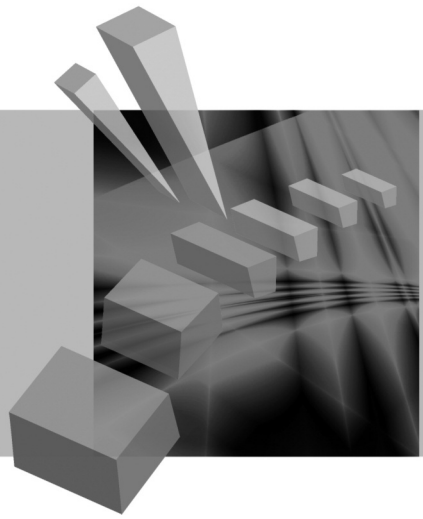
PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu
RESEARCH PAPERS
of Wrocław University of Economics

242

Taksonomia 19.

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2012

Recenzenci: Eugeniusz Gatnar, Elżbieta Gołata, Tadeusz Kufel, Józef Pociecha,
Mirosław Szreder, Feliks Wysocki

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Tytuł sfinansowano ze środków Sekcji Klasyfikacji i Analizy Danych PTS
i Uniwersytetu Ekonomicznego we Wrocławiu

Publikacja jest dostępna na stronie www.ibuk.pl

Streszczenia opublikowanych artykułów są dostępne w międzynarodowej bazie danych
The Central European Journal of Social Sciences and Humanities http://cejsh.icm.edu.pl
oraz w The Central and Eastern European Online Library www.ceeol.com,
a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon http://kangur.uek.krakow.pl/
bazy_ae/bazekon/nowy/index.php

Informacje o naborze artykułów i zasadach recenzowania znajdują się
na stronie internetowej Wydawnictwa
www.wydawnictwo.ue.wroc.pl

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM
Nakład: 320 egz.

Spis treści

Summaries

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU nr 242
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS 2012

Taksonomia 19. Klasyfikacja i analiza danych – teoria i zastosowania ISSN 1899-3192

Joanna Banaś, Małgorzata Machowska-Szewczyk
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

OCENA INTENSYWNOŚCI WYKORZYSTANIA SKRZYNEK POCZTY ELEKTRONICZNEJ ZA POMOCĄ UPORZĄDKOWANEGO MODELU PROBITOWEGO

Streszczenie: Celem artykułu jest analiza opinii studiujących użytkowników kont pocztowych na temat posiadanych skrzynek poczty. Źródło danych stanowiły kwestionariusze ankiety wypełniane przez studentów wybranych wydziałów szczecińskich uczelni, które zawierały pytania dotyczące zarówno charakterystyki skrzynki poczty elektronicznej, jak i intensywności jej używania. Do oceny obsługi poczty elektronicznej wykorzystano uporządkowany model probitowy, w którym występuje jakościowa zmienna zależna, opisująca kategorie uporządkowane przy założeniu rozkładu normalnego składnika losowego.

Słowa kluczowe: uporządkowany model probitowy, badanie ankietowe, metoda regresji krokowej.

1. Wstęp

Poczta elektroniczna jest obecnie jedną z najczęściej używanych form komunikacji między ludźmi i instytucjami. Większość portali społecznościowych oferuje bezpłatne założenie konta i dostęp do poczty elektronicznej przez stronę internetową WWW. Wiele szkół, wyższych uczelni czy firm również umożliwia upoważnionej grupie zakładanie własnych skrzynek poczty elektronicznej. Podstawowym celem artykułu jest analiza opinii studiujących użytkowników kont pocztowych na temat posiadanych skrzynek poczty. Badanie zostało przeprowadzone na podstawie kwestionariuszy ankiety wypełnianych przez studentów wybranych wydziałów szczecińskich uczelni. Kwestionariusz zawierał pytania dotyczące zarówno charakterystyki skrzynki poczty elektronicznej, jak i intensywności jej używania. Każdy respondent mógł także ocenić sposób prezentacji danych w serwisie internetowym i częstość występowania problemów z wykorzystaniem niektórych usług.

