

Małgorzata FURMANKIEWICZ, Piotr ZIUZIAŃSKI
Uniwersytet Ekonomiczny w Katowicach
Wydział Informatyki i Komunikacji
Koło Naukowe Scientia Ingenium przy Katedrze Inżynierii Wiedzy
malgorzata.furmankiewicz@gmail.com, piotrziuzianski@gmail.com

Małgorzata K. KRZCIUK
Uniwersytet Ekonomiczny w Katowicach
Wydział Zarządzania
malgorzata.krzciuk@uekat.pl

ZASTOSOWANIE METODY REPREZENTACYJNEJ W ANALIZIE DANYCH ANKIETOWYCH. STUDIUM PRZYPADKU MONITOROWANIA EPIDEMII PRZEZ INTERNAUTÓW

Streszczenie. Artykuł porusza tematykę monitorowania zjawisk epidemiologicznych wśród internautów. Sklasyfikowano rodzaje źródeł informacji dotyczących epidemiologii w Polsce i zaprezentowano ich przykłady. Autorzy przedstawili wyniki przeprowadzonych badań ankietowych z lat 2013-2015, które nawiązują do tej tematyki. Do analizy danych statystycznych z 2015 roku zastosowano metodę reprezentacyjną. Populację internautów podzielono na warstwy ze względu na: zamieszkiwane województwo, płeć oraz wiek.

Słowa kluczowe: metoda reprezentacyjna, epidemiologia, e-zdrowie, pakiet R, badania ankietowe.

THE USE OF SURVEY SAMPLING IN SURVEY DATA ANALYZING. CASE STUDY OF MONITORING EPIDEMIC BY INTERNET USERS

Summary. The article describes topic of Internet usage to monitor epidemic. Polish epidemiological source of information in this area have been classified and its examples have been given. Authors present results of their research conducted in 2013-2015 connected with this subject. Sampling survey method has been used to analyze data obtained in 2015. The division of the population of Internet users was made according: the voivodship, sex and age group.

Keywords: Survey sampling, epidemiology, e-health, R language, survey research.

1. Wstęp

Internet jako coraz łatwiej dostępne medium stanowi źródło różnorodnych informacji dla coraz szerszego odsetka społeczeństwa w Polsce. Autorzy artykułu skupili swoją uwagę na wykorzystaniu sieci Internet do monitorowania zjawisk epidemiologicznych. Należy zwrócić uwagę, że monitorowanie zachorowalności społeczeństwa pełni ważną funkcję w profilaktyce zdrowotnej i stanowi istotny temat. Dzięki monitorowaniu sytuacji epidemiologicznej możliwe jest przedsięwzięcie odpowiednich środków zaradczych, które mogą powstrzymać rozprzestrzenianie się konkretnego zjawiska epidemiologicznego.

Na samym wstępie warto przytoczyć definicję epidemiologii, którą można określić jako dyscyplinę medyczną, której zadaniem jest badanie wpływu rozmaitych determinant i warunków środowiskowych na częstość występowania i szerzenie się zarówno chorób zakaźnych, jak i niezakaźnych lub ogólnie różnych stanów fizjologicznych ludności [1].

Celem niniejszego artykułu jest zaprezentowanie możliwości monitorowania zjawisk epidemiologicznych, z uwzględnieniem rządowych i prywatnych internetowych źródeł informacji. Ponadto, celem jest zaprezentowanie wyników przeprowadzonych badań ankietowych z lat 2013-2015, dotyczących tematyki monitorowania zjawisk epidemiologicznych wśród internautów.

2. Internet jako źródło danych epidemiologicznych

Informacje na temat zjawisk epidemiologicznych w Internecie można pozyskać ze źródeł prywatnych lub rządowych. W Polsce odpowiednie służby publikują na swoich stronach internetowych tzw. meldunki epidemiologiczne. Natomiast do prywatnych źródeł informacji można zaliczyć: portale informacyjne, wortale medyczne i poświęcone tematyce zdrowia, a także wyspecjalizowane systemy, które na swoich stronach publikują dane epidemiologiczne. Tabela 1 prezentuje podział źródeł informacji dotyczących zjawisk epidemiologicznych wraz z charakterystyką i nazwą podmiotu, który dane te publikuje na swoich stronach.

