dwóch statystyk testowych .....	276

27. <b>Małgorzata Szerszunowicz</b> - Analiza przestrzenna bezrobocia w Polsce z wykorzystaniem technik ESDA .....	289
28. <b>Miroslaw Zajdel</b> - Zachowanie tłumu modelowanego w oparciu o behavior owadów społecznych w obliczu różnych konfiguracji wyjść ewakuacyjnych.....	298
29. <b>Tomasz Żądło</b> - O szacowaniu parametrów modeli dla danych przekrojowo-czasowych z pakietem R.....	309

## VII. ANALIZY SPOŁECZNO-EKONOMICZNE

30. <b>Sebastian Chmielewski</b> - Sektor TSL w Polsce .....	325
31. <b>Krzysztof Czubocho</b> - Wybrane aspekty transformacji społeczno-gospodarczej w Polsce w kontekście integracji Polski z Unią Europejską (UE): refleksje socjologa.....	347
32. <b>Łukasz Furman</b> - Optymalizacja podatkowa w ramach specjalnej strefy ekonomicznej .....	361
33. <b>Maciej Jagódka</b> - Kreowanie gospodarki opartej na wiedzy jako główny instrument podnoszenia konkurencyjności regionów .....	371
34. <b>Maciej Kłodawski</b> - Pojęcie informacji w naukach teoretycznoprawnych.....	379
35. <b>Krzysztof Kocurek</b> - Programowanie rozwoju regionalnego – kluczowe wyzwania .....	390
36. <b>Weronika Kumańska</b> - Wykluczenie społeczne osób niepełnosprawnych .....	399
37. <b>Agnieszka Kwiatkowska</b> - Cyberterrorizm problemem współczesnego świata.....	410
38. <b>Aneta Lipczyńska</b> - Pojęcie, rodzaje i znaczenie obligacji dla polskiej gospodarki .....	416
39. <b>Tomasz Marciniak, Ewa Kopeć</b> - Analiza załadunku i sposoby zabezpieczenia wagonów z surowcem drzewnym oraz znaczenie informatyzacji w monitorowaniu załadunku .....	430
40. <b>Lidia Pokrzycka</b> - Analiza zawartości prasy. Wnioski z badań nad prasą województwa lubelskiego.....	441
41. <b>Dominika Polko</b> - Zastosowanie metody statystycznej kontroli jakości w podejmowaniu decyzji giełdowych .....	449
42. <b>Artur A. Trzebiński</b> - Fundusze inwestycyjne a oszczędności Polaków.....	461
43. <b>Grzegorz Wilk-Jakubowski</b> - Media a terrorizm .....	472
44. <b>Miroslaw Zajdel</b> - Przemiany zatrudnienia w regionie łódzkim w okresie transformacji (wybrane aspekty).....	480
45. <b>Dariusz Grzegorz Żak</b> - Ochrona patentowa na wynalazek .....	492



Małgorzata Krzciuk  
Piotr Ziuziański

## O teście niezależności trzech zmiennych na pewnym przykładzie empirycznym

**Streszczenie:** *LimeSurvey to wolny i otwarty, rozbudowany system zarządzania badaniami ankietowymi. W pracy zostały przedstawione podstawowe jego funkcje, a także możliwości wykorzystania. Opisano również jego budowę i zastosowane technologie. System ten został wykorzystany w badaniu ankietowym studentów Uniwersytetu Ekonomicznego w Katowicach. W artykule omówiono także podstawy teoretyczne analizy pełnej oraz częściowej niezależności trzech zmiennych. Na przykładzie empirycznym – wybranych danych pochodzących z wspomnianego badania zaprezentowane zostało jej zastosowanie z wykorzystaniem funkcji wbudowanych i samodzielnie zaprogramowanych w języku R.*

**Słowa kluczowe:** CAWI, CAII, LimeSurvey, niezależność stochastyczna, test Chi kwadrat niezależności

### 1. Badanie ankietowe studentów UE Katowice

W maju 2012 r. Koło Naukowe Statystyków Uniwersytetu Ekonomicznego w Katowicach przeprowadziło badanie ankietowe z wykorzystaniem LimeSurvey w wersji 1.91+. Brał w nim udział studenci pierwszego roku kierunku Analityka gospodarcza na Wydziale Zarządzania Uniwersytetu Ekonomicznego w Katowicach w roku akademickim 2011/2012.

Przed przystąpieniem do badania ustalono cele główne i poboczne. Nad opracowaniem kwestionariusza ankiety od sformułowania pytań po ustalenie odpowiedniej kolejności i wdrożeniem go do systemu pracował zespół studentów. Uwzględniono skale pomiarowe i zaplanowano metody analizy zebranych zmiennych, postawiono także hipotezy badawcze.

Celem badania było poznanie motywów wyboru kierunku Analityka gospodarcza, a także stworzenie typowego profilu studenta tegoż kierunku. Ponadto pobocznymi celami była ocena materiałów promocyjnych kierunku, a także poznanie planów studentów odnośnie studiów drugiego stopnia i oczekiwań dotyczących przyszłej kariery zawodowej oraz zainteresowanie kołami naukowymi. W kwestionariuszu zostało zawarte także pytanie dotyczące korzystania ze świadczeń socjalnych (tzw. pytanie o kwestie wrażliwe).

Przed sformułowaniem kwestionariusza ankiety postawiono następujące hipotezy badawcze odnośnie studentów kierunku Analityka gospodarcza:

- Badana grupa studentów preferuje przedmioty ścisłe;
- Studenci dokonali świadomego wyboru kierunku;
- Głównym źródłem informacji o kierunku były materiały promocyjne Uczelni;
- Większość studentów ma zamiar kontynuować kierunek na studiach uzupełniających.

Ostateczne badanie zostało poprzedzone badaniem pilotażowym, w celu weryfikacji poprawności i przejrzystości kwestionariusza ankiety. Dzięki wykorzystaniu LimeSurvey kwestionariusz ankiety po przygotowaniu i badaniu pilotażowym był łatwy w modyfikacji. Zebrane dane dzięki możliwościom eksportu systemu do programu R były poddane analizie w tymże programie.

