

Beata Bieszk-Stolorz

Uniwersytet Szczeciński
e-mail: beatus@wneiz.pl

BADANIE CZASU TRWANIA W BEZROBOCIU Z WYKORZYSTANIEM MODELI REGRESJI DLA ZDARZEŃ POWTARZAJĄCYCH SIĘ

ANALYSIS OF THE DURATION IN UNEMPLOYMENT WITH USE OF THE REGRESSION MODELS FOR THE RECURRENT EVENTS

DOI: 10.15611/pn.2018.507.02

JEL Classification: C41, J64

Streszczenie: W analizie czasu trwania badaniu podlegają często procesy wielokrotnie generujące określone zdarzenia – procesy zdarzeń powtarzających się. Celem artykułu jest analiza wielokrotnych epizodów czasu trwania w bezrobociu rejestrowanym oraz porównanie z wynikami dla pojedynczego epizodu w zależności od płci. W badaniu wykorzystano warstwowy model Coxa. Kolejne epizody, oddzielone okresami wyrejestrowania z urzędu, tworzą nieciągłe przedziały ryzyka. Zastosowano podejście warunkowe, w którym czas do zajścia kolejnego zdarzenia wyznaczono, korzystając z pojęcia luki czasowej (początek każdego epizodu zostaje zrestartowany do zera). W badaniu wykorzystano dane indywidualne osób zarejestrowanych w Powiatowym Urzędzie Pracy w Szczecinie. Płeć była silną determinantą intensywności pierwszych wyrejestrowań z dowolnego powodu, podjęcia pracy i wykreślenia, natomiast nie determinowała intensywności czwartych i kolejnych wyrejestrowań oraz wyrejestrowania z pozostałych powodów dla każdego z epizodów.

Słowa kluczowe: zdarzenia powtarzające się, warstwowy model Coxa, nieciągłe przedziały ryzyka, bezrobocie.

Summary: In the survival analysis, processes generating the particular events repeatedly are analysed (the recurrent events). The goal of the article is the analysis of multiple episodes of the duration in the registered unemployment and comparison with the results for the single episode with respect to gender by means of the stratified Cox model. Subsequent episodes, separated by the deregistration periods create discontinuous risk intervals. Therefore, the conditional approach was used in which the time to the subsequent event was estimated by using the concept of time gap. The individual data for persons registered in the Poviats Labour Office in Szczecin were used. Gender was the strong determinant of intensity of the first deregistration because of any reason, accepting the job and refusal, but it was not the determinant of intensities of the fourth and subsequent deregistrations. It was not also the determinant of intensities of deregistration because of other reasons.

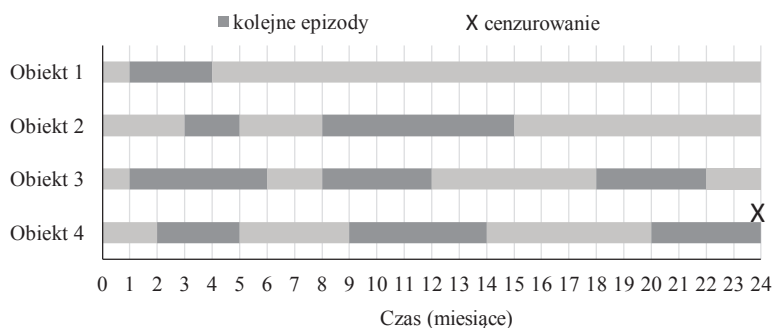
Keywords: recurrent events, stratified Cox model, discontinuous risks intervals, unemployment.

1. Wstęp

W analizie przeżycia badaniu podlega czas trwania jednostki w danym stanie, aż do momentu wystąpienia określonego zdarzenia kończącego obserwację. Zdarza się, że analizowane są procesy, w trakcie których jednostka może kilkakrotnie znaleźć się w określonym przez badanie stanie. Procesy zdarzeń powtarzających się (*recurrent events*) definiuje się jako procesy, które wielokrotnie generują określone zdarzenia [Cook, Lawless 2007]. Można je analizować wykorzystując do tego celu wybrane metody z obszaru analizy przeżycia. Metody te w naukach technicznych służą do badania przestojów na liniach montażowych lub analizy procesów wykrywania i usuwania błędów oprogramowania. W naukach medycznych badaniu podlega czas do nawrotu objawów choroby [Sagara i in. 2014]. W naukach ekonomiczno-społecznych analizować można czas kolejnych wejść i wyjść w sferę ubóstwa [Sączewska-Piotrowska 2015] a także czas kolejnych roszczeń gwarancyjnych lub ubezpieczeniowych. Analizowaną zmienną losową w takich badaniach jest czas do wystąpienia zdarzenia.