Ponieważ większość odpowiedzi na pytania występujące w kwestionariuszu ankiety dostarcza informacji o charakterze jakościowym, wyrażonej w skali nominalnej lub porządkowej, to do oceny obsługi poczty elektronicznej wykorzystano uporządkowany model probitowy, w którym występuje jakościowa zmienna zależna, opisu-

jąca kategorie uporządkowane, przy założeniu normalnego rozkładu składnika losowego. Model taki umożliwia zbadanie wpływu poszczególnych zmiennych na poziom oceny posiadanej skrzynki poczty elektronicznej.

2. Opis metody

Uporządkowany model probitowy służy do opisu jakościowej zmiennej zależnej, której warianty można uporządkować. Przykładem takiej zmiennej jest subiektywna ocena sytuacji finansowej rodziny z wartościami: 1 – zła, 2 – średnia, 3 – dobra, 4 – bardzo dobra. Mogą to być również zmienne ilościowe, których wartości nie są dokładnie znane, a dostępna jest informacja, z jakiego przedziału pochodzą. Model ten po raz pierwszy został przedstawiony w pracy Zavoina i McKelveya [1975]. Aby dokonać formalnego zapisu modelu probitowego, zakłada się istnienie nieobserwowalnej zmiennej Y^* [Kostrzewska 2010; Dudek 2007]. Na podstawie jej wartości otrzymanych dla *i*-tej obserwacji *n*-elementowej próbki przypisuje się kategorie zmiennej zależnej *Y* następująco:

$$y_i = \begin{cases} 0, & \text{gdy } y_i^* \le \mu_1 \\ 1, & \text{gdy } \mu_1 < y_i^* \le \mu_2 \\ 2, & \text{gdy } \mu_2 < y_i^* \le \mu_3 \\ \vdots & \\ J, & \text{gdy } y_i^* > \mu_J \end{cases}, \tag{1}$$

gdzie $\mu_1, \mu_2, ..., \mu_J$ to wartości progowe, umożliwiające nadawanie kodów zmiennej *Y*, Y^* to nieobserwowalna zmienna, o której zakłada się, że $Y^* = \beta_0 + \mathbf{X}\boldsymbol{\beta} + U$, gdzie $\mathbf{X} = [X_1, ..., X_K]$ jest wektorem losowym obserwowanych zmiennych, $\boldsymbol{\beta} = [\beta_1, ..., \beta_K]^T$ – wektorem parametrów modelu, β_0 to wyraz wolny, U – składnik losowy o rozkładzie normalnym standaryzowanym.

Do modelowania prawdopodobieństwa zajścia zdarzenia, że zmienna *Y* przyjmie ustaloną wartość $y_i \in \{0,1,2, ..., J\}$ pod warunkiem, że zmienne $X_1 = x_{i1}, ..., X_K = x_{iK}$ można wykorzystać założenie, że składnik losowy U ma rozkład normalny standaryzowany i wyznaczyć je za pomocą dystrybuanty rozkładu normalnego standaryzowanego Φ:

$$\begin{aligned} P(Y = y_i / X_1 = x_{i1}, ..., X_K = x_{iK}) &= P(y_i = j / X_1 = x_{i1}, ..., X_K = x_{iK}) = \\ &= P(\mu_j < y_i^* \le \mu_{j+1}) = P(\mu_j < \mathbf{x}_i\mathbf{b} + b_0 + u_i \le \mu_{j+1}) = \\ &= P(\mu_j - \mathbf{x}_i\mathbf{b} - b_0 < u_i \le \mu_{j+1} - \mathbf{x}_i\mathbf{b} - b_0) = \\ &= \Phi(\mu_{j+1} - \mathbf{x}_i\mathbf{b} - b_0) - \Phi(\mu_j - \mathbf{x}_i\mathbf{b} - b_0). \end{aligned} \tag{2}$$