W ubiegłym roku przeprowadzono podobne badanie wykorzystując opracowany samodzielnie kwestionariusz ankiety. Koło Naukowe Statystyków planuje powtórzenie badania dla kierunku Analityka gospodarcza w kolejnych latach dla studentów pierwszego roku

z wykorzystaniem LimeSurvey, zyskując w ten sposób materiał porównawczy umożliwiający pogłębioną analizę.

Szczegóły dotyczące systemu LimeSurvey zostaną zaprezentowane w drugiej części opracowania. Następnie zostaną przedstawione wybrane problemy analizy zebranych danych w programie R.

## 2. System zarządzania badaniami ankietowymi LimeSurvey

Badania ankietowe są jedną z metod pozyskiwania danych. W dobie Internetu na znaczeniu bardzo zyskały internetowe badania ankietowe zwane CAII lub CAWI (ang. Computer Assisted Internet/Web Interview).

Ankieter chcący przeprowadzić badanie CAWI staje przed wyborem pomiędzy gotowym rozwiązaniem w postaci systemu zarządzania ankietami a przygotowaniem własnego rozwiązania. Stworzenie własnej ankiety wymaga znajomości odpowiedniego języka programowania (np. PHP), baz danych (np. MySQL), a także podstaw w tworzeniu aplikacji internetowych (m. in. HTML, CSS). Zaletą takiego rozwiązania jest pełna elastyczność a ograniczeniem są jedynie możliwości programistyczne tworzącego ankietę.

Wykorzystanie gotowego rozwiązania (komercyjnego czy darmowego) zmusza do posługiwania się narzędziem w określonym zakresie, ale zaletą jest brak wymaganej znajomości w zakresie programowania.

W niniejszej pracy autorzy pragną zaprezentować system do zarządzania badaniami ankietowymi LimeSurvey.

LimeSurvey jest bardzo popularnym oprogramowaniem ze względu na swoje liczne zalety, częste aktualizację, a zwłaszcza dlatego iż jest to oprogramowanie wolne i otwarte (licencja GNU GPL) przez co może być wykorzystywane nawet komercyjnie bezpłatnie. Jest to bardzo popularne oprogramowanie stosowane na szeroką skalę, świadczy o tym np. ilość pobrań z zasobów SourceForge (w 2011 r. ponad 150 000 razy<sup>1</sup>).

Obecną stabilną wersją LimeSurvey jest 1.92+, ze strony <http://www.limesurvey.org/en/download> można także pobrać system w wersji beta 2.0 beta 1 (stan na 12 maja 2012 r.). Do prawidłowego działania potrzebuje serwera wyposażonego w odpowiednią ilość wolnej przestrzeni dyskowej a także w odpowiedniej wersji: bazę danych oraz interpreter PHP z odpowiednimi bibliotekami i modułami. Instalację można przeprowadzić lokalnie lub na zdalnym serwerze. Dla instalacji lokalnej należy pobrać ze strony LimeSurvey wersję zintegrowaną z XAMPP (X – Cross-platform, Apache, MySQL, PHP, Perl) lub samodzielnie zainstalować lokalnie uprzednio serwer Apache z odpowiednimi rozszerzeniami. Proces instalacyjny został dokładnie omówiony w dokumentacji oprogramowania i jest on niezwykle prosty zwłaszcza od wersji 2.0. Do poprawnego działania LimeSurvey w wersji 2.0 wykorzystuje 33 tabele bazy danych.

LimeSurvey posiada liczne zalety, niektóre z nich zostały wymienione poniżej<sup>2</sup>:

- Nielimitowana liczba:
  - badań ankietowych prowadzonych w jednym czasie,
  - pytań w ankiecie (ograniczone jedynie przez bazę),
  - respondentów,
- Ankiety wielojęzyczne;
- Zarządzanie użytkownikami;
- 28 różnych typów pytań, podzielone na następujące kategorie:
  - Pytania maski:

---

<sup>1</sup> <http://www.limesurvey.org/component/nbill/?action=orders&task=order&cid=1>

<sup>2</sup> <http://www.limesurvey.org/en/about-limesurvey/features>

- Data/czas, Płeć, Ranking, Równanie, Tak/Nie, Wartość liczbową, Wartość językowa, Wielokrotna wartość liczbową, Wyświetlanie tekstu, Ładowanie pliku;
- Pytania jednego wyboru:
  - Lista (rozwijalna), Lista (radio), Lista z komentarzem, Skala 5-punktowa;
- Tabele:
  - Tabela, Tabela (Liczby), Tabela (Tak/Nie/Nie wiem), Tabela (Teksty), Tabela (Wzrost, Bez zmian, Spadek), Tabela (skala 10-punktowa), Tabela (skala 5-punktowa), Tabela dwóch skal, Tabela w kolumnach;
- Pytania tekstowe:
  - Tekst dowolny (bardzo długi), Tekst dowolny (długi), Tekst dowolny (krótki), Wielokrotny krótki tekst;
- Pytania wielokrotnego wyboru:
  - Wielokrotny wybór, Wielokrotny wybór z komentarzami;
- Integracja zdjęć i filmów w ankiecie;
- Pytania warunkowe i pytania zależne od poprzednich odpowiedzi;
- Edytor szablonów ankiet;
- Opcja dla respondentów pozwalająca dokończyć badanie później;
- Ankiety anonimowe i nieanonimowe;

Należy także nadmienić, że dzięki złożoności systemu i swoim możliwościom może być on wykorzystywany w wielu obszarach, np.: ogólne zbieranie danych (np. przez telefon w callcenter), badania marketingowe, badania satysfakcji klientów, ale także: testy psychologiczne, zarządzanie jakością, badanie kwalifikacji kandydatów. Obecnie zastosowanie LimeSurvey wykracza poza badania społeczne, bo może być wykorzystywane nawet jako: formularz zamówienia dla próbek produktów czy zarządzanie zaproszeniami dla wydarzeń (np. rezerwacja noclegów)<sup>3</sup>.

LimeSurvey został przetłumaczony na wiele języków całego świata (w tym polski). Według podanych danych<sup>4</sup> system został przetłumaczony średnio w 80% całości w 60 językach (95% całości systemu dla polskiego). Tłumaczeniem zajmują się wolontariusze. LimeSurvey wykorzystuje GNU gettext<sup>5</sup>. Jest to pakiet oprogramowania (w ramach projektu GNU) wykorzystywany do umiędzynarodowienia programów komputerowych.