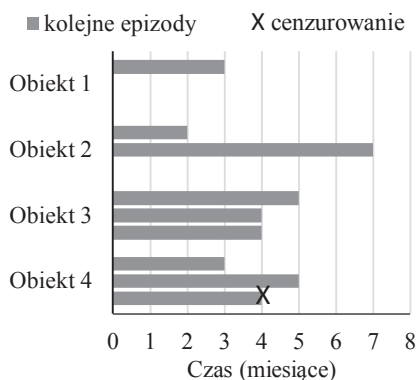
Celem artykułu jest analiza wielokrotnych epizodów czasu trwania w bezrobociu rejestrowanym oraz porównanie otrzymanych wyników z wynikami dla pojedynczego epizodu. Korzystając z warstwowego modelu regresji Coxa wyznaczono intensywność względną wyrejestrowania z urzędu kobiet w stosunku do mężczyzn w poszczególnych epizodach. Za zdarzenia kończące epizod przyjęto wyrejestrowanie z dowolnego powodu, podjęcie pracy, wykreślenie oraz pozostałe przyczyny.

W naukach medycznych w analizie czasu do wystąpienia objawów przewlekłego schorzenia mówi się o ciągłych przedziałach ryzyka. Tak jest w przypadku ataków astmy, napadów padaczkowych oraz złamań w osteoporozie. W badaniu czasu trwania w bezrobociu należy uwzględnić fakt, że kolejne epizody mogą być oddzielone okresami wyrejestrowania z urzędu (rys. 1). Tworzą więc nieciągłe przedziały ryzyka [Guo, Gill, Allore 2008]. W prezentowanym badaniu zastosowano zatem po-



Rys. 1. Przykład obiektów z nieciągłymi przedziałami ryzyka

Źródło: opracowanie własne.



Rys. 2. Zdarzenia powtarzające się zrestartowane do 0

Źródło: opracowanie własne.

dejskie warunkowe [Hosmer, Lemeshow 1999, s. 308–311; Machin, Cheung, Parmar 2006, s. 247; Aalen, Borgan, Gjessing 2008, s. 473]. Czas do zajścia kolejnego zdarzenia wyznaczono, korzystając z pojęcia luki czasowej [Prentice, Williams, Peterson 1981]. Początek każdego epizodu został zrestartowany do zera (rys. 2).

2. Dane wykorzystane w badaniu

W badaniu wykorzystano dane indywidualne osób bezrobotnych zarejestrowanych po raz pierwszy w 2013 roku w Powiatowym Urzędzie Pracy w Szczecinie. Obserwowano całą historię zarejestrowań jednostek do końca 2014 roku. Każda z historii składa się z epizodów będących okresami zarejestrowania w urzędzie. Analizowano czas trwania kolejnych epizodów. W rejestrach urzędów pracy podaje się kilkadziesiąt powodów wyrejestrowania osoby bezrobotnej. Korzystając z wcześniejszych badań dotyczących zdarzeń konkurujących [Bieszk-Stolorz 2017a; Bieszk-Stolorz 2017b], powody te pogrupowano i rozpatrywano trzy grupy przyczyn wyrejestrowania z urzędu: podjęcie pracy (Praca), wykreślenie z urzędu z przyczyn będących po stronie osoby zarejestrowanej (Wykreślenie) oraz z pozostałych przyczyn (Pozostałe). Na przyczynę Praca składa się podjęcie pracy subsydiowanej, niesubsydiowanej i działalności gospodarczej. Przez Wykreślenie rozumie się usunięcie z rejestru z powodów: niestawiennictwa w urzędzie w wyznaczonym terminie lub odmowy przyjęcia propozycji zatrudnienia. Na powód Pozostałe składa się wyrejestrowanie, m.in. z powodu przyznania renty, zasiłku lub emerytury, wyjazdu za granicę, śmierci.