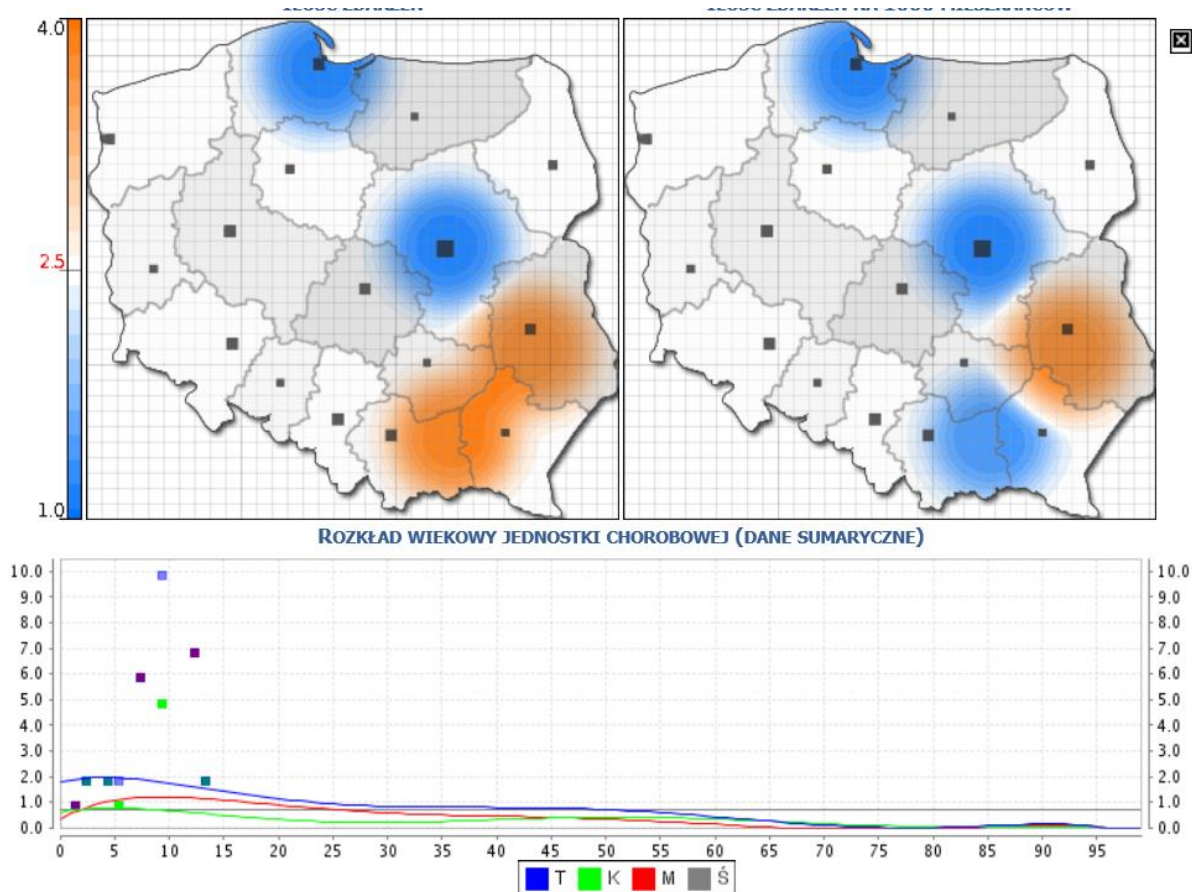
Warto nadmienić, że dane dotyczące zjawisk epidemiologicznych mogą być zaprezentowane w różnorodny sposób. Przykładowo, przedsiębiorstwo informatyczne Kamssoft S.A. w ramach Ogólnopolskiego Systemu Ochrony Zdrowia (OSOZ) publikuje tzw. Mapy Zdrowotne Kraju. Na rys. 1 zaprezentowano przykładowy zrzut ekranu Map Zdrowotnych Kraju, dotyczący zachorowalności na ostre zapalenie gardła.

Tabela 1

Źródła informacji dotyczących zjawisk epidemiologicznych wraz z charakterystyką

Rodzaj źródła	Podmiot	Źródło informacji	Charakterystyka
Rządowe	Inspekcja sanitarna	Meldunki epidemiologiczne	Meldunki o zachorowaniach na choroby zakaźne, zakażeniach i zatruciach w Polsce gromadzi Pracownia Monitorowania i Analizy Sytuacji Epidemiologicznej Zakładu Epidemiologii Narodowego Instytutu Zdrowia Publicznego. Następnie są one wysyłane do Zakładu Epidemiologii Państwowego Zakładu Higieny Narodowego Instytutu Zdrowia Publicznego przez Wojewódzkie Stacje Sanitarно-Epidemiologiczne. Dane te dotyczą lat od 1996 r. do bieżącego roku, dostępne są natomiast w ujęciach rocznym, półrocznym, kwartalnym oraz dwutygodniowym. W meldunkach zawarte są informacje dotyczące liczby zachorowań oraz zachorowalności na 100 tys. ludności w danym okresie, w odniesieniu do poprzedniego okresu, na daną jednostkę chorobową.
	WHO	Repozytorium danych WHO Global Health Observatory	Udostępnianie danych epidemiologicznych dotyczących np. zakażenia wirusem HIV na świecie.
Prywatne	Google Inc.	Google Flu Trends	Udostępnienie w formie kartogramu informacji o tendencjach aktywności wirusa grypy na świecie. Istnieje możliwość przeglądania danych od 2004 roku. Monitorowanie zachorowań bazuje na wyszukiwanych frazach w najpopularniejszej na świecie przeglądarce (Google). W Polsce projekt dostępny jest od 2009 roku. Dane dostępne są także w formacie CSV.
		Google Dengue Trends	Przedstawienie w formie kartogramu informacji o tendencjach rozprzestrzeniania się wirusa dengi na świecie.
	Kamssoft S.A.	Mapy zdrowotne kraju w ramach OSOZ	OSOZ to kompleksowy projekt, którego zadaniem jest zarządzanie opieką zdrowotną w Polsce dzięki wykorzystaniu metod i technik sterowania, zaczerpniętych m.in. z cybernetyki: technicznej, ekonomicznej oraz medycznej. Dane można przeglądać w ujęciu terytorialnym, demograficznym (rozkład wiekowy danej jednostki chorobowej dla mężczyzn, kobiet, kobiet i mężczyzn oraz średniej), a także czasowym (dane roczne, miesięczne, tygodniowe, dzienne).