W wersji 2.0 został wykorzystany wzorzec projektowy MVC (ang. *Model-View-Controller*, Model-Widok-Kontroler), skorzystano z frameworka "Yii"<sup>6</sup>. Obecnie trwa zbieranie funduszy (docelowa kwota to 50 000 \$) w celu opłacenia szybszego rozwoju LimeSurvey<sup>7</sup>.

Bardzo istotnym aspektem dla każdego oprogramowania jest integracja i współpraca z innymi programami. LimeSurvey oferuje zarówno opcję importu i eksportu do najpopularniejszych pakietów statystycznych i programów do analizy danych. Wyniki badania ankietowego (odpowiedzi respondentów można wyeksportować do:

- Microsoft Word (.doc)
- Microsoft Excel (.xls)
- Pliku .csv
- SPSS (plik poleceń .sps, plik danych .csv)
- R (plik poleceń .R, plik danych .csv)

<sup>3</sup> <http://www.limesurvey.org/en/about-limesurvey/features>

<sup>4</sup> <http://www.limesurvey.org/en/contribute/translations-status>

<sup>5</sup> <http://www.gnu.org/software/gettext/>

<sup>6</sup> <http://www.limesurvey.org/en/component/content/article/1-general-news/237-limesurvey-192-stable-and-limesurvey-20-alpha-2-released>

<sup>7</sup> <http://www.limesurvey.org/component/nbill/?action=orders&task=order&cid=1>



Należy nadmienić iż LimeSurvey posiada wbudowane narzędzia do przeglądania wyników. Istnieje możliwość wyświetlenia zagregowanych odpowiedzi respondentów z uwzględnieniem filtrów (m.in. zakres numerów identyfikacyjnych odpowiedzi, uwzględnienie zakończonych i niedokończonych odpowiedzi, wybranie poszczególnych pytań), a także prezentacja wyników np. na wykresie kołowym. Formatem wyników może być: HTML, PDF lub Microsoft Excel.

Natomiast kwestionariusz ankiety można wyeksportować do plików o następujących rozszerzeniach:

- .lss
- .zip (zawierające pliki .lss i .lssr)
- .xml
- .xls
- queXML (jako spakowane archiwum .zip)

Wyeksportowane archiwum queXML zawiera plik PDF badania, plik queXML badania i plik queXF podziału XML, który można użyć z queXF: <http://quexf.sourceforge.net/> do przetwarzania skanowanych badań.

Ponadto istnieje także możliwość wyświetlenia kwestionariusza w postaci do druku.

Interfejs graficzny LimeSurvey jest przejrzysty, mimo ogromnych możliwości i wielości opcji, dzięki „hierarchicznej” budowie. Po zalogowaniu użytkownik widzi opcje administracyjne dla systemu, tworząc nową ankietę widzi panel administracyjny oraz moduł do zarządzania ankietami. Kolejnymi modułami są grupy pytań oraz pytania. W pytaniach można zarządzać tzw. subpytaniami, czyli wariantami odpowiedzi.

Bardzo ważnym aspektem są możliwości wyświetlenia kwestionariusza ankiety dla respondenta. LimeSurvey dysponuje domyślnie kilkoma szablonami kwestionariusza ankiet. Przygotowanie własnego szablonu nie jest prostym zadaniem, wymaga bardzo dobrej znajomości organizacji szablonów w systemie i CSS. Każdy szablon ma własny katalog, w którym zawarte są 3 podstawowe rodzaje plików: szablony (.pstpl), pliki arkuszy styli (.css), obrazki (np. .png). Fragmenty kodu HTML przechowywane są w plikach .pstpl, wykorzystywane są do konstrukcji strony dla użytkownika końcowego w trakcie wywołań jego akcji na ankiecie. W celu odwoływania się do poszczególnych elementów ankiety należy posługiwać się słowami kluczowymi umieszczonymi w nawiasach klamrowych (np. {SURVEYNAME})<sup>8</sup>.

Możliwości LimeSurvey nie ograniczają się do stworzenia kwestionariusza i eksportu wyników. Jest to system wspomagający cały proces prowadzenia badań ankietowych. Istnieje możliwość stworzenia nowych użytkowników (z odpowiednimi prawami) lub nowej grupy użytkowników i rozesłanie do nich informacji. Jest więc to także platforma wymiany informacji pomiędzy członkami zespołu prowadzącymi badania<sup>9</sup>.

Ankieter ma także wpływ na to, kto może wypełniać ankiety dzięki wykorzystanemu systemowi tokenów. Tokeny można wprowadzić jedynie dla nieanonimowych badań. Ankieter ma możliwość importu pliku .csv lub kwerendy LDAP listy uczestników badania (nazwa, adresy email) lub dodania ręcznie nowych tokenów. Ponadto ankieter może wysłać zaproszenia do badania lub przypomnienia dla uczestników badania, którzy jeszcze nie wzięli w nim udziału.

Kolejnym ograniczeniem odnośnie grupy respondentów są tzw. kwoty. Można w nich zdefiniować limit odpowiedzi respondentów na dane pytanie. Kolejni respondenci, którzy przekroczą ustalony limit odpowiedzi zostaną o tym poinformowani odpowiednim komunikatem, a badanie zostanie przerwane<sup>10</sup>.

Zaawansowaną funkcjonalnością LimeSurvey jest zależność pytań od udzielonych uprzednio odpowiedzi. Od wersji 1.92+ wprowadzono tzw. Expression Manager umożliwiający odwoływanie się do pytania wyświetlanego na tej samej stronie.

---

<sup>8</sup> <http://docs.limesurvey.org/The+template+editor>

<sup>9</sup> <http://docs.limesurvey.org/Zarządzanie+użytkownikami>

<sup>10</sup> <http://docs.limesurvey.org/Quotas>

### 3. Podstawy teoretyczne analizy niezależności zmiennych

Analiza zbiorowości statystycznej oparta jest w dużej mierze na badaniu nie jednej a wielu cech ją charakteryzujących. Istotnym zagadnieniem staje się zatem możliwość sprawdzenia, czy między analizowanymi zmiennymi zachodzi zależność.