Każdej osobie zarejestrowanej w urzędzie pracy można przyporządkować jej historię zarejestrowania. Na każdą taką historię składają się epizody, czyli kolejne okresy bezrobocia rejestrowanego. Po wstępnej analizie liczby epizodów w historiach zarejestrowań podjęto decyzję o ich podziale na cztery grupy: z jednym,

dwoma, trzema oraz czterema lub więcej epizodami. Wyodrębnienie tej ostatniej grupy wynikało z małej liczebności historii osób zarejestrowanych z co najmniej czterema epizodami. Łącznie analizowano 5418 historii, z czego 2644 dotyczyły kobiet. Najwięcej z nich (4100) składało się z jednego, a 1078 z dwóch epizodów (tab. 1). Pierwsze epizody kończyły się najczęściej podjęciem pracy, a następnie wykreśleniem (zarówno dla kobiet, jak i mężczyzn). W przypadku drugich epizodów sytuacja dla kobiet była analogiczna, a wśród mężczyzn dominowało wykreślenie, a następnie podjęcie pracy (tab. 2). Ze względu na podział zdarzeń kończących kolejne epizody pojawiły się dwa rodzaje obserwacji cenzurowanych prawostronnie. W przypadku dowolnego powodu wyrejestrowania badana jednostka zniknęła z pola widzenia, co związane było z końcem okresu obserwacji. Był to pierwszy rodzaj cenzurowania. W analizie epizodów według przyczyn wyrejestrowania dochodził jeszcze drugi rodzaj – wyrejestrowanie z powodów innych niż analizowany.

Tabela 1. Historie zarejestrowań według liczby epizodów

Liczba epizodów	Liczba historii (kobiety)
1	4100 (2073)
2	1078 (476)
3	193 (73)
4 i więcej	47 (22)
Razem	5418 (2644)

Źródło: opracowanie własne.

Tabela 2. Liczebność kolejnych epizodów

Numer epizodu	Razem (kobiety)	Typ obserwacji	Dowolny powód (kobiety)	Praca (kobiety)	Wykreślenie (kobiety)	Pozostałe (kobiety)
1	5418 (2644)	Pełne	5 048 (2448)	2 790 (1359)	1862 (710)	696 (379)
		Cenzurowane	370 (196)	2 628 (1285)	3556 (1934)	4722 (2265)
2	1318 (571)	Pełne	1035 (428)	466 (232)	472 (156)	97 (40)
		Cenzurowane	283 (143)	852 (339)	846 (415)	1221 (531)
3	240 (95)	Pełne	162 (64)	82 (39)	65 (17)	15 (8)
		Cenzurowane	78 (31)	158 (56)	175 (78)	225 (87)
4 lub kolejny	47 (22)	Pełne	32 (18)	19 (11)	11 (6)	2 (1)
		Cenzurowane	15(4)	28 (11)	36 (16)	45 (21)
Razem	7023 (3332)	Pełne	6 277 (2958)	3 357 (1641)	2410 (889)	810 (428)
		Cenzurowane	746 (374)	3 666 (1691)	4613 (2443)	6213 (2904)

Źródło: opracowanie własne.

3. Metodyka badania

W badaniu wykorzystano wybrane metody analizy trwania. W pierwszym etapie analizy do oceny intensywności wyrejestrowania z urzędu z dowolnego powodu oraz trzech głównych przyczyn dla pojedynczego epizodu wykorzystano model hazardu Coxa [Cox 1972; Bieszk-Stolorz, Markowicz 2012]:

$$h(t, X) = h_0(t) \exp\left(\sum_{i=1}^n \beta_i X_i\right), \quad (1)$$

gdzie: t – czas, $X = [X_1, X_2, \dots, X_n]$ – wektor zmiennych, $h_0(t)$ – hazard bazowy.

W badaniach rynku pracy model ten wykorzystywany jest m.in. do analizy aktywności ekonomicznej ludności [Landmesser 2013] i oceny determinant czasu pozostawania bez pracy [Bieszk-Stolorz 2013; Grzenda 2017].

Drugi etap badania obejmował analizę wielokrotnych epizodów z wykorzystaniem warstwowego modelu Coxa [Kleinbaum, Klein 2005, s. 352]:

$$h_g(t, X) = h_{0g}(t) \exp\left(\sum_{i=1}^n \beta_{ig} X_i\right), \quad (2)$$

gdzie: t – czas, $X = [X_1, X_2, \dots, X_n]$ – wektor zmiennych, $g = 1, 2, \dots, k$ – warstwy, $h_{0g}(t)$ – hazard bazowy w warstwie g .

