

HENRYK OGRYZKO-WIEWIÓROWSKI

Zastosowanie analizy logliniowej w badaniach społecznych

---

Application of a loglinear analysis in social research

Do badania zależności zmiennych nominalnych najczęściej stosuje się test  $\chi^2$ . O ile jest on w miarę efektywny dla tabel dwuwymiarowych (jedna zmienna niezależna i jedna zależna), o tyle jest mało przydatny do badania większych wiązek zmiennych. Nie całkiem jest konkluzyjny zwłaszcza w sytuacji, gdy interesują nas wzajemne interakcje między wieloma zmiennymi nominalnymi. Chcąc, na przykład, wyjaśnić akceptację metody *in vitro*, możemy pytać, który z czynników ma większy wpływ – przekonania polityczne, religijność, płeć? Z jednej strony słyszymy z mediów, że za dopuszczalnością tej metody zapłodnienia opowiadają się niemal jednogłośnie osoby o poglądach liberalnych, a z drugiej – widzimy, że w ramach obu antagonistycznych partii w naszym kraju (największa rządząca i największa opozycyjna) występują zarówno zwolennicy, jak i przeciwnicy *in vitro*. Wiadomo też, że we wszystkich partiach mamy zarówno ludzi wierzących, jak i niewierzących, osoby o poglądach bardziej liberalnych i bardziej konserwatywnych. Ustalenie zatem, który z tych czynników ma bardziej lub mniej realny wpływ na akceptację metody *in vitro* nie jest łatwe w oparciu o tradycyjne metody analizy statystycznej. Do takich celów bardziej nadaje się analiza logliniowa (*loglinear analysis*). Jej podstawy opracowali w latach 50. M. S. Bartlett<sup>1</sup> oraz S. N. Roy wraz z M. A. Ka-

---

<sup>1</sup> M. S. Bartlett, *Contingency table interactions*, „Journal of the Royal Statistical Society Sup.”, 1953, s. 248–252.

stenbaumem.<sup>2</sup> Istotny wkład w rozwój nowej metody analizy danych wniosły też prace B. S. Everitta<sup>3</sup>, G. J. G. Upton<sup>4</sup>, a także S. E. Fienberga i A. Rinaldo<sup>5</sup>.

Analiza logliniowa jest zbieżna z analizą wariancji stosowaną dla danych ilościowych. Zarówno w jednej, jak i drugiej mówi się o efektach głównych (*main effects*) i o interakcji (*interactions*) pomiędzy zmiennymi. Obecność interakcji wymaga zazwyczaj reinterpretacji efektu głównego. Celem analizy logliniowej jest znalezienie modelu, który najlepiej opisuje oddziaływanie zmiennych nominalnych na zmienną dychotomiczną. Jeśli zmienne niezależne mają charakter ilościowy, to mówimy o modelu regresji logistycznej. Natomiast gdy zmienne niezależne mają charakter jakościowy, wówczas model regresji logistycznej jest równoznaczny z modelem logliniowym.<sup>6</sup> Przy analizie dwu zmiennych, dopasowanie modelu jest równoznaczne z obliczeniem wartości oczekiwanych dla każdej z komórek tabeli krzyżowej i porównanie ich z wartościami obserwowanymi.

Dla testu niezależności  $\chi^2$  obliczanie wartości oczekiwanych w tabeli  $2 \times 2$  polega na pomnożeniu wartości marginesowych przecinających się w danej komórce wierszy i kolumn oraz podzielenie otrzymanego iloczynu przez liczbę badanych osób. Załóżmy, że zbadałiśmy 200 osób (100 kobiet i 100 mężczyzn), którym zadano pytanie (dychotomiczne) o zadowolenie z materialnych warunków życia: „Czy jest P. zadowolona(y) z materialnych warunków życia swojej rodziny?” Rozkład odpowiedzi zawiera tabela 1.