Zatem funkcja prawdopodobieństwa przyjęcia wariantu *i* przez zmienną *Y* dla konkretnej obserwacji jest następująca:

$$P(Y=y_i \,/\, \mathbf{X}=\mathbf{x}_i)=\begin{cases}\Phi(\mu_1-\mathbf{x}_i\mathbf{b}-b_0) & \text{, gdy } \ y_i=0\\ \Phi(\mu_2-\mathbf{x}_i\mathbf{b}-b_0)-\Phi(\mu_1-\mathbf{x}_i\mathbf{b}-b_0) & \text{, gdy } \ y_i=1\\ \Phi(\mu_3-\mathbf{x}_i\mathbf{b}-b_0)-\Phi(\mu_2-\mathbf{x}_i\mathbf{b}-b_0) & \text{, gdy } \ y_i=2\\ \vdots & \\ \Phi(\mu_J-\mathbf{x}_i\mathbf{b}-b_0)-\Phi(\mu_{J-1}-\mathbf{x}_i\mathbf{b}-b_0) & \text{, gdy } \ y_i=J-1\\ 1-\Phi(\mu_J-\mathbf{x}_i\mathbf{b}-b_0), & \text{, gdy } \ y_i=J\end{cases}, \quad (3)$$

gdzie $\mathbf{x}_i=\left[x_{i1}, x_{i2}, ..., x_{iK}\right]$ – wektor wartości zmiennych $X_1, ..., X_K$ zaobserwowanych dla *i*-tej jednostki, $\mathbf{b}=[b_1, ..., b_K]^T$ oraz b_0 – wartości estymatorów parametrów $\boldsymbol{\beta}$ oraz β_0 odpowiednio uzyskane z próbki.

Po oszacowaniu za pomocą metody największej wiarygodności parametrów modelu $\mathbf{b}$ oraz b_0 (przy czym wartości estymatorów nie interpretuje się bezpośrednio) warto dokonać predykcji efektów zmian wartości zmiennych objaśniających na wartość prawdopodobieństwa przynależności do każdej z tych grup. Owe efekty krańcowe wpływu zmian wartości zmiennej X_k, $k\in\{1,2, ..., K\}$ na prawdopodobieństwo, że zmienna zależna *Y* przyjmie dla *i*-tej obserwacji ustaloną wartość porządkową $y_i\in\{0,1,2, ..., J\}$, wyznacza się, wykorzystując pochodną cząstkową [Kostrzewska 2010]:

$$\frac{\partial P(Y=y_i \,/\, \mathbf{X}=\mathbf{x}_i)}{\partial x_{ik}}=\begin{cases}b_k\ \varphi(\mu_1-\mathbf{x}_i\mathbf{b}-b_0) & \text{, gdy } \ y_i=0\\ b_k\left(\varphi(\mu_2-\mathbf{x}_i\mathbf{b}-b_0)-\varphi(\mu_1-\mathbf{x}_i\mathbf{b}-b_0)\right) & \text{, gdy } \ y_i=1\\ b_k\left(\varphi(\mu_3-\mathbf{x}_i\mathbf{b}-b_0)-\varphi(\mu_2-\mathbf{x}_i\mathbf{b}-b_0)\right) & \text{, gdy } \ y_i=2\\ \vdots & \\ b_k\left(\varphi(\mu_J-\mathbf{x}_i\mathbf{b}-b_0)-\varphi(\mu_{J-1}-\mathbf{x}_i\mathbf{b}-b_0)\right) & \text{, gdy } \ y_i=J-1\\ -b_k\varphi(\mu_J-\mathbf{x}_i\mathbf{b}-b_0) & \text{, gdy } \ y_i=J\end{cases}, \quad (4)$$

gdzie φ oznacza funkcję gęstości rozkładu normalnego standaryzowanego.