Badano intensywność wychodzenia z bezrobocia w zależności od płci osoby wyrejestrowanej. Dychotomiczna zmienna X przyjęła wartość 1 dla kobiet i 0 dla mężczyzn. Model (1) ma w związku z tym postać:

$$h(t, X) = h_0(t) \exp \beta X. \quad (3)$$

W modelu (2) warstwy (g) zdefiniowano następująco: 1 – pierwszy epizod, 2 – drugi epizod, 3 – trzeci epizod, 4 – czwarty i kolejny epizod. Przyjął on postać:

$$h_g(t, X) = h_{0g}(t) \exp \beta_g X \text{ dla } g = 1, 2, 3, 4. \quad (4)$$

Hazard względny, wyrażający intensywność wychodzenia z bezrobocia kobiet w stosunku do mężczyzn, dla obu modeli wyznaczono, korzystając odpowiednio ze wzorów:

$$HR = \exp \beta, \quad (5)$$

$$HR_g = \exp \beta_g \text{ dla } g = 1, 2, 3, 4. \quad (6)$$

4. Analiza intensywności wyrejestrowań z urzędu pracy

Właściwe badanie poprzedzono analizą średniego czasu trwania epizodów (tab. 3). W większości przypadków ich średni czas trwania był dłuższy dla kobiet niż dla mężczyzn. Wyjątek stanowił trzeci epizod dla dowolnego powodu wyrejestrowania i podjęcia pracy oraz drugi epizod dla wykreślenia. Kobiety dłużej czekały na podjęcie pracy, ale też miały dłuższy czas do wykreślenia, czyli rezygnacji ze współpracy z urzędem. Średnie czasy trwania były krótsze dla kolejnych epizodów. Kobiety miały krótszy czas trwania dowolnego epizodu z powodu podjęcia pracy (4,90 miesiąca) niż wykreślenia (5 miesięcy). W przypadku mężczyzn sytuacja była odwrotna: dłuższy średni czas do podjęcia pracy (4,36 miesiąca) niż wykreślenia (4,01 miesiąca).

Tabela 3. Średni czas trwania epizodów według przyczyn wyrejestrowania i płci

Grupy	Czas zarejestrowania (miesiące)				
	Dowolny epizod	1 epizod	2 epizod	3 epizod	4 epizod i więcej
Dowolny powód wyrejestrowania					
Ogółem	4,63	5,07	2,96	2,20	1,36
Kobiety	5,07	5,51	3,14	2,11	1,54
Mężczyźni	4,24	4,66	2,84	2,26	1,13
Praca					
Ogółem	4,65	5,06	2,98	2,31	1,14
Kobiety	4,90	5,28	3,34	1,97	0,94
Mężczyźni	4,36	4,80	2,62	2,62	1,41
Wykreślenie					
Ogółem	4,38	4,86	2,88	1,94	1,87
Kobiety	5,00	5,59	2,72	1,95	2,84
Mężczyźni	4,01	4,40	2,96	1,93	0,70
Pozostałe					
Ogółem	5,34	5,70	3,30	2,79	0,74
Kobiety	5,87	6,18	3,61	3,14	0,33
Mężczyźni	4,76	5,12	3,08	2,40	1,15

Źródło: opracowanie własne.

W pierwszym etapie badania wyznaczono estymatory modeli hazardu Coxa (1) dla pojedynczych epizodów (tab. 4). Kobiety mniej intensywnie niż mężczyźni były wyrejestrowywane z dowolnego powodu (o 26%) i z powodu wykreślenia (o 45%), ale bardziej intensywnie podejmowały pracę (o 10%). Płeć nie determinowała intensywności wyrejestrowania z pozostałych powodów (brak istotności parametru β).