Działem statystyki obejmującym swym zakresem analizę związków między badanymi zmiennymi jest teoria współzależności. Wyróżnia ona dwa rodzaje współzależności między cechami populacji. Pierwsza to zależność funkcyjna, której istotę stanowi to, iż określonej wartości jednej zmiennej, nazywanej zmienną niezależną i oznaczanej najczęściej jako  $X$ , odpowiada dokładnie jedna wartość drugiej zmiennej – zależnej ( $Y$ ). Drugim rodzajem jest zależność stochastyczna. Ma ona miejsce, gdy zmiana jednej ze zmiennych wpływa na rozkład prawdopodobieństwa drugiej zmiennej. Jej szczególnym przypadkiem jest zależność korelacyjna, nazywana także statystyczną, pozwalająca określić jak średnio rzecz biorąc zmienia się wartość analizowanej cechy w zależności od wartości innej zmiennej<sup>11</sup>.

W przypadku, gdy poddawane analizie cechy są mierzalne, można zastosować pojęcie korelacji lub regresji jednakże, kiedy choć jedna z nich jest niemierzalna istnieje konieczność posłużenia się kategorią niezależności stochastycznej<sup>12</sup>.

W przypadku badania niezależności stochastycznej jednym z najczęściej stosowanych testów jest test niezależności  $\chi^2$ . Podstawę do analizy stanowi tablica kontyngencji, o  $r$  wierszach odpowiadających wariantom pierwszej zmiennej oraz  $k$  kolumnach dla kategorii drugiej cechy. Wnętrze tabeli wypełniają, zatem liczebności  $n_{ij}$  odpowiednich kombinacji wartości zmiennych dla  $i$ -tego wiersza oraz  $j$ -tej kolumny. Najczęściej zakłada się, iż liczebności te powinny być co najmniej równe 5. Hipotezę zerową tego testu formalnie można zapisać jako  $H_0 : E(n_{ij}) = E(\hat{n}_{ij})$ , natomiast alternatywną  $H_1 : E(n_{ij}) \neq E(\hat{n}_{ij})$ , gdzie przez  $\hat{n}_{ij}$  oznaczono liczebności teoretyczne spełniające warunek stochastycznej niezależności, a  $E$  jest operatorem wartości oczekiwanej<sup>13</sup>. Liczebności teoretyczne wyznaczone są z następującego wzoru:

$$\hat{n}_{ij} = \frac{n_i \cdot n_j}{n}, \quad (1)$$

gdzie jako  $n_i$  oraz  $n_j$  oznaczono liczebności brzegowe, będące sumą liczebności odpowiednio  $i$ -tego wiersza lub  $j$ -tej kolumny<sup>14</sup>. Statystyka omawianego testu ma postać:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^r \frac{(n_{ij} - \hat{n}_{ij})^2}{\hat{n}_{ij}}. \quad (2)$$

Przy założeniu prawdziwości  $H_0$  ma ona asymptotyczny rozkład  $\chi^2$  o  $(r-1)(k-1)$  stopniach swobody i prawostronnym obszarze krytycznym określonym nierównością  $\chi^2 \geq \chi_\alpha^2$ . Przez  $\chi^2$  z indeksem  $\alpha$  oznaczona została wartość krytyczna odczytana z tablic rozkładu  $\chi^2$  dla ustalonego poziomu istotności np. 0,05 lub 0,01<sup>15</sup>.

Test ten znajduje jednakże zastosowanie tylko w przypadku analizy jedynie dwóch cech. By możliwe było zbadanie łącznie zależności trzech lub więcej zmiennych konieczne jest zastosowanie wielowymiarowych tablic kontyngencji oraz uogólnionej postaci testu niezależności  $\chi^2$ .

Ogromne znaczenie możliwości badania współzależności między więcej niż dwoma zmiennymi potwierdza paradoks Simpsona. Uwidacznia on, iż w wielu przypadkach kierunek lub siła związku między dwiema cechami może zależeć od kolejnych zmiennych. Przykładowo związek

<sup>11</sup> Sobczyk M.: Statystyka, PWN, Warszawa 2001, s. 207

<sup>12</sup> Greń J.: Statystyka matematyczna, modele i zadania, PWN, Warszawa 1978, s. 130

<sup>13</sup> Sobczyk M.: Statystyka, PWN, Warszawa 2001, s. 213-214

<sup>14</sup> Por. Greń J.: Statystyka matematyczna, modele i zadania, PWN, Warszawa 1978, s. 131

<sup>15</sup> Sobczyk M.: Statystyka. PWN, Warszawa 2001, s. 213-214

między skutecznością leczenia pacjentów (określoną kategoriami sukces, porażka) oraz metodą leczenia, oznaczoną przez A lub B, może w dużej mierze zależeć także od miejsca leczenia pacjentów (związanego z tym doświadczenia lekarzy)<sup>16</sup>.

W niniejszym artykule poruszony zostanie jedynie przypadek uogólnienia testu niezależności  $\chi^2$  dla trzech zmiennych. Bazą do analizy związku między trzema cechami zbiorowości jest tablica kontyngencji składająca się z  $r \times k \times l$  komórek. Odpowiednio kolejne z wymiarów tabeli oznaczają liczbę wariantów zmiennej rzędu, kolumny oraz warstwy<sup>17</sup>.

W analizie trzech cech można wyróżnić siedem wariantów konceptualizacji relacji między nimi. Pierwszy z nich zakłada, iż wszystkie z analizowanych zmiennych są względem siebie niezależne. W kolejnych trzech jedna z cech jest niezależna od pozostałych dwóch zmiennych, które względem siebie nie muszą być już niezależne. Przykładowo do tej grupy zaliczany jest przypadek, gdy zmienna kolumny jest niezależna od zmiennej rzędu oraz warstwy. Ostatnie trzy warianty to niezależność warunkowa jednej ze zmiennych względem kolejnych<sup>18</sup>.