Jeżeli zmienna niezależna X_k, $k\in\{1,2, ..., K\}$ jest dychotomiczna, to jej wpływ na zmianę prawdopodobieństwa, że zmienna objaśniana *Y* przyjmie dla *i*-tej obserwacji ustaloną wartość porządkową $y_i\in\{0,1,2, ..., M\}$, wyznaczany jest jako różnica odpowiednich prawdopodobieństw według wzoru (5):

$$\begin{aligned}&P(Y=y_i \,/\, X_1=x_{i1},...,X_k=1,...,X_K=x_{iK})-\\ &P(Y=y_i \,/\, X_1=x_{i1},...,X_k=0,...,X_K=x_{iK}).\end{aligned} \quad (5)$$

Wpływy cząstkowe oblicza się zazwyczaj w punkcie wyznaczonym jako wektor wartości średnich poszczególnych zmiennych niezależnych. Oszacowane parametry modelu wskazują na taki sam wpływ konkretnej zmiennej objaśniającej na prawdopodobieństwo każdego ze stanów osiąganych przez zmienną zależną. Wpływy cząstkowe umożliwiają ocenę oddziaływania każdej ze zmiennych indywidualnie na zmiany prawdopodobieństwa każdego ze stanów zmiennej objaśnianej.

3. Materiał badawczy

Aby przeprowadzić ocenę intensywności wykorzystania skrzynek poczty elektronicznej przez studentów, zgromadzono materiał badawczy techniką CAWI [Mazurek-Łopacińska 2005]. Na podstawie wypełnionych przez studentów kilku szczecińskich uczelni kwestionariuszy ankiety [Górski 2011] od marca do maja 2011 r. zbudowano uporządkowany model probitowy. Spośród dostępnych informacji wyodrębniono grupę 252 studentów, którzy zadeklarowali posiadanie adresu poczty elektronicznej. Zmienną zależną zdefiniowano na podstawie odpowiedzi na pytanie 4. w kwestionariuszu ankiety:

Ile wiadomości e-mail wysyłasz w ciągu dnia?

Zmienna zależna *Y* może dla *i*-tej obserwacji w próbce przyjmować wartości y_i określone następująco:

$$y_i = \begin{cases} 0, & \text{gdy } y_i^* \le 2 \quad \text{(nie więcej niż 2)} \\ 1, & \text{gdy } 2 < y_i^* \le 5 \quad \text{(od 3 do 5)} \\ 2, & \text{gdy } 5 < y_i^* \le 9 \quad \text{(od 6 do 9)} \\ 3, & \text{gdy } y_i^* > 9 \quad \text{(10 lub więcej)} \end{cases} .$$

Przy analizie ocen skrzynek poczty elektronicznej na podstawie kwestionariusza ankiety przyjęto 31 potencjalnych zmiennych objaśniających X_k. Na podstawie każdej zmiennej o charakterze jakościowym, której liczba możliwych kategorii była większa niż 2, utworzono odpowiednią liczbę nowych zmiennych binarnych, odzwierciedlających poszczególne kategorie. Wprowadzono oznaczenie $X_{k.j}$ dla zmiennych binarnych utworzonych na podstawie zmiennej X_k, gdzie $k \in \{1, ..., 31\}$ jest numerem zmiennej, *j* zaś oznacza numer wariantu wyjściowej zmiennej. Zmiennej $X_{k.j}$ przypisuje się wartość jeden, jeżeli wystąpił wybrany wariant (zmienna X_k przyjęła *j*-ty wariant cechy), natomiast w przeciwnym przypadku wartość zero. Tabela 1 przedstawia zbiór zmiennych potencjalnie objaśniających (zmienne $X_{18.3}, X_{26.1}, X_{31.3}, X_{31.4}$ usunięto, gdyż żaden z badanych nie wybrał takiej kategorii, co zaznaczono przez przekreślenie).

Tabela 1. Zmienne objaśniające (wytłuszczono zmienne bazowe).