Źródło: opracowanie własne na podstawie: [2], [3], [4], [5], [6], [7].



Rys. 1. Mapy zdrowotne kraju dla ostrego zapalenia gardła (dane dzienne)

Fig. 1. Health maps for pharyngitis (daily data)

Źródło: <http://www.osoz.pl/osoz-mzk/servlet/main/indexhtml>.

Kwestią dyskusyjną pozostaje wiarygodność zarówno jednych, jak i drugich źródeł informacji.

3. Procedura pozyskiwania i analizy danych

W procesach pozyskania i analizy zebranych danych wykorzystano dwa narzędzia informatyczne – system LimeSurvey oraz program R. Przygotowanie zbioru danych ankietowych obejmowało kilka etapów, które zaprezentowano na rys. 2.

LimeSurvey to system służący do zarządzania badaniami ankietowymi. Jego możliwości wykraczają daleko poza samo opracowanie pojedynczego kwestionariusza. Na potrzeby tego badania został wykorzystany do: opracowania kwestionariusza, śledzenia postępu wypełniania go przez respondentów, dzięki graficznej, podstawowej analizie danych, automatycznego zamknięcia badania i eksportu wyników. Ponadto, LimeSurvey pozwala m.in. na: tworzenie kwestionariuszy wielojęzycznych, osadzanie multimediiów w ankiecie, zarządzanie użytkownikami [8]. LimeSurvey jest popularnym oprogramowaniem z uwagi na

liczne zalety. Przede wszystkim należy wskazać, iż jest to oprogramowanie wolne i otwarte (licencja GNU GPL), przez co z powodzeniem może być wykorzystywane w celach komercyjnych, bezpłatnie [9]. System ten oparty jest na bazie danych MySQL i serwerze PHP [10].



Rys. 2. Etapy pozyskania i analizy danych
Fig. 2. Stages of collecting and analyzing data
Źródło: opracowanie własne.

Jak już wspomniano, do analizy danych wykorzystano program R. Należy zwrócić uwagę, że pod nazwą R kryje się zarówno nazwa języka programowania, jak i nazwa platformy programistycznej, która wyposażona jest w interpretator tego języka. W szerokim ujęciu R to także nazwa projektu, w ramach którego rozwijane są język i środowisko [11]. Należy zaznaczyć, iż program ten pozwala na opracowania własnego skryptu, samodzielne napisanie funkcji, co zostało wykorzystane w przeprowadzonych analizach.

4. Wyniki badań z lat 2013–2015

Z uwagi na rosnącą popularność badań ankietowych online [12], w celu uzyskania wyników każdorazowo posłużono się metodą CAWI (ang. *Computer Assisted Web Interview*), czyli wykorzystano Internet jako medium do zbierania danych [13]. Dodatkowymi zaletami badań ankietowych online są: duży zasięg badania oraz stosunkowo niski koszt i szybki dostęp do wyników [14].

Pierwsze badanie dotyczące tej tematyki przeprowadzono w marcu 2013 roku [15]. W badaniu, które miało charakter pilotażowy, wzięło udział 212 respondentów, z tego 69,81% procent stanowiły kobiety. Ankietowani odpowiadali na pytanie dotyczące śledzenia za pomocą Internetu informacji epidemiologicznych. W badaniu tym nieco ponad 10% respondentów odpowiedziało, że poszukuje takich wiadomości [16].

W 2014 roku powtórzono badanie zwiększając liczbę respondentów do 526 osób, głównie mieszkańców województwa śląskiego. Badanie trwało miesiąc, w okresie od 10 kwietnia do 10 maja. Zdecydowaną większość próby (72,62%) stanowiły kobiety, przy 27,38% udziale mężczyzn [17]. W 2014 roku uzyskano 12,74% odpowiedzi twierdzących na pytanie dotyczące śledzenia epidemii za pomocą sieci Internet.

Rok później, w 2015 roku, przeprowadzono trzeci etap badania rozszerzając próbkę do 1023 respondentów z całej Polski [18]. W badaniu tym wzięły udział 762 kobiety, co stanowiło 74,49% próby oraz 261 mężczyzn, stanowiących odpowiednio 25,51% ogółu respondentów. W 2015 roku uzyskano znaczny wzrost zainteresowania wśród respondentów tematyką epidemii. Wśród ankietowanych 21,51% potwierdziło, że śledzi informacje o epidemiach z wykorzystaniem Internetu. Podział respondentów ze względu na płeć przedstawiono w tabeli 2, natomiast wyniki badania dotyczącego monitorowania epidemiologii w tabeli 3.

Tabela 2

Podział respondentów ze względu na płeć w latach 2013-2015

Rok \ Płeć	2013	2014	2015
Kobieta	69,81%	72,62%	74,49%
Mężczyzna	30,19%	27,38%	25,51%

Źródło: opracowanie własne na podstawie: Furmankiewicz M., Sołtysik-Piorunkiewicz A., Ziuziański P.: Wykorzystanie technologii ICT w społeczeństwie informacyjnym w świetle badań systemów zarządzania wiedzą w e-zdrowiu, [w:] Białas A. (red.): Informatyka w Województwie Śląskim – innowacyjne trendy rozwoju. Rozprawy i Monografie, Instytut Technik Innowacyjnych EMAG, Katowice 2015.