Analiza uwzględniająca pierwszy ze wspomnianych powyżej wariantów określana jest modelem całkowitej niezależności. Hipoteza zerowa oraz alternatywna dla tego wariantu są modyfikacją hipotez dla testu niezależności  $\chi^2$  dla tablic kontyngencji  $r \times k$ , zatem korzystając z wcześniej zastosowanych oznaczeń, można je zapisać w następującej postaci:

$$H_0 : E(n_{ijm}) = E(\hat{n}_{ijm})$$

$$H_1 : E(n_{ijm}) \neq E(\hat{n}_{ijm}).$$

Odpowiednio dostosowując również wzór (2) do potrzeb omawianej analizy otrzymujemy statystykę testu dla trzech zmiennych:

$$\chi^2 = \sum_{i=1}^k \sum_{j=1}^r \sum_{m=1}^l \frac{(n_{ijm} - \hat{n}_{ijm})^2}{\hat{n}_{ijm}}, \quad (3)$$

gdzie odpowiednio przez  $n_{ijm}$  i  $\hat{n}_{ijm}$  oznaczone zostały liczebności obserwowane i teoretyczne dla kolejnych kombinacji  $i$ -tego wiersza,  $j$ -tej kolumny oraz  $m$ -tej warstwy. Liczebności niezbędne do wyznaczenia powyższej miary obliczyć można na podstawie wzoru:

$$\hat{n}_{ijm} = \frac{n_{i..} n_{.j.} n_{..m}}{n^2}, \quad (4)$$

$n_{i.}$ ,  $n_{.j.}$  oraz  $n_{..m}$  są liczebnościami brzegowymi analogicznie jak dla testu dla dwóch zmiennych  $i$ -tego wiersza,  $j$ -tej kolumny i  $m$ -tej warstwy. Statystyka  $\chi^2$  dla wzoru (3) ma rozkład  $\chi^2$  o  $[(rkl) - r - k - l + 2]$  stopniach swobody<sup>19</sup>.

Model częściowej niezależności w odróżnieniu od modelu całkowitej niezależności pozwala stwierdzić czy jedna, wybrana zmienna jest niezależna w stosunku do pozostałych. Zatem analizując trzy cechy zbiorowości możliwe jest określenie czy:

- a) zmienna rzędu jest niezależna względem zmiennej kolumny oraz warstwy;
- b) zmienna kolumny jest niezależna względem zmiennej rzędu oraz warstwy;
- c) zmienna warstwy jest niezależna względem zmiennej rzędu oraz kolumny.

Istotną różnicę w stosunku do wcześniej wspomnianego modelu pełnej niezależności stanowi sposób wyznaczenia liczebności teoretycznych. Dla kolejnych wariantów korzystamy odpowiednio z następujących wzorów:

<sup>16</sup> Sheskin D. J.: Handbook of Parametric and nonparametric statistical procedures, CHAPMAN & HALL/CRC, Boca Raton London New York Washington, D.C. 2000, s. 421-422

<sup>17</sup> Por. Sheskin D. J.: Handbook of Parametric and nonparametric statistical procedures, CHAPMAN & HALL/CRC, Boca Raton London New York Washington, D.C. 2000, s. 425

<sup>18</sup> Christensen R.: Log-linear models, Springer-Verlag, New York 1990, 63-64

<sup>19</sup> Por. Sheskin D. J.: Handbook of Parametric and nonparametric statistical procedures, CHAPMAN & HALL/CRC, Boca Raton London New York Washington, D.C. 2000, s. 426-428

$$\hat{n}_{ijm} = \frac{n_{i.}(n_{.j}n_{.m})}{n} \quad (5a)$$

$$\hat{n}_{ijm} = \frac{n_{.j}(n_{i.}n_{.m})}{n} \quad (5b)$$

$$\hat{n}_{ijm} = \frac{n_{.m}(n_{i.}n_{.j})}{n} \quad (5c)$$

Fragmentem w znaczący sposób różniącym wzory (5a-5c) od ich odpowiednika dla całkowitej analizy niezależności trójwymiarowej tablicy kontyngencji jest element objęty nawiasem. W odróżnieniu od wzoru (4) gdzie iloczynowi podlegały trzy całkowite liczebności brzegowe dla zadanej kategorii zmiennej rzędu, kolumny i warstwy, we wzorach oznaczonych od (5a) do (5c) w liczniku umieszczane są dwa czynniki. Pierwszym elementem jest brzegowa liczebność całkowita dla zmiennej, której dotyczy założenie niezależności względem pozostałych. W nawiasie ujęto natomiast liczebność brzegową łączną dla danej kombinacji kategorii pozostałych dwóch cech zbiorowości. W pierwszym przypadku wymnażana jest całkowita liczebność brzegowa zmiennej rzędu oraz odpowiednia łączna liczebność brzegowa dla kombinacji  $j$ -tej kategorii zmiennej kolumny i  $m$ -tej – zmiennej warstwy, w drugim będzie to analogicznie całkowita liczebność brzegowa zmiennej kolumny i łączna liczebność brzegowa dla kombinacji  $i$ -tego wariantu zmiennej rzędu i  $m$ -tego dla zmiennej warstwy. Stosując ostatni z podanych wzorów w liczniku należałoby umieścić całkowitą liczebność brzegową zmiennej warstwy i łączną liczebność brzegową dla kombinacji  $i$ -tej kategorii zmiennej rzędu i  $j$ -tej – zmiennej kolumny. W każdym ze wzorów w mianowniku umieszczamy jednakże  $n$  – liczebność całkowitą. Przykładowo analizując trzy zmienne, z których każda posiada dwa warianty, w każdym z trzech omawianych przypadków otrzymamy osiem wartości liczebności teoretycznych. Po cztery odpowiadające poszczególnym wariantom zmiennej, której niezależność badamy. W każdym z nich ujęta powinna zostać inna kombinacja kategorii pozostałych dwóch zmiennych.

Kolejny etap analizy – wyznaczenie statystyki testu przebiega w analogiczny sposób jak w modelu całkowitej niezależności. Statystyka  $\chi^2$  dla wzorów (5a)-(5c) ma, więc rozkład  $\chi^2$  o:

a)  $[(rkl)-kl-r+1]$

b)  $[(rkl)-rl-k+1]$

c)  $[(rkl)-rk-l+1]$

stopniach swobody<sup>20</sup>. Zarówno w przypadku pełnej jak i częściowej niezależności stosowany jest prawostronny obszar krytyczny. Zatem jeżeli znajdzie nierówność  $\chi^2 \geq \chi^2_{\alpha}$  należy odrzucić hipotezę  $H_0$  o niezależności wszystkich analizowanych zmiennych lub wybranej względem pozostałych dwóch.