Zmienna	Opis zmiennych binarnych
1	2
X_1	Czas posiadania adresu poczty elektronicznej. Zmienne binarne $\boldsymbol{X_{1.1}}, X_{1.2}, X_{1.3}$ odpowiadające kategoriom: **krócej niż 5 lat**, 5-9 lat, 10 lat lub więcej
X_2	Liczba adresów poczty elektronicznej. Zmienne binarne $X_{2.1}, X_{2.2}, \boldsymbol{X_{2.3}}$ odpowiadające kategoriom: 1, 2-3, **4 lub więcej**
X_3	Częstość sprawdzania poczty elektronicznej w ciągu dnia. Zmienne binarne $X_{3.1}, \boldsymbol{X_{3.2}}, X_{3.3}$ odpowiadające kategoriom: rzadko, **średnio**, często
X_4	Główni adresaci wysyłanych wiadomości e-mail. Zmienne binarne $X_{4.1}, X_{4.2}, X_{4.3}, X_{4.4}, \boldsymbol{X_{4.5}}$ odpowiadające kategoriom: rodzina, osoby związane z pracą, znajomi, osoby związane ze szkołą, **inni**
X_5	Posiadanie *uczelnianej* skrzynki pocztowej (mail-box) (1 − tak, 0 – nie)
X_6	Posiadanie skrzynki pocztowej (mail-box) na portalu *wp* (1 − tak, 0 – nie)
X_7	Posiadanie skrzynki pocztowej (mail-box) na portalu *onet* (1 − tak, 0 – nie)
X_8	Posiadanie skrzynki pocztowej (mail-box) na portalu *interia* (1 − tak, 0 – nie)
X_9	Posiadanie skrzynki pocztowej (mail-box) na portalu *o2* (*tlen*) (1 − tak, 0 – nie)
X_{10}	Posiadanie skrzynki pocztowej (mail-box) na portalu *gmail* (1 − tak, 0 – nie)
X_{11}	Posiadanie *innej* skrzynki pocztowej (mail-box) (1 − tak, 0 – nie)
X_{12}	Serwery, na których znajduje się najczęściej używana skrzynka pocztowa. Zmienne binarne $X_{12.1}, X_{12.2}, X_{12.3}, X_{12.4}, X_{12.5}, X_{12.6}, \boldsymbol{X_{12.7}}$ odpowiadające kategoriom: *uczelniany*, *wp*, *onet*, *interia*, *o2* (*tlen*), *gmail*, ***inny***
X_{13}	Ocena częstotliwości pojawiania się problemów dotyczących wykorzystania niektórych usług w serwisie *uczelnianej* poczty elektronicznej. Zmienne binarne $X_{13.1}, X_{13.2}, \boldsymbol{X_{13.3}}$ odpowiadające kategoriom: nigdy, rzadko, **często**
X_{14}	Ocena częstotliwości pojawiania się problemów dotyczących wykorzystania niektórych usług w serwisie *wp* posiadanej poczty elektronicznej. Zmienne binarne $X_{14.1}, X_{14.2}, \boldsymbol{X_{14.3}}$ odpowiadające kategoriom: nigdy, rzadko, **często**
X_{15}	Ocena częstotliwości pojawiania się problemów dotyczących wykorzystania niektórych usług w serwisie *onet* posiadanej poczty elektronicznej. Zmienne binarne $X_{15.1}, X_{15.2}, \boldsymbol{X_{15.3}}$ odpowiadające kategoriom: nigdy, rzadko, **często**
X_{16}	Ocena częstotliwości pojawiania się problemów dotyczących wykorzystania niektórych usług w serwisie *interia* posiadanej poczty elektronicznej. Zmienne binarne $X_{16.1}, X_{16.2}, \boldsymbol{X_{16.3}}$ odpowiadające kategoriom: nigdy, rzadko, **często**
X_{17}	Ocena częstotliwości pojawiania się problemów dotyczących wykorzystania niektórych usług w serwisie *o2* posiadanej poczty elektronicznej. Zmienne binarne $X_{17.1}, X_{17.2}, \boldsymbol{X_{17.3}}$ odpowiadające kategoriom: nigdy, rzadko, **często**
X_{18}	Ocena częstotliwości pojawiania się problemów dotyczących wykorzystania niektórych usług w serwisie *gmail* posiadanej poczty elektronicznej. Zmienne binarne $X_{18.1}, \boldsymbol{X_{18.2}}, \sout{X_{18.3}}$ odpowiadające kategoriom: nigdy, **rzadko**, ~~często~~
X_{19}	Ocena częstotliwości pojawiania się problemów dotyczących wykorzystania niektórych usług w *innym* serwisie posiadanej poczty elektronicznej. Zmienne binarne $X_{19.1}, X_{19.2}, \boldsymbol{X_{19.3}}$ odpowiadające kategoriom: nigdy, rzadko, **często**
X_{20}	Ocena sposobu prezentacji danych (interfejs) w serwisie *uczelnianej* poczty elektronicznej. Zmienne binarne $\boldsymbol{X_{20.1}}, X_{20.2}, X_{20.3}$ odpowiadające kategoriom: **nieprzyjazny**, zadowalający, przyjazny