Drugi etap analizy, czyli estymacja parametrów warstwowego modelu Coxa (3), pozwolił na ocenę epizodów wielokrotnych (tab. 5). Kobiety mniej intensywnie niż

Tabela 4. Wyniki estymacji modelu regresji Coxa dla pojedynczych epizodów

Powód wyrejstrowania	β	Błąd standardowy	Statystyka Walda	p	Iloraz hazardu
Dowolny powód	$\chi^2 = 46,4918, p = 0,0000$				
	-0,1725	0,0253	46,3904	0,0000	0,8415
Praca	$\chi^2 = 6,5721, p = 0,0104$				
	0,0931	0,0363	6,5587	0,0104	1,0975
Wykreślenie	$\chi^2 = 200,9263, p = 0,0000$				
	-0,5882	0,0423	193,4751	0,0000	0,5553
Pozostałe	$\chi^2 = 0,1438, p = 0,7045$				
	0,0267	0,0705	0,1437	0,7046	1,0271

Źródło: opracowanie własne.

Tabela 5. Wyniki estymacji warstwowego modelu Coxa dla epizodów wielokrotnych

Powód wyrejstrowania	Numer epizodu	β	Błąd standardowy	Statystyka Walda	p	Iloraz hazardu	
1	2	3	4	5	6	7	
Dowolny powód	1	$\chi^2 = 33,0174, p = 0,0000$					
		-0,1620	0,0282	32,9958	0,0000	0,8505	
	2	$\chi^2 = 11,0894, p = 0,0009$					
		-0,2100	0,0634	10,9747	0,0009	0,8106	
	3	$\chi^2 = 0,1141, p = 0,7355$					
		-0,0543	0,1612	0,1137	0,7360	0,9471	
	4 i więcej	$\chi^2 = 3,1948, p = 0,0739$					
		0,6466	0,3624	3,1824	0,0744	1,9090	
Praca	1	$\chi^2 = 4,5463, p = 0,0330$					
		0,0858	0,0403	4,5353	0,0332	1,0896	
	2	$\chi^2 = 2,0499, p = 0,1522$					
		0,1333	0,0930	2,0517	0,1520	1,1426	
	3	$\chi^2 = 1,6820, p = 0,1947$					
		0,2897	0,2226	1,6937	0,1931	1,3360	
	4 i więcej	$\chi^2 = 2,7102, p = 0,0997$					
		0,7627	0,4671	2,6656	0,1025	2,1440	
	Wykreślenie	1	$\chi^2 = 153,1293, p = 0,0000$				
			-0,5816	0,0478	148,2220	0,0000	0,5590
2		$\chi^2 = 34,6758, p = 0,0000$					
		-0,5635	0,0982	32,9287	0,0000	0,5692	
3		$\chi^2 = 6,1898, p = 0,0129$					
		-0,6706	0,2824	5,6409	0,0176	0,5114	

Tabela 5, cd.

1	2	3	4	5	6	7
	4 i więcej	$\chi^2 = 0,5801, p = 0,4462$				
		0,4853	0,6341	0,5857	0,4441	1,6246
Pozostałe	1	$\chi^2 = 0,4632, p = 0,4962$				
		0,0518	0,0762	0,4625	0,4965	1,0532
	2	$\chi^2 = 1,3466, p = 0,2451$				
		-0,2391	0,2073	1,3309	0,2486	0,7873
	3	$\chi^2 = 0,7845, p = 0,3758$				
		0,4583	0,5178	0,7832	0,3762	1,5813
	4 i więcej	$\chi^2 = 0,0687, p = 0,7932$				
		0,3722	1,4159	0,0691	0,7927	1,4509

Źródło: opracowanie własne.

mężczyźni opuszczali rejestr podczas pierwszego i drugiego epizodu (odpowiednio o 25% i 29%), przy czym intensywniej podejmowały pracę w trakcie trwania pierwszego epizodu (o 9%) oraz mniej intensywnie były wykreślane podczas pierwszego (o 44%), drugiego (o 43%) i trzeciego (o 49%) epizodu. W pozostałych przypadkach płeć nie była determinantą intensywności wychodzenia z bezrobocia (brak istotności parametrów).