Tabela 1. Zadowolenie z materialnych warunków życia wg płci

	Płeć		Ogółem
	K	M	
T	35 45	55 45	90
N	65 55	45 55	110
	100	100	200

<sup>2</sup> S. N. Roy, M. A. Kastenbaum, *On the hypothesis of no "interaction" in a multi-way contingency table*, Ann. Math. Stat., Vol. 27, 1956, s. 749–757.

<sup>3</sup> B. S. Everitt, *The Analysis of Contingency Tables*, Chapman and Hall, London 1977.

<sup>4</sup> G. J. G. Upton, *The Analysis of Cross-Tabulated Data*, John Wiley, Chichester 1978.

<sup>5</sup> S. E. Fienberg, R. Alessandro, *Three centuries of categorical data analysis: Log-linear models and maximum likelihood estimation*, „Journal of Statistical Planning and Inference”, Vol. 137, November 2007, s. 3430–3445.

<sup>6</sup> A. Stanisław, *Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny*, t. 2, *Modele liniowe i nieliniowe*, StatSoft, Kraków 2007, s. 218.

Wartość oczekiwaną dla pierwszej od lewej górnej komórki otrzymujemy z wyrażenia:

$$\frac{90 \cdot 100}{200} = 45$$

Analogicznie postępujemy z pozostałymi komórkami tabeli. Liczby napisane wytłuszczonym drukiem oznaczają wartości obserwowane, a normalnym – wartości oczekiwane.

Ze wzoru na chi-kwadrat

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e}$$

wynika, że im większa jest różnica pomiędzy wartościami oczekiwanymi i obserwowanymi, tym bardziej prawdopodobna jest zależność pomiędzy obiema zmiennymi. Ze szkolnej trygonometrii wiadomo, że logarytm iloczynu jest równy sumie logarytmów składników iloczynu. Zatem logarytm wartości w komórkach tabeli może być traktowany jako liniowa (addytywna) funkcja logarytmów wartości znajdujących się w poszczególnych komórkach. Wartości zlogarytmizowane, w przeciwieństwie do surowych, nadają się do przekształceń liniowych. W przysłowiowej „czarnej” skrzynce analizy logliniowej realizowany jest proces znajdowania (budowania) takiego modelu interakcji wartości badanych zmiennych, który testując każdą zmianę wartości w komórkach tabeli (poprzez każdorazowe wyliczenie testu dopasowania) powoduje najmniejszy przyrost wartości *chi-kwadrat* po kolejnej rotacji.<sup>7</sup> W pakiecie statystycznym SPSS nazywa się to modelem, który uzyskuje najlepszy wynik w testach dobroci dopasowania (*Goodness-of-fit test statistics*). Dzięki temu otrzymujemy wiedzę na temat roli poszczególnych zmiennych niezależnych w objaśnianiu (warunkowaniu) wartości zmiennej zależnej. Model jest tym lepszy, im wyższy uzyskuje iloraz wiarygodności (*Likelihood ratio chi square*) przy możliwie najniższej wartości  $\chi^2$  (poziom istotności *p* ma być wyższy od 0,05). Mamy zatem sytuację odwrotną do tej, która występuje przy tradycyjnym zastosowaniu testu *chi-kwadrat*. Tam chodzi o to, by różnica między wartościami obserwowanymi i oczekiwanymi była możliwie najwyższa. Dzięki temu uzyskuje się wysoką wartość  $\chi^2$ , a przez to maksymalizację pewności, że wartości zmiennej zależnej (istotnie) warunkowane są przez wartości zmiennej niezależnej. Tutaj natomiast mamy założenia odwrotne: im mniejsze są w komórkach tabeli różnice między wartościami obserwowanymi i oczekiwanymi, tym bardziej dopasowany jest model i większy jest jego iloraz wiarygodności. Minimalizacja różnic między wartościami obserwowanymi i oczekiwanymi oznacza,

<sup>7</sup> P. R. Kinnear, D. G. Colin, *SPSS for Windows Made Simple*, Psychology Press Ltd, East Sussex 1999, s. 323.