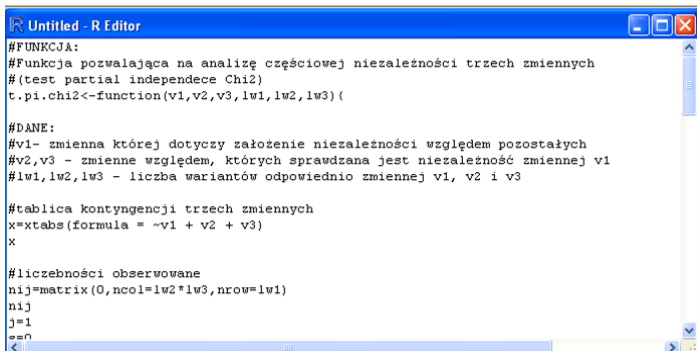
#### 4. Zastosowanie programu R w badaniu niezależności zmiennych

Program R pozwala na dosyć szeroką analizę w zakresie badania niezależności cech. Dzięki wbudowanej w pakiet *stats* funkcji *chisq.test* możliwa jest analiza zależności między dwiema zmiennymi – zgodna z podstawowym wariantem testu  $\chi^2$  niezależności (wzory (1) oraz (2)). Jednakże znacznie większe możliwości daje użytkownikowi programu funkcja *wtd.chi.sq* będąca jednym ze składników pakietu *weights*. Umożliwia ona badanie zależności zachodzących pomiędzy trzema zmiennymi, zgodnie z wzorami (3) i (4). Odpowiada ona zatem modelowi pełnej niezależności trzech zmiennych. Dodatkowo może ona zostać również zastosowana do zbadania zależności jedynie dwóch cech.

By możliwe było jednakże poddanie analizie częściowej niezależności badanych zmiennych konieczne jest zaprogramowanie przez użytkownika dodatkowej funkcji w języku R. Tworzenie

<sup>20</sup> Por. Sheskin D. J.: Handbook of Parametric and nonparametric statistical procedures, CHAPMAN & HALL/CRC, Boca Raton London New York Washington, D.C. 2000, s. 428-434

własnych funkcji umożliwiła użytkownikowi procedura *function*, której zastosowanie zostało przedstawione na rysunku 1.



```
Untitled - R Editor
#FUNKCJA:
#Funkcja pozwalająca na analizę częściowej niezależności trzech zmiennych
#(test partial independence Chi2)
t.pi.chi2<-function(v1,v2,v3,lw1,lw2,lw3){

#DANE:
#v1- zmienna której dotyczy założenie niezależności względem pozostałych
#v2,v3 - zmienne względem, których sprawdzana jest niezależność zmiennej v1
#lw1,lw2,lw3 - liczba wariantów odpowiednio zmiennej v1, v2 i v3

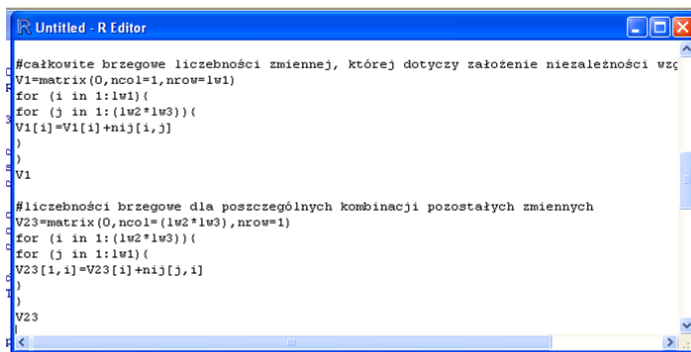
#tablica kontyngencji trzech zmiennych
x=xtabs(formula = ~v1 + v2 + v3)
x

#liczebności obserwowane
nij=matrix(0,ncol=lw2*lw3,nrow=lw1)
nij
j=1
:=0
```

Rysunek 1. Okno nowego skryptu – zastosowanie procedury *function*.

Zmiennej *t.pi.chi2* przypisana została funkcja o pięciu argumentach, z których pierwsze trzy odpowiadają analizowanym cechom – *v1*, *v2*, *v3*, pozostałe natomiast liczbie wariantów poszczególnych cech (*lw1*, *lw2*, *lw3*). Dodatkowo istotnym założeniem dla omawianej funkcji jest, to, iż jako pierwsza podawana jest zmienna, której niezależność względem pozostałych będzie badana.

Kolejnym etapem jest utworzenie tablicy kontyngencji dla wszystkich zmiennych z wykorzystaniem funkcji *xtabs* (wchodzącej w skład pakietu *stats*) i przekształcenie na macierz o *lw1* wierszach i *lw2\*lw3* kolumnach. W następnym kroku wyznaczane są całkowite liczebności brzegowe dla zmiennej, której dotyczy założenie niezależności oraz liczebności brzegowe dla poszczególnych kombinacji pozostałych zmiennych (rysunek 2).



```
Untitled - R Editor
#całkowite brzegowe liczebności zmiennej, której dotyczy założenie niezależności w z
V1=matrix(0,ncol=1,nrow=lw1)
for (i in 1:lw1){
  for (j in 1:(lw2*lw3)){
    V1[i]=V1[i]+nij[i,j]
  }
}
V1

#liczebności brzegowe dla poszczególnych kombinacji pozostałych zmiennych
V23=matrix(0,ncol=(lw2*lw3),nrow=1)
for (i in 1:(lw2*lw3)){
  for (j in 1:lw1){
    V23[1,i]=V23[i]+nij[j,i]
  }
}
V23
```

Rysunek 2. Okno nowego skryptu – wyznaczenie liczebności brzegowych.

Ich wyznaczenie pozwala na wyliczenie wartości teoretycznych, a także statystyki testu zgodnie z wzorami (4) oraz (5a-5c). Ostatnimi elementami, które wymagają zaprogramowania są liczba stopni swobody oraz *p*-wartość. Deklaracja tych zmiennych została przedstawiona na rysunku 3.

```

Chi2=sum(X)
Chi2

#liczba stopni swobody
df=1*w1*lw2*lw3-lw2*lw3-lw1+1
df

#p-wartość
p_wartość=pchisq(Chi2,df,lower.tail = FALSE)

wynik<-matrix(c(Chi2,df, p_wartość),ncol=1,nrow=3)
nazwa<-list(c("Chi2","df","p-wartość"),c("Wartość"))
dimnames(wynik)<-nazwa
wynik
)
t.p1.chi2 (v1,v2,v3,lw1,lw2,lw3)

```

Rysunek 3. Okno nowego skryptu – wyznaczenie liczby stopni swobody i  $p$ -wartości oraz deklaracja zmiennej `wynik`.