Tabela 3

Zainteresowanie respondentów informacjami dotyczącymi epidemii w latach 2013-2015

Rok \ Płeć	2013	2014	2015
Razem, w tym:	10,38%	12,74%	21,51%
Kobieta	7,55%	8,56%	15,05%
Mężczyzna	2,83%	4,18%	6,45%

Źródło: opracowanie własne.

Z przedstawionych danych wynika, że z roku na rok wśród internautów, zarówno kobiet, jak i mężczyzn, rośnie zainteresowanie tematyką epidemii.

5. Zastosowanie metody reprezentacyjnej bazując na narzędziu R

W metodzie reprezentacyjnej wyróżnić możemy dwa główne podejścia – randomizacyjne oraz modelowe. Pierwsze z nich – randomizacyjne – zakłada, że źródłem losowości ocen analizowanych charakterystyk populacji lub domen jest plan losowania, czyli sposób

losowania elementów do próby. Ponadto, przyjmuje się, że wartości badanej cechy są z góry ustalone, a zatem nie są losowe. Podejście to rozważane jest m.in. w pracach: Brachy [20], Pawłowskiego [21], Steczkowskiego [22]. W podejściu modelowym natomiast wektor wartości cechy badanej w populacji pojmowany jest jako realizacja wektora zmiennych losowych, o łącznym rozkładzie oznaczanym jako ξ . Przyjmowane założenia, a dotyczące tego rozkładu określają model nadpopulacji. Podejście to szerzej opisane zostało m.in. w pracach Cassela, Särndala, Wretmana [23], Valianta, Dorfmana, Royalla [24]. Należy zaznaczyć, iż podejście randomizacyjne ma zastosowanie jedynie w przypadku, gdy próba jest losowa, natomiast modelowe może być stosowane zarówno dla prób losowych, jak i prób nielosowych często wykorzystywanych w praktyce (por. Żądło [25], s. 18-19, 27).

Metoda reprezentacyjna znajduje zastosowanie w wielu dziedzinach takich jak analizy rynkowe, badania demograficzne, badania przedsiębiorstw. Może ona zatem zostać wykorzystana także w przypadku analizy danych pochodzących z badania ankietowego. Problem ten poruszony został m.in. w [26], Domański, Krzciuk, Miłek, Ziuziański. W niniejszym artykule, z uwagi na to, iż dane dostępne w wyniku przeprowadzonego badania pochodzą z próby nielosowej, wymagane jest zastosowanie podejścia modelowego. Rozważanym problemem jest predykcja na podstawie modelu warstwowego. Zakładamy zatem podział badanej populacji U , na H rozłącznych warstw U_h o liczebnościach odpowiednio N_h ($h=1, \dots, H$, $N = \sum_{h=1}^H N_h$). Ponadto, przez s_h oznaczamy informacje z próby i h -tej warstwy, a zatem próbki z warstwy ($s_h = s \cap U_h$), o liczebności n_h .

Analizowany model nadpopulacji określony jest przez poniższe założenia:

$$\begin{cases} E_{\xi}(Y_{ih}) = \mu_h, D_{\xi}^2(Y_{ih}) = \sigma_h^2 \\ \text{Cov}(Y_i, Y_t) = 0 \end{cases}, \quad (1)$$

gdzie: $i \in U_h$, $h=1, \dots, H$, $i=1, \dots, N$, $t=1, \dots, N$, $i \neq t$ (por. Valiant, Dorfman, Royall [25], s. 168, Wywiół [27], s.113).

Do oszacowania frakcji w populacji zastosowano predyktor (por. Wywiół [27], s. 56):

$$\hat{Y} = N^{-1} \sum_{h=1}^H N_h \bar{Y}_{sh}, \quad (2)$$

gdzie

$$\bar{Y}_{sh} = \hat{Y}_h = n_h^{-1} \sum_{i=1}^{n_h} Y_i. \quad (3)$$

Należy zauważyć, że na potrzeby oceny frakcji badana zmienna powinna być zmienną zero-jedynkową, co rozważane było m.in. w [28], Domański, Krzciuk, Miłek, Ziuziański. Ponadto, powyższy wzór można także zastosować do predykcji średniej w populacji. Powinno zostać podkreślone, iż ξ - błąd średniokwadratowy predyktora danego wzorem (2),

uwzględniając jego nieobciążoność, równy jest wariancji predykcji. Może on zatem być szacowany za pomocą nieobciążonej statystyki:

$$M\hat{S}E_{\xi}(\hat{Y}) = \hat{D}_{\xi}^2(\hat{Y}) = N^{-2} \sum_{h=1}^H n_h^{-1} N_h (N_h - n_h) \hat{\sigma}_h^2, \quad (4)$$

gdzie

$$\hat{\sigma}_h^2 = (n_h - 1)^{-1} \sum_{i=1}^{n_h} \left(Y_i - n_h^{-1} \sum_{i=1}^{n_h} Y \right)^2. \quad (5)$$

Do oszacowania frakcji w warstwie wykorzystać można predyktor dany wzorem (3). W tym przypadku również, ze względu na nieobciążoność predyktora ξ – błąd średniokwadratowy równy jest wariancji predykcji. Jego ocena wyznaczona może być na podstawie poniższej statystyki:

$$M\hat{S}E(\hat{Y}_h) = \hat{D}_{\xi}^2(\hat{Y}_h) = \frac{N_h - n_h}{N_h n_h} \hat{\sigma}_h^2, \quad (6)$$

gdzie $\hat{\sigma}_h^2$ dane jest wzorem (5) (por. Wywiół [27], s. 56).