że model bazujący na wartościach oczekiwanych jest możliwie najbardziej „dopasowany” do wartości obserwowanych.<sup>8</sup> Poszukiwanie takiego modelu odbywa się w kolejnych krokach analitycznych.

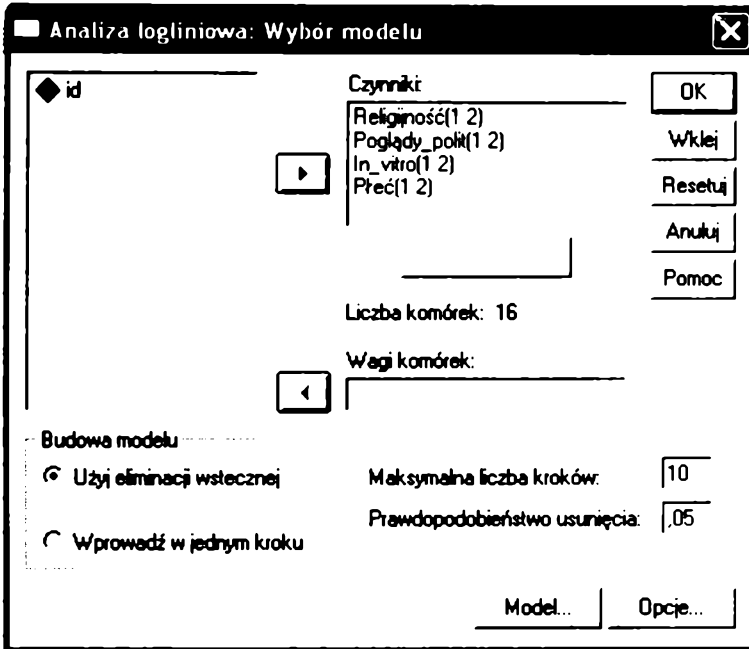
Analizę logliniową zastosowałem w badaniach 350-osobowej grupy studentów – 183 kobiet i 167 mężczyzn. Zadano im trzy pytania: Czy uważa siebie za osobę wierzącą i praktykującą (osoby deklarujące się jako wierzące, ale niepraktykujące nie były brane pod uwagę)?; Czy uważa siebie za osobę o poglądach bardziej liberalnych czy bardziej konserwatywnych?; Czy metoda zapładniania *in vitro* powinna być w naszym kraju zalegalizowana? Chociaż istnieje pokusa, by na każde z pytań udzielić odpowiedzi „średkowej” (np. wierzący, ale niepraktykujący, umiarkowane poglądy polityczne, legalizacja metody *in vitro*), to jednak – w tym sondażu – chodziło o „wyostrzenie” stanowisk. Osobom, którym trudno było zdecydować się na odpowiedź jednoznaczną, albo grzecznie dziękowano za udział w badaniach, albo wykluczano je z badanej zbiorowości w późniejszym etapie procesu badawczego. Łącznie mamy zatem cztery zmienne: opinia na temat legalizacji *in vitro*, poglądy polityczne, religijność oraz płeć. Powstają zatem pytania: Jakie są interakcje pomiędzy tymi zmiennymi? Które z trzech zmiennych mają istotny wpływ na poglądy dotyczące legalizacji metody *in vitro*? W tabeli 2 zamieszczam rozrzut uzyskanych odpowiedzi.

Tabela 2. Rozkład odpowiedzi na zadane pytania

Poglądy polityczne	Religijność	Akceptacja <i>in vitro</i>	Płeć		Ogółem
			Kobieta	Mężczyzna	
Liberalne	Tak	Tak	62	42	104
		Nie	12	10	22
	Nie	Tak	39	62	101
Konserwatywne	Nie	Nie	6	5	11
		Tak	19	13	32
	Tak	Nie	22	12	34
		Tak	10	8	18
		Nie	13	15	28
Ogółem			183	167	350

Przy czterech zmiennych dwuwartościowych otrzymujemy 16 komórek zawierających liczby odzwierciedlające poglądy respondentów. Zastosowanie w tym przypadku analizy logliniowej jest jak najbardziej uzasadnione, a nawet wskazane. Mam nadzieję, że zamieszczony poniżej zrzut ekranowy z pakietu statystycznego SPSS 12 PL ułatwi osobom zainteresowanym korzystanie z omawianej metody analizy danych we własnych badaniach.