1	2
X_{21}	Ocena sposobu prezentacji danych (interfejs) w serwisie posiadanej poczty elektronicznej na *wp*. Zmienne binarne $\boldsymbol{X_{21.1}}, X_{21.2}, X_{21.3}$ odpowiadające kategoriom: **nieprzyjazny**, zadowalający, przyjazny
X_{22}	Ocena sposobu prezentacji danych (interfejs) w serwisie posiadanej poczty elektronicznej na *onet*. Zmienne binarne $\boldsymbol{X_{22.1}}, X_{22.2}, X_{23.3}$ odpowiadające kategoriom: **nieprzyjazny**, zadowalający, przyjazny
X_{23}	Ocena sposobu prezentacji danych (interfejs) w serwisie posiadanej poczty elektronicznej *interia*. Zmienne binarne $\boldsymbol{X_{23.1}}, X_{23.2}, X_{23.3}$ odpowiadające kategoriom: **nieprzyjazny**, zadowalający, przyjazny
X_{24}	Ocena sposobu prezentacji danych (interfejs) w serwisie posiadanej poczty elektronicznej na *o2*. Zmienne binarne $\boldsymbol{X_{24.1}}, X_{24.2}, X_{24.3}$ odpowiadające kategoriom: **nieprzyjazny**, zadowalający, przyjazny
X_{25}	Ocena sposobu prezentacji danych (interfejs) w serwisie posiadanej poczty elektronicznej na *gmail*. Zmienne binarne $\boldsymbol{X_{25.1}}, X_{25.2}, X_{25.3}$ odpowiadające kategoriom: **nieprzyjazny**, zadowalający, przyjazny
X_{26}	Ocena sposobu prezentacji danych (interfejs) w serwisie innej posiadanej poczty elektronicznej. Zmienne binarne ~~$X_{26.1}$~~, $\boldsymbol{X_{26.2}}, X_{26.3}$ odpowiadające kategoriom: ~~nieprzyjazny~~, **zadowalający**, przyjazny
X_{27}	Korzystanie z poczty elektronicznej za pomocą urządzenia przenośnego (1 − tak, 0 – nie)
X_{28}	Płeć (1 − kobieta, 0 – mężczyzna)
X_{29}	Wiek. Zmienne binarne $X_{29.1}, X_{29.2}, \boldsymbol{X_{29.3}}$ odpowiadające kategoriom: 18-20 lat, 21-23, **24 lub więcej lat**
X_{30}	Kierunek studiów. Zmienne binarne $X_{30.1}, \boldsymbol{X_{30.2}}, X_{30.3}, X_{30.4}$ odpowiadające kategoriom: Informatyka (ZUT w Szczecinie), **Lekarski (PUM)**, Zarządzanie i inżynieria produkcji (ZUT w Szczecinie), Ekonomia (US)
X_{31}	Rodzaj studiów. Zmienne binarne $X_{31.1}, \boldsymbol{X_{31.2}}$, ~~$X_{31.3}$~~, ~~$X_{31.4}$~~ odpowiadające kategoriom: stacjonarne I stopnia, **stacjonarne II stopnia**, ~~niestacjonarne I stopnia~~, ~~niestacjonarne II stopnia~~ .

Źródło: opracowanie własne na podstawie kwestionariusza ankiety [Górski 2011].