Dodatkowo, dla dogłębniejszej oceny jakości predykcji, możliwe jest wyznaczenie pierwiastka statystyk oznaczonych jako (4) oraz (6) (RMSE), równego w tym przypadku ocenie średniego błędu predykcji, oraz względnej jego wartości (RRMSE) – w stosunku do wartości predyktora.

6. Wyniki badań 2015

W niniejszej części artykułu przedstawione zostaną rezultaty analiz przeprowadzonych na podstawie danych z próby, zebranych w wyniku badania ankietowego w 2015 roku.

Dane dotyczące śledzenia informacji epidemiologicznych w Internecie, uzyskane z próby, sprowadzone zostały do postaci zmiennej zero-jedynkowej, gdzie wartość 1 przypisana została odpowiedzi twierdzącej.

W badaniu uwzględnione zostały trzy zmienne warstwujące: płeć, wiek oraz zamieszkiwane województwo. Zatem w pierwszym przypadku rozważano podział na dwie warstwy, w trzecim natomiast z populacji wyodrębniono szesnaście warstw. Gdy populację podzielono ze względu na wiek, wyodrębnionych zostało pięć warstw. Uwzględnione przedziały wiekowe to: 15-24, 25-34, 35-44, 45-54 oraz 55 i więcej lat. Dane dotyczące liczebności poszczególnych warstw w populacji, użyte w analizie, pochodzą z Banku Danych Lokalnych Głównego Urzędu Statystycznego.

W tabeli 4 zaprezentowane zostały wyniki przeprowadzonych analiz dla frakcji osób śledzących za pomocą Internetu informacje epidemiologiczne w populacji, przy uwzględnieniu wszystkich rozważanych czynników grupujących.

Tabela 4

Tabela wynikowa dla oszacowań frakcji w populacji

Podział populacji na warstwy ze względu na:	Wartość predyktora frakcji w populacji	MSE	RMSE	RRMSE
Płeć	0,2264	0,0002	0,015	0,0661
Wiek	0,3586	0,0014	0,038	0,1065
Miejsce zamieszkania	0,2223	0,0002	0,014	0,0642

Źródło: opracowanie własne.

W przypadku podziału na warstwy, ze względu na płeć oraz miejsce zamieszkania, uzyskano zbliżoną wartość predyktora. Szacowana wartość odsetku osób śledzących informacje epidemiologiczne wyniosła ok. 22%. Gdy warstwy wydzielone zostały na podstawie wieku, to szacowany odsetek był znacznie wyższy – przekroczył on 35%. Pierwiastek błędu średniokwadratowego analizowanego predyktora (RMSE) dla pierwszej i trzeciej z rozważanych zmiennych warstwujących również przyjął zbliżoną wartość. Wartości frakcji osób śledzących informacje epidemiologiczne odchylają się od badanej frakcji w populacji o przeciętnie 1,5 punktu procentowego. W obu przypadkach stanowi to ok. 6,5% wartości predyktora. Jeśli chodzi o wiek, to względny pierwiastek błędu średniokwadratowego (RRMSE) stanowi nieco ponad 10% wartości predyktora.

Druga część przeprowadzanych analiz dotyczyła oszacowania badanej frakcji w poszczególnych warstwach, wyodrębnionych na podstawie omówionych powyżej trzech czynników grupujących. Tabela 5 zawiera zestawienie uzyskanych wyników, gdy za zmienną na podstawie, której określono warstwy, przyjęto płeć.

Tabela 5

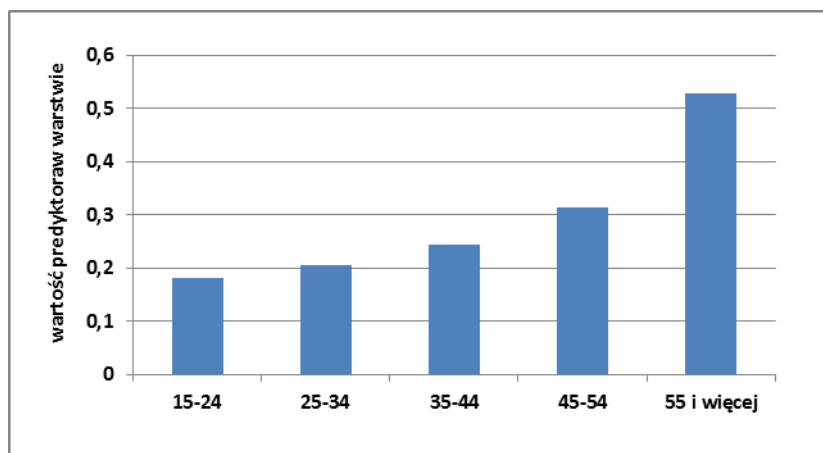
Tabela wynikowa dla oszacowań frakcji w warstwach – podział ze względu na płeć

Podział populacji na warstwy ze względu na płeć	Wartość predyktora frakcji w populacji	MSE	RMSE	RRMSE
Kobieta	0,202	0,0002	0,015	0,072
Mężczyzna	0,253	0,0007	0,027	0,107

Źródło: opracowanie własne.