Do wyznaczenia  $p$ -wartości wykorzystana została funkcja `pchisq`, której argumentami są wartość statystyki testu oraz liczba stopni swobody –  $df$ .

Po odwołaniu się do opisywanej funkcji zwracane są trzy wartości przypisane do zmiennej `wynik`: statystyka testu, liczba stopni swobody oraz  $p$ -wartość. Dla zwiększenia przejrzystości i czytelności otrzymywanych wyników każda z uzyskiwanych wartości opatrzona została etykietą.

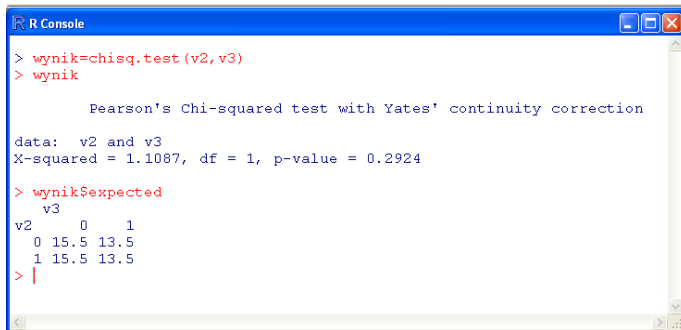
Pełny kod omawianej funkcji zawarty został w załączniku 1. W kolejnej części opracowania zostanie przedstawiony przykład wykorzystania tej funkcji na danych empirycznych.

## 5. Przykład empiryczny

Spośród danych zgromadzonych dzięki ankiecie, opisaney w pierwszej części artykułu, do niniejszej analizy wybrane zostały trzy zmienne:

- $v_1$  – odnosząca się do wyboru matematyki jako przedmiotu zdawanego na maturze (wartość 1 przyporządkowana gdy przedmiot ten był zdawany na poziomie rozszerzonym, 0 – jeśli student zdawał go na poziomie podstawowym lub nie zdawał w ogóle)
- $v_2$  – dotycząca oceny przydatności informacji o dostępie do specjalistycznego oprogramowania podczas zajęć na kierunku Analityka Gospodarcza (1 – gdy przyznano ocenę najwyższą („bardzo ważne”), 0 – w pozostałych przypadkach – zastosowana została skala pięciostopniowa : bardzo ważne, raczej ważne, nie mam zdania, raczej nieważne, zupełnie nieważne)
- $v_3$  – wyrażająca chęć podjęcia przez studenta pracy na stanowisku analityk w przedsiębiorstwach krajowych i międzynarodowych (kategoria 1 przypisana gdy wybrany został wariant „bardzo”, 0 – dla odpowiedzi: „przeciętnie”, tylko z konieczności”, „wcale”)

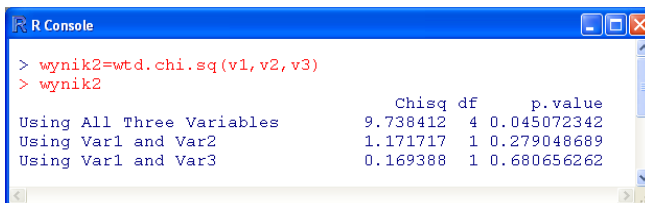
W pierwszym etapie zbadana została zależność między zmiennymi  $v_2$  oraz  $v_3$ . W tym celu zastosowano gotową funkcję `chisq.test` – rysunek 4.



Rysunek 4. Wyniki analizy zależności dwóch zmiennych – funkcja `chisq.test`.

W wyniku przeprowadzonej analizy można stwierdzić, iż nie ma podstaw do odrzucenia hipotezy zerowej testu, uzyskana p-wartość jest wyższa od zakładanego poziomu 0,05. Nie można, zatem stwierdzić zależności między badanymi zmiennymi – najwyższą oceną przydatności informacji o dostępie do specjalistycznego oprogramowania a chęcią podjęcia pracy na stanowisku analityk. Dodatkowo sprawdzone zostało założenie dotyczące liczebności teoretycznych, po przez odwołanie `wynik$expected` – zmiennej `wynik` przypisano zastosowaną funkcję. Dla wszystkich kombinacji wariantów zmiennych liczebności te są wyższe niż 5.

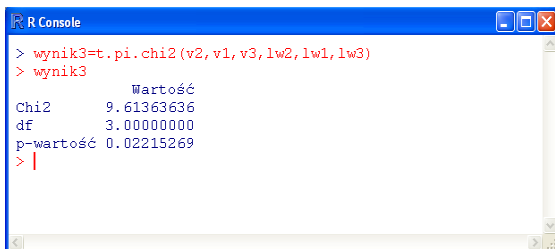
W kolejnym kroku analizy zbadana została zależność między wszystkimi trzema badanymi cechami. Do analizy pełnej niezależności zastosowano funkcję `wtd.chi.sq`. Otrzymane rezultaty przedstawiono na rysunku 5.



Rysunek 5. Wyniki analizy pełnej niezależności trzech zmiennych – funkcja `wtd.chi.sq`

Uzyskana p-wartość (niższa od przyjętego poziomu  $\alpha=0,05$ ) pozwala na stwierdzenie, iż należy odrzucić hipotezę  $H_0$  o braku zależności między analizowanymi zmiennymi. Również w tym przypadku zostało sprawdzone i spełnione założenie o liczebnościach teoretycznych nie mniejszych niż 5.

Do zbadania częściowej niezależności wykorzystano funkcję `t.pi.chi2` opisaną w części poruszającej problematykę zastosowania programu R. Sprawdzona została niezależność zmiennej  $v_2$  względem zmiennych  $v_1$  oraz  $v_3$  – rysunek 6.



Rysunek 6. Wyniki analizy częściowej niezależności dla zmiennej  $v_2$  – funkcja `t.pi.chi2`

W związku z niewielką liczebnością badanej zbiorowości (liczba ankietowanych studentów wyniosła 58), w przypadku analizy częściowej zależności dla zmiennych  $v_1$ ,  $v_3$  nie zostało spełnione założenie dotyczące liczebności teoretycznych. Gdyby założenie to byłoby spełnione, na podstawie uzyskanych wyników można byłoby stwierdzić istnienie zależności między zmienną  $v_2$ , a pozostałymi cechami podlegającymi analizie. Uzyskana p-wartość nie przekroczyła założonego poziomu 0,05.