4. Wyniki estymacji uporządkowanego modelu probitowego

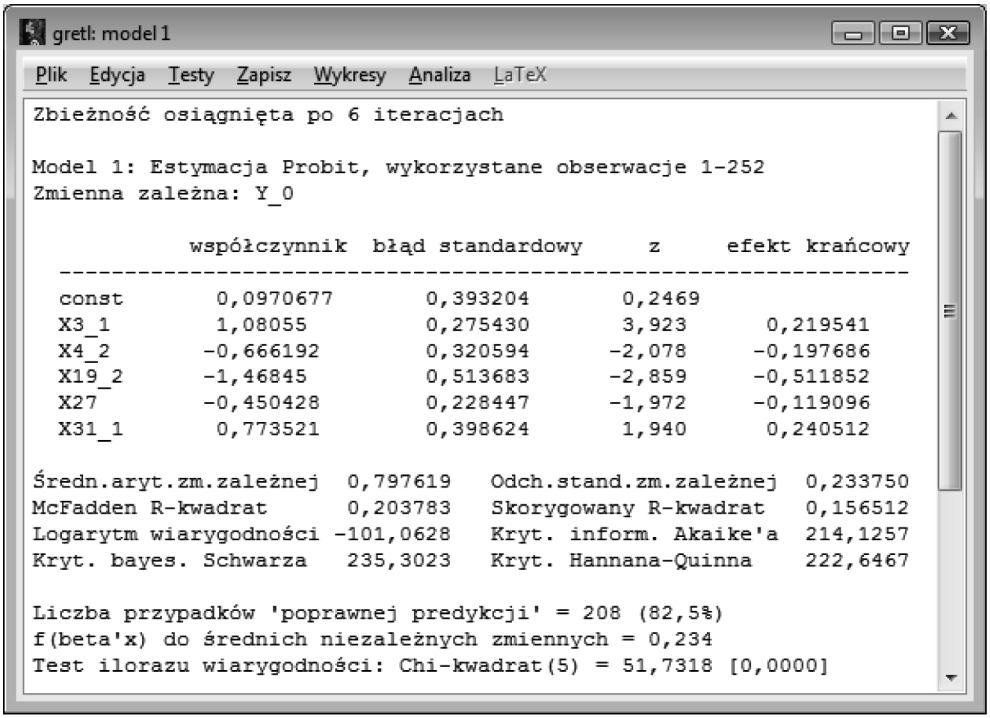
Liczności występowania kolejnych wartości zmiennej zależnej oraz wskaźniki struktury przedstawiono w tab. 2. Widać, że najczęściej (w niemal 80% przypadków) badani wysyłają przeciętnie nie więcej niż 2 wiadomości e-mail w ciągu dnia. Prawie 17% respondentów wysyła dziennie średnio od 3 do 5 wiadomości. Niecałe 3% ankietowanych – od 7 do 9, a powyżej 10 – mniej niż 1% badanych.

Współczynniki uporządkowanego modelu probitowego zostały wyznaczone w niekomercyjnym programie gretl 1.9.5cvs [http://www.kufel.torun.pl/] metodą regresji krokowej wstecz (por. rys. 1). Statystyczną łączną istotność (na poziomie 0,05) oszacowanych parametrów modelu sprawdzono testem ilorazu wiarygodności.

Tabela 2. Rozkład wartości zmiennej zależnej

Wartości	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$
Liczność	201	42	7	2
Odsetek	0,798	0,167	0,028	0,008

Źródło: opracowanie własne na podstawie danych [Górski 2011].



Rys. 1. Okno wynikowe dla modelu probitowego

Źródło: obliczenia własne w programie gretl.

Z otrzymanych rezultatów wynika, że zmiany w przeciętnej liczbie wysyłanych dziennie wiadomości e-mail można wyjaśnić częstością sprawdzania poczty elektronicznej, kategorią odbiorcy, częstością występowania problemów w serwisach mniej popularnych, korzystaniem z poczty na urządzeniu przenośnym i rodzajem studiów.

Współczynników modelu probitowego nie można interpretować wprost, dlatego obliczono prawdopodobieństwa warunkowe przyjęcia kolejnych wartości (por. wzory (2)-(3) i tab. 3) jako średnie z teoretycznych prawdopodobieństw warunkowych w próbie oraz wpływy cząstkowe istotnych zmiennych objaśniających na te prawdopodobieństwa, obliczone w punkcie średnich wartości tych zmiennych (por. wzór (4) i tab. 3).