Na podstawie uzyskanych wyników ocenia się, iż co piąta kobieta i co czwarty mężczyzna śledzą w Internecie informacje dotyczące epidemii. Wartości tych predyktorów odchylają się od frakcji w warstwie dla populacji o przeciętnie 1,5 punktu procentowego w przypadku kobiet, i 2,7 w przypadku mężczyzn. Odpowiednio pierwiastek błędu średniokwadratowego w rozważanych warstwach stanowił nieco ponad 7 i ok. 11% wyznaczonego predyktora.

Rysunek 3 przedstawia wyniki predykcji analizowanej frakcji w warstwach, gdy populacja podzielona została ze względu na wiek. Szacuje się, że największy odsetek osób śledzących informacje epidemiologiczne odnotowuje się wśród osób w wieku powyżej 54 lat – ponad 50%. Ocenia się natomiast, iż odsetek ten jest najmniejszy w grupie osób od 15 do 24 lat. W tej warstwie wartość predyktora wyniosła mniej niż 20%.



Rys. 3. Wartości predyktora frakcji dla czynnika grupującego wiek

Fig. 3. Values of fractions predictor for the factor age grouping

Źródło: opracowanie własne.

Poniższa tabela zawiera zbiorcze zestawienie wartości błędów średniokwadratowych, wraz z ich pierwiastkami i wartościami względnymi w odniesieniu do oceny frakcji w warstwie.

Tabela 6

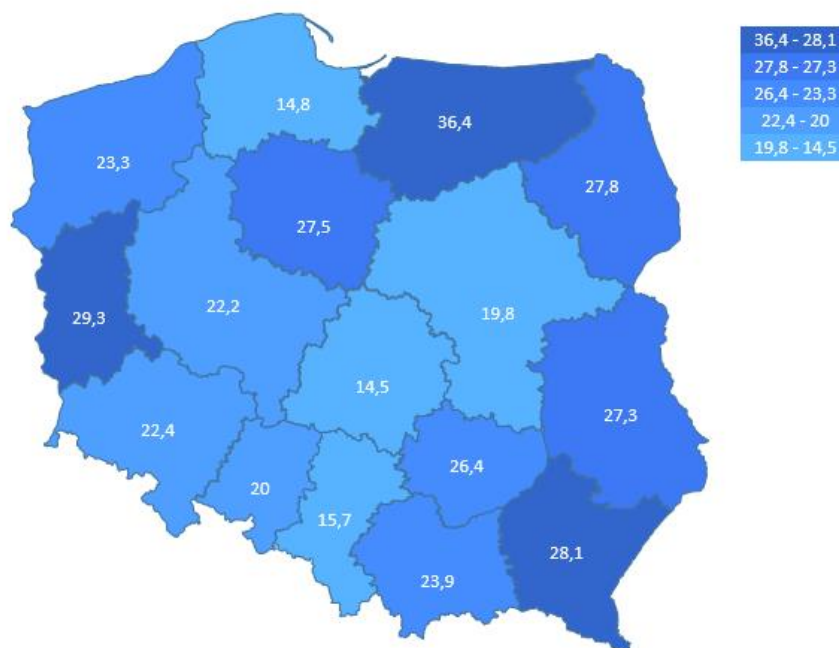
Tabela wynikowa dla oszacowań frakcji w warstwach – podział ze względu na wiek

Przedział wiekowy (w latach)	MSE	RMSE	RRMSE
15-24	0,0003	0,018	0,098
25-34	0,0005	0,022	0,107
35-44	0,0015	0,038	0,157
45-54	0,0043	0,066	0,202
55 i więcej	0,0075	0,087	0,167

Źródło: opracowanie własne.

Najmniejszą wartość błędu średniokwadratowego zaobserwowano w przypadku przedziału wiekowego 15-24 lata. Wartości frakcji osób śledzących informacje epidemiologiczne odchylają się od badanej frakcji w tej warstwie, w populacji o przeciętnie 1,8 punktu procentowego, co stanowi niespełna 10% wartości predyktora. Również w przypadku kolejnej z analizowanych warstw wartość RRMSE nie przekroczyła 12%. Największy względny pierwiastek MSE odnotowano dla grupy wiekowej 45-54 lata, wyniósł on ok. 20%.

Ostatnim z rozważanych przypadków jest predykcja badanej frakcji w warstwach, gdy określone zostały one na podstawie miejsca zamieszkania. Rysunek 4 prezentuje, uzyskane na podstawie próby, wartości predyktorów.



Rys. 4. Wartości predyktora frakcji w warstwie dla zmiennej warstwującej miejsce zamieszkania
Fig. 4. Values of fraction predictor for variable layering place of residence

Źródło: opracowanie własne.

Najwyższą wartość predyktora frakcji w warstwie uzyskano dla województwa warmińsko-mazurskiego. Szacuje się, że ok. 36% mieszkańców tego województwa śledzi informacje w Internecie o epidemiach. W przypadku czterech województw ocena badanej frakcji była niższa niż 20%. Były to województwa: łódzkie, mazowieckie, pomorskie oraz śląskie. Jednakże w przypadku województwa łódzkiego odsetek ten był najniższy – ocenia się, że jedynie nieco ponad 14% mieszkańców śledzi informacje dotyczące epidemii.

W tabeli 7 zaprezentowano wartości błędów średniokwadratowych i względnych pierwiastków tychże błędów dla poszczególnych województw.