<sup>8</sup> Tamże, s. 324.



Tym razem proces obliczeniowy (cztery zmienne) realizowany był w 10 krokach. Liczba kroków jest zawsze różna i zależna zarówno od liczby badanych zmiennych, jak i wzajemnych powiązań. Poniżej zamieszczam fragment tekstu, który pojawił się w dziesiątym kroku analizy.

Step 10

The best model has generating class

Płeć\*Religijność

Poglądy polit.\*In\_vitro

Goodness-of-fit test statistics

Likelihood ratio chi square = 8,29373 DF = 9 P = ,505

Pearson chi square = 7,97611 DF = 9 P = ,537

Z ostatecznego modelu o najwyższym ilorazie wiarygodności (8,29 373) wynika, że w bezpośredniej interakcji pozostają dwie pary zmiennych: „poglądy polityczne i opinia na temat *in vitro*” oraz „płeć i religijność”. Badania wprawdzie nie mają charakteru reprezentatywnego, niemniej trzeba przyznać, że jest to wynik nieco zaskakujący. Okazuje się, że na poglądy w sprawach legalizacji metody *in vitro* nie mają wpływu ani płeć ani religijność respondenta.

Wygenerowany przez model logliniowy wynik można zilustrować testem niezależności  $\chi^2$ , którego kolejne wartości i poziom istotności podane są w poniższych tabelach.

Tabela 3. Zależność akceptacji *in vitro* od poglądów politycznych

Akceptacja in_vitro	Poglądy polityczne		Ogółem
	liberalne	konserwatywne	
tak	205	50	255
nie	33	62	95
Ogółem	238	112	350

$$\chi^2 = 66,3 \quad df = 1 \quad p < 0,001$$

Tabela 4. Zależność akceptacji *in vitro* od płci

Akceptacja in_vitro	Płeć		Ogółem
	liberalne	konserwatywne	
tak	130	125	255
nie	53	42	95
Ogółem	183	167	350

$$\chi^2 = 0,64 \quad df = 1 \quad p < 0,05$$

Tabela 5. Zależność akceptacji *in vitro* od religijności

Akceptacja in_vitro	Religijność		Ogółem
	liberalne	konserwatywne	
tak	136	119	255
nie	56	39	95
Ogółem	192	158	350

$$\chi^2 = 0,88 \quad df = 1 \quad p < 0,05$$

Tabela 6. Zależność religijności od płci

Religijność	Płeć		Ogółem
	kobieta	mężczyzna	
tak	115	77	192
nie	68	90	158
Ogółem	183	167	350

$$\chi^2 = 9,873 \quad df = 1 \quad p < 0,05$$

Wyliczone wartości  $\chi^2$  dla kolejnych par zmiennych potwierdzają trafność wyniku uzyskanego w analizie logliniowej. Wynika z tego, że modele logliniowe stanowią znakomite rozszerzenie i uzupełnienie metod analitycznych dla zmiennych jakościowych. Miejmy nadzieję, że niniejszy tekst ośmieli badaczy życia społecznego do częstszego sięgania po nowsze i często bardziej efektywne metody analizy danych empirycznych.

---

## SUMMARY

To analyse the association of two nominal variables we can use chi-square test. If we have more nominal variables (multi-way contingency tables) and dichotomous dependent variable, better is loglinear analysis. The advent of modern loglinear analysis in the last decades has made it possible to tease out the relationships among the variables in a way that was not possible in applications of traditional methods. In an investigation of the relationships between political views, religiousness, sex and acceptance of method *in vitro* we obtained two pairs of interactions: political views – method *in vitro* and sex – religiousness.