Uzyskany model probitowy jest dobrze dopasowany do danych. Wskazuje na to m.in. brak istotnych różnic między częstościami występowania poszczególnych wartości zmiennej zależnej (por. tab. 2) a średnimi prawdopodobieństw warunkowych (por. tab. 3).

Tabela 3. Prawdopodobieństwa warunkowe $P(Y = j)$ i efekty krańcowe

Średnia liczba wysyłanych wiadomości w ciągu dnia	$Y = 0$ (nie więcej niż 2)	$Y = 1$ (od 3 do 5)	$Y = 2$ (od 6 do 9)	$Y = 3$ (10 lub więcej)
$P(Y = j)$	0,795	0,169	0,030	0,006
Wpływ $X_{3.1}$	0,220	–0,191		
Wpływ $X_{4.2}$	–0,198	–0,022	0,210	
Wpływ $X_{19.2}$	–0,512	0,493		
Wpływ X_{27}	–0,119	0,086	0,003	
Wpływ $X_{31.1}$	0,241	–0,058	–0,022	

Źródło: obliczenia własne.

5. Wnioski

Interpretacja wpływów cząstkowych na prawdopodobieństwa warunkowe (por. tab. 3) pozwala sformułować następujące konkluzje:

- przyrostowi prawdopodobieństwa wysyłania do 2 wiadomości dziennie najbardziej sprzyja studiowanie na studiach stacjonarnych I stopnia; rzadkie sprawdzanie poczty sprzyja nieco słabiej;
- na spadek tego prawdopodobieństwa ma wpływ rzadkość pojawiania się problemów z wykorzystaniem usług w serwisach innych niż popularne, pisanie w sprawach zawodowych oraz sprawdzanie poczty na urządzeniach przenośnych;
- rzadkość pojawiania się problemów z wykorzystaniem usług w serwisach innych niż popularne ma wpływ na przyrost prawdopodobieństwa wysyłania od 3 do 5 wiadomości dziennie, natomiast rzadkie sprawdzanie poczty obniża to prawdopodobieństwo;
- pisanie wiadomości związanych z pracą wywołuje przyrost prawdopodobieństwa wysyłania od 6 do 9 wiadomości dziennie.

Interesujące jest również to, że zaproponowany model probitowy nie wykazuje zależności częstości wysyłania wiadomości elektronicznych w ciągu dnia od: okresu posiadania adresu poczty elektronicznej, liczby posiadanych skrzynek pocztowych, rodzaju serwisu pocztowego, problemów z wykorzystaniem usług na portalach popularnych, sposobu prezentacji danych w serwisie, płci i wieku respondenta oraz kierunku studiów.

Literatura

Dudek H., *An Identification of Farmers' Households in Danger of Poverty on the Ground of Ordered Logit Model*, [w:] Taksonomia 14, red. K. Jajuga, M. Walesiak, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu nr 1169, UE, Wrocław 2007.

Górski Ł., *Analiza statystyczna poczty internetowej*, Praca inżynierska, ZUT, Szczecin 2011.

http://www.kufel.torun.pl/.

Kostrzewska J., *Wpływ cech społeczno-demograficznych na wysokość wynagrodzenia zamężnych kobiet*, Zeszyty Naukowe UEK, Seria: Metody Analizy Danych, Kraków 2010.

Mazurek-Łopacińska K., *Badania marketingowe. Teoria i praktyka*, Wydawnictwo Naukowe PWN, Warszawa 2005.

Zavoina W., McKelvey R., *A statistical model for the analysis of ordinal level dependent variables*, „Journal of Mathematical Sociology" 1975.

EVALUATION OF INTENSITY OF MAILBOXES USING WITH THE ORDERED PROBIT MODEL

Summary: This article aims to analyze the opinion of studying users of mailboxes on their mailboxes. The data source survey questionnaires were completed by students of selected departments of Szczecin universities, which included questions about both the characteristics of the mailbox, and the intensity of its use. For the evaluation of e-mail an ordered probit model is used in which there is a qualitative dependent variable, describing structured categories assuming a normal distribution of the random component.

Keywords: ordered probit model, questionnaire survey, stepwise regression.