Tabela 7

Wartości błędów predyktora frakcji w warstwie dla zmiennej warstwującej miejsce zamieszkania

Województwo	MSE	RMSE	RRMSE
dolnośląskie	0,003	0,055	0,246
kujawsko-pomorskie	0,005	0,071	0,260
lubelskie	0,006	0,079	0,289
lubuskie	0,005	0,072	0,246
łódzkie	0,002	0,045	0,311
małopolskie	0,003	0,051	0,213
mazowieckie	0,002	0,045	0,225
opolskie	0,004	0,060	0,301

cd. tabeli 7

podkarpackie	0,007	0,081	0,287
podlaskie	0,006	0,076	0,273
pomorskie	0,002	0,049	0,329
śląskie	0,001	0,023	0,147
świętokrzyskie	0,004	0,061	0,231
warmińsko-mazurskie	0,005	0,073	0,220
wielkopolskie	0,002	0,046	0,209
zachodniopomorskie	0,004	0,65	0,280

Źródło: opracowanie własne.

Na podstawie powyższej tabeli (7) można zaobserwować, że wartości błędów średniokwadratowych wyznaczonych predyktorów frakcji w warstwie nie przekroczyły poziomu 0,008. Najniższy błąd otrzymano dla województwa śląskiego (0,001). Wartości RRMSE przyjmowały wartości od ok. 15% – dla województwa śląskiego, do niespełna 33% wartości predyktora w przypadku województwa pomorskiego.

7. Podsumowanie

Jak wynika to badań, monitorowanie sytuacji epidemiologicznej dla Internautów nie jest zjawiskiem obcym. Niniejszy artykuł może stanowić punkt wyjścia do pogłębienia badań w zakresie monitorowania zjawisk epidemiologicznych wśród Internautów. Badania należałoby rozszerzyć, wzbogacając je np. o wskazanie źródła informacji epidemiologicznej.

Zastosowanie podejścia modelowego metody reprezentacyjnej pozwoliło na oszacowanie badanego odsetku osób śledzących za pomocą Internetu informacje dotyczące epidemii (w populacji oraz wyodrębnionych warstwach), na podstawie danych uzyskanych z przeprowadzonego badania ankietowego. Dokonano również oceny jakości predykcji. Poprawa jakości oszacowań, a zatem niższe wartości błędów średniokwadratowych oraz ich względnych pierwiastków możliwe byłyby do uzyskania przy zwiększonej liczebności próby w warstwach.

Bibliografia

1. Jędrzychowski W.: Epidemiologia – wprowadzenie i metody badań. Wydawnictwo Lekarskie PZWL, Warszawa 1999, s. 11.
2. Sołtysik-Piorunkiewicz A., Furmankiewicz M., Ziuziański P.: Kokpit menedżerski jako narzędzi do wspomaganie decyzji prosumenta w e-zdrowiu, [w:] Pańkowska M. (red.): Uwarunkowania technologiczno-społeczne i modele prosumpcji, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice 2015.

3. Furmankiewicz M., Sołtysik-Piorunkiewicz A., Ziuziański P.: Zaawansowane techniki graficznej analizy danych epidemiologicznych na kokpicie menadżerskim, [w:] Nycz M. (red.): Informatyka Ekonomiczna (Business Informatics), nr 2 (32), Uniwersytet Ekonomiczny we Wrocławiu, Wrocław 2014.
4. Furmankiewicz M., Ziuziański P.: Internet jako źródło danych epidemiologicznych, [w:] Zieliński Z.E. (red.): Rola informatyki w naukach ekonomicznych i społecznych. Innowacje i implikacje interdyscyplinarne, 2/2013, Wydawnictwo Wyższej Szkoły Handlowej w Kielcach, Kielce 2013.
5. Ziuziański P., Furmankiewicz M., Sołtysik-Piorunkiewicz A.: E-Health Artificial Intelligence System Implementation: Case Study of Knowledge Management Dashboard of Epidemiological Data in Poland. "International Journal of Biology and Biomedical Engineering", vol. 8, 2014, pp. 164-171.
6. Bukowska-Piestrzyńska A.: Pacjent jako e-klient, [w:] Zieliński Z.E. (red.): Rola informatyki w naukach ekonomicznych i społecznych. Innowacje i implikacje interdyscyplinarne. Tom 1/2009. Wydawnictwo Wyższej Szkoły Handlowej w Kielcach, Kielce 2012, s. 139-140.
7. Kaczmarek B.: Analiza strategii biznesowych spółki Google Inc. – studium przypadku, [w:] Antonowicz P. (red.): Zarządzanie i finanse nr 1/4, Sopot 2013, s. 239-240.
8. Furmankiewicz M., Ziuziański P.: Internetowy kwestionariusz ankiety jako nowoczesne narzędzie prowadzenia badań w naukach społecznych, [w:] Gołuchowski J., Spyra Z. (red.): Nowe Media i technologie we współczesnej komunikacji marketingowej. Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice 2015 (w druku).
9. Krzciuk M., Ziuziański P.: O teście niezależności trzech zmiennych na pewnym przykładzie empirycznym, [w:] Zieliński Z.E. (red.): Rola informatyki w naukach ekonomicznych i społecznych. Innowacje i implikacje interdyscyplinarne. Tom 2/2012, Wydawnictwo Wyższej Szkoły Handlowej w Kielcach, Kielce 2012.
10. Ziuziański P., Furmankiewicz M.: Rola Internetu w autodiagnozie i samoleczeniu: szanse i zagrożenia, [w:] Zieliński Z.E. (red.): Rola informatyki w naukach ekonomicznych i społecznych. Innowacje i implikacje interdyscyplinarne. Zeszyt 2/2014, Wydawnictwo Wyższej Szkoły Handlowej w Kielcach, Kielce 2014.
11. Biecek P.: Przewodnik po pakiecie R, Wrocław 2008, s. 2.
12. Zdonek D., Zdonek I., Spalek S.: Czynniki motywujące ekspertów do udziału w badaniach ankietowych, [w:] Zeszyty Naukowe Politechniki Śląskiej, s. Organizacja i Zarządzania, nr 67, Wydawnictwo Politechniki Śląskiej, Gliwice 2013, s.145-154.
13. Furmankiewicz M., Ziuziański P.: Popularność Europejskiego Certyfikatu Umiejętności Komputerowych – studium empiryczne, [w:] Zieliński Z.E. (red.): Rola informatyki w naukach ekonomicznych i społecznych. Innowacje i implikacje interdyscyplinarne, 1/2013, Wydawnictwo Wyższej Szkoły Handlowej w Kielcach, Kielce 2013.