5. Zakończenie

Model regresji Coxa pozwolił na ocenę wpływu płci na intensywność wychodzenia z bezrobocia dla dowolnego pojedynczego epizodu. Wykazano, że płeć była determinantą wyrejestrowań z urzędu z dowolnego powodu, z powodu podjęcia pracy i wykreślenia. Warstwowy model regresji Coxa umożliwił ocenę wpływu płci na intensywność wyrejestrowania z urzędu podczas kolejnych epizodów – kolejnych okresów zarejestrowań w urzędzie. Dla pierwszych epizodów płeć była silną determinantą intensywności wyrejestrowań z dowolnego powodu, podjęcia pracy i wykreślenia. W przypadku drugich epizodów płeć determinowała intensywność wyrejestrowania z dowolnego powodu i wykreślenia. Dla trzecich epizodów płeć była determinantą tylko wykreślenia z rejestru. Natomiast nie determinowała intensywności czwartych i kolejnych wyrejestrowań oraz wyrejestrowania z pozostałych powodów dla każdego z czterech epizodów. Przeprowadzona analiza wskazuje na to, że w analizowanym okresie, biorąc pod uwagę dowolny powód wyrejestrowania, mężczyźni intensywniej niż kobiety opuszczali rejestr. Dokładne badanie przyczyn wyrejestrowania informuje o tym, że na wynik ten duży wpływ ma wykreślenie z rejestru. Mężczyźni intensywniej niż kobiety byli wykreśleni, natomiast kobiety intensywniej podejmowały pracę. Prawidłowością tą charakteryzowały się pierwsze trzy okresy zarejestrowania w urzędzie.

Literatura

- Aalen O.O., Borgan O., Gjessing H.K., 2008, *Survival and Event History Analysis. A Process Point of View*, Springer, New York.
- Bieszk-Stolorz B., 2013, *Analiza historii zdarzeń w badaniu bezrobocia*, Volumina.pl Daniel Krzanowski, Szczecin.
- Bieszk-Stolorz B., 2017a, *Cumulative Incidence Function in Studies on the Duration of the Unemployment Exit Process*, Folia Oeconomica Stetinensia, vol. 17 (2017), iss. 1, s. 138–150, DOI: 10.1515/fofi-2017-0011.
- Bieszk-Stolorz B., 2017b, *Funkcja skumulowanej częstości i modele hazardu w ocenie konkurujących form wyjścia z bezrobocia*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 469, Taksonomia, nr 29, *Klasyfikacja i analiza danych – teoria i zastosowania*, s. 21–31, DOI: 10.15611/pn.2017.469.02.
- Bieszk-Stolorz B., Markowicz I., 2012, *Modele regresji Coxa w analizie bezrobocia*, CeDeWu, Warszawa.
- Cook R.J., Lawless J.F., 2007, *The Statistical Analysis of Recurrent Events*, Springer, New York.
- Cox D.R., 1972, *Regression models and life-tables*, Journal of the Royal Statistical Society, Series B, vol. 34, no. 2, s. 187–220.
- Grzenda W., 2017, *Modelling the Duration of the First Job Using Bayesian Accelerated Failure Time Models*, Folia Oeconomica. Acta Universitatis Lodzianensis, vol. 4, no. 330, s. 19–38, DOI: 10.18778/0208-6018.330.02.
- Guo Z., Gill T.M., Allore H.G., 2008, *Modeling repeated time-to-event health conditions with discontinuous risk intervals: an example of a longitudinal study of functional disability among older persons*, Methods of Information in Medicine, vol. 47, iss. 2, s. 107–116.
- Hosmer D.W., Lemeshow S., 1999, *Applied Survival Analysis. Regression Modeling of Time to Event Data*, John Wiley & Sons, New York.
- Kleinbaum D., Klein M., 2005, *Survival Analysis. A Self-Learning Text*, Springer, New York.
- Landmesser J., 2013, *Wykorzystanie metod analizy czasu trwania do badania aktywności ekonomicznej ludności w Polsce*, Wydawnictwo SGGW, Warszawa.
- Machin D., Cheung Y.B., Parmar M.K.B., 2006, *Survival Analysis. A Practical Approach*, 2 ed., John Wiley & Sons, Chichester.
- Prentice R.L., Williams B.J., Peterson A.V., 1981, *On the regression analysis of multivariate failure time data*, Biometrika, vol. 68, iss. 2, s. 373–379.
- Sagara I., Giorgi R., Doumbo O.K., Piarroux R., Gaudart J., 2014, *Modelling recurrent events: comparison of statistical models with continuous and discontinuous risk intervals on recurrent malaria episodes data*, Malaria Journal, vol. 13, s. 293.
- Sączewska-Piotrowska A., 2015, *Badanie ubóstwa z zastosowaniem nieparametrycznej estymacji funkcji przeżycia dla zdarzeń powtarzających się*, Przegląd Statystyczny, R. LXII – zeszyt 1, s. 29–51.