## 6. Podsumowanie

Artykuł prezentuje problematykę analizy zależności trzech zmiennych z zastosowaniem wbudowanych oraz samodzielnie zaprogramowanych funkcji programu R. W przykładzie empirycznym wykorzystano dane dotyczące studentów kierunku Analityka Gospodarcza na Uniwersytecie Ekonomicznym w Katowicach zebrane przez Koło Naukowe Statystyków z wykorzystaniem systemu zarządzania badaniami ankietowymi LimeSurvey.

## Bibliografia

1. Christensen, R.: *Log-linear models*, Springer-Verlag, New York 1990.
2. Greń J.: *Statystyka matematyczna, modele i zadania*, PWN, Warszawa 1978.
3. Sheskin D. J.: *Handbook of Parametric and nonparametric statistical procedures*, CHAPMAN & HALL/CRC, Boca Raton London New York Washington, D.C. 2000.
4. Sobczyk M.: *Statystyka*. PWN, Warszawa 2001.

## Netografia

1. <http://docs.limesurvey.org/Quotas>
2. <http://docs.limesurvey.org/The+template+editor>
3. <http://docs.limesurvey.org/Zarzadzanie+uzytkownikami>
4. <http://www.gnu.org/software/gettext/>
5. <http://www.limesurvey.org/component/nbill/?action=orders&task=order&cid=1>
6. <http://www.limesurvey.org/en/about-limesurvey/features>
7. <http://www.limesurvey.org/en/component/content/article/1-general-news/237-limesurvey-192-stable-and-limesurvey-20-alpha-2-released>
8. <http://www.limesurvey.org/en/contribute/translations-status>

## On Test of Independence of Three Variables Based on Some Real Data

*LimeSurvey is a free and open source powerful survey management system. In the paper the functions of the system and the areas of its application are presented. Authors described construction of LimeSurvey and technologies used within it. The system was used in the survey research of students at the University of Economics in Katowice. Analysis of independence for three dimensional contingency tables – model of complete and partial independence – are presented as well including theoretical bases of the problem and an empirical example. The purpose of the article is to present a function in R language– built-in and prepared by the authors.*

**Keywords:** CAWI, CAII, LimeSurvey, stochastic independence, Chi-square test of independence



## Załącznik 1.

```
#FUNKCJA:
#Funkcja pozwalająca na analizę częściowej
niezależności trzech zmiennych (test partial
independence Chi2)
t.pi.chi2<-function(v1,v2,v3,lw1,lw2,lw3){

#DANE:
#v1- zmienna której dotyczy założenie
niezależności względem
pozostałych
#v2,v3 - zmienne względem, których sprawdzana
jest niezależność zmiennej
v1
#lw1,lw2,lw3 - liczba wariantów odpowiednio
zmiennej v1, v2 i v3

#tablica kontyngencji trzech zmiennych
x=xtabs(formula = ~v1 + v2 + v3)
x

#alfa - poziom istotności
alfa=0.05

#liczebności obserwowane
nij=matrix(0,ncol=lw2*lw3,nrow=lw1)
nij
j=1
s=0
while (j<=(lw2*lw3)){
for (i in 1:lw1){
nij[i,j]=x[i+s]
}
j=j+1
s=s+lw1
}
nij

# całkowita liczebność
N=sum(nij)

#całkowite brzegowe liczebności zmiennej, której
dotyczy założenie niezależności względem
pozostałych
V1=matrix(0,ncol=1,nrow=lw1)
for (i in 1:lw1){
for (j in 1:(lw2*lw3)){
V1[i]=V1[i]+nij[i,j]
}
}
}
```

```
V1
#liczebności brzegowe dla poszczególnych
kombinacji pozostałych
zmiennych
V23=matrix(0,ncol=(lw2*lw3),nrow=1)
for (i in 1:(lw2*lw3)){
for (j in 1:lw1){
V23[1,i]=V23[i]+nij[j,i]
}
}
V23

#liczebności teoretyczne
nijt=matrix(0,ncol=lw2*lw3,nrow=lw1)
for (i in 1:lw1){
for (j in 1:(lw2*lw3)){
nijt[i,j]=(V1[i]*V23[1,j])/N
}}
nijt

#wartość statystyki testu Chi2 dla częściowej
niezależności 3 zmiennych
X=matrix(0, ncol=(lw2*lw3), nrow=lw1)
X
for (i in 1:lw1){
for (j in 1:(lw2*lw3)){
X[i,j]=((nij[i,j]-nijt[i,j])^2)/nijt[i,j]
}}
X

Chi2=sum(X)
Chi2

#liczba stopni swobody
df=lw1*lw2*lw3-lw2*lw3-lw1+1
df

#p-wartość
p_wartość=pchisq(Chi2,df,lower.tail = FALSE)

wynik<-matrix(c(Chi2,df,
p_wartość),ncol=1,nrow=3)
nazwa<-list(c("Chi2","df","p-
wartość"),c("Wartość"))
dimnames(wynik)<-nazwa
wynik
}
t.pi.chi2(v1,v2,v3,lw1,lw2,lw3)
```



**Małgorzata Krzciuk**

Posiada tytuł zawodowy licencjat uzyskany na uniwersytecie Ekonomicznym na kierunku Informatyka i Ekonometria – specjalność Ekonometria i statystyka. Jest studentką studiów stacjonarnych drugiego stopnia na tej samej uczelni na Wydziale Finansów i ubezpieczeń, na kierunku Finanse i Rachunkowość – specjalność Analityk finansowy. Należy do Koła Naukowego Statystyków przy Katedrze Statystyki Uniwersytetu Ekonomicznego w Katowicach.



**Piotr Ziuziański**

Posiada tytuł zawodowy licencjat uzyskany na Uniwersytecie Ekonomicznym w Katowicach na Wydziale Zarządzania na kierunku Informatyka i Ekonometria – specjalność Ekonometria i Statystyka. Jest studentem studiów stacjonarnych drugiego stopnia kierunku Informatyka i ekonometria na Wydziale Informatyki i komunikacji w Katowicach oraz studentem pierwszego stopnia kierunku Informatyka na tym samym Wydziale. Jest członkiem Koła Naukowego Statystyków przy Katedrze Statystyki Uniwersytetu Ekonomicznego w Katowicach.