14. Naramski M., Szmerek A.R., Herman K.: Analiza porównawcza wybranych stron internetowych służących do przeprowadzania badań ankietowych, [w:] Zeszyty Naukowe Politechniki Śląskiej, s. Organizacja i Zarządzania, nr 68, Wydawnictwo Politechniki Śląskiej, Gliwice 2014, s. 355-365.
15. Furmankiewicz M., Sołtysik-Piorunkiewicz A., Ziuziański P.: Czynniki determinujące zachowanie internautów w obszarze e-zdrowia na podstawie wyników badań z lat 2013-2015, [w:] Studia Ekonomiczne, Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach, Katowice 2015 (w druku).
16. Furmankiewicz M., Ziuziański P.: Ocena wykorzystania technologii teleinformatycznych w procesie autodiagnozy i samoleczenia w świetle badania opinii internautów, [w:] Sroka H., Porębska-Miąc T. (red.): Systemy Wspomagania Organizacji SWO 2013, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice 2013.
17. Ziuziański P., Furmankiewicz M.: Rola Internetu w autodiagnozie i samoleczeniu: szanse i zagrożenia, [w:] Zieliński Z.E. (red.): Rola informatyki w naukach ekonomicznych i społecznych. Innowacje i implikacje interdyscyplinarne. Zeszyt 2/2014, Wydawnictwo Wyższej Szkoły Handlowej w Kielcach, Kielce 2014.
18. Furmankiewicz M., Sołtysik-Piorunkiewicz A., Ziuziański P.: Badanie: Internet i zdrowie. „Ogólnopolski System Ochrony Zdrowia”, nr 9/2015, s. 42-45.
19. Furmankiewicz M., Sołtysik-Piorunkiewicz A., Ziuziański P.: Wykorzystanie technologii ICT w społeczeństwie informacyjnym w świetle badań systemów zarządzania wiedzą w e-zdrowiu, [w:] Białas A. (red.): Informatyka w Województwie Śląskim – innowacyjne trendy rozwoju. Rozprawy i Monografie, Instytut Technik Innowacyjnych EMAG, Katowice 2015.
20. Bracha C.: Teoretyczne podstawy metody reprezentacyjnej. PWN, Warszawa 1996.
21. Pawłowski Z.: Wstęp do statystycznej metody reprezentacyjnej. PWN, Warszawa 1972.
22. Steczkowski J.: Metoda reprezentacyjna w badaniach zjawisk ekonomiczno-społecznych. PWN, Warszawa 1995.
23. Cassel C.M., Särndal C.E., Wretman J.H.: Foundations of Inference in Survey Sampling. John Willey & Sons, New York-London-Sydney-Toronto 1977.
24. Valliant R., Dorfman A.H., Royall R.M.: Finite Population Sampling and Inference. A Prediction Approach. John Willey & Sons, New York 2000.
25. Żądło T.: Elementy statystyki małych obszarów z programem R. Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice 2008.
26. Domański P., Krzciuk M., Miłek M., Ziuziański P.: Badania ankietowe w oparciu o próby nielosowe z wykorzystaniem programu SPSS, [w:] Zieliński Z.E. (red.): Rola informatyki w naukach ekonomicznych i społecznych. Innowacje i implikacje interdyscyplinarne. Zeszyt 2/2011, Wydawnictwo Wyższej Szkoły Handlowej w Kielcach, Kielce 2011.
27. Wywiół J.: Wprowadzenie do metody reprezentacyjnej. Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice 2010.

28. Domański P., Krzciuk M., Miłek M., Ziuziański P.: Badania ankietowe w praktyce, [w:] Kuczera M. (red.): Rola dokonań studentów a możliwości osiągnięcia sukcesu po zakończeniu studiów, Wydawnictwo CreativeTime, Kraków 2011, s. 142.

Abstract

Nowadays, Internet as a medium became a very important source of knowledge for information society. Government institutions and private corporation on their web pages give access to information about epidemiological situation in country to prevent pandemic in order to protect society. Authors present some examples of this information sources. Article discuss topic of monitoring epidemic by Internet users and presents results of conducted survey research in this area in 2015. The process of collecting data using Computer Assisted Web Interview method has been described. Obtained data has been analyzed with usage of sampling survey method which has been described in detail.