

Analiza danych ankietowych
Lista 8

1. Pokazać, że relacja $Y \perp Z|X$ nie implikuje relacji $Y \perp Z$. Dla dowodu wystarczy podać przykład.

Przykład: Tabela 1 przedstawia prawdopodobieństwa dla układu trzech cech:

X – wykształcenie s – ścisłe, h – humanistyczne,
 Y – płeć k – kobieta, m – mężczyzna
 Z – zarobki w – wysokie, n – niskie.

Tabela 1:

X	Y	Z	
		w	n
s	k	0.08	0.02
	m	0.32	0.08
h	k	0.12	0.18
	m	0.08	0.12

2. Pokazać, że następujące modele logarytmiczno-liniowe reprezentują podane w Tabeli 2 typy zależności między cechami

Tabela 2:

Symbol	Typ zależności	π_{ijk}
(X, Y, Z)	$X \perp Y \perp Z$	$\pi_{i++}\pi_{+j+}\pi_{++k}$
(XY, Z)	$(X, Y) \perp Z$	$\pi_{ij+}\pi_{++k}$
(XZ, YZ)	$X \perp Y Z$	$\pi_{i+k}\pi_{+jk}/\pi_{++k}$

3. W Tabeli 3 podano ciąg hierarchicznie podporządkowanych modeli.
 (a) Wyznaczyć stopnie swobody rozważanych modeli.

Tabela 3:

Oznaczenie	Symbol
M_0	: (XYZ)
M_1	: (XY, YZ, XZ)
M_2	: (XY, YZ)
M_3	: (XY, Z)
M_4	: (X, Y, Z)

(b) A następnie sprawdzić stopnie swobody rozkładu χ^2 statystyk ilorazu wiarogodności

$$\begin{aligned} DF(G^2(M_1|M_0)) &= (I-1)(J-1)(K-1), \\ DF(G^2(M_2|M_1)) &= (I-1)(K-1), \\ DF(G^2(M_3|M_2)) &= (J-1)(K-1), \\ DF(G^2(M_4|M_3)) &= (I-1)(J-1). \end{aligned}$$

4. Pokazać, że estymatory największej wiarogodności $\mu_{ijk}^{(r)}$ w modelach hierarchicznych M_r (patrz Tabela 3) wyrażają się wzorami

Tabela 4:

Oznaczenie	Symbol	Estymator μ_{ijk}
M_2	(XY, YZ)	$\hat{\mu}_{ijk}^{(2)} = \frac{n_{ij+n+jk}}{n_{+j+}}$
M_3	(XY, Z)	$\hat{\mu}_{ijk}^{(3)} = \frac{n_{ij+n++k}}{n}$
M_4	(X, Y, Z)	$\hat{\mu}_{ijk}^{(4)} = \frac{n_{i++n+j+n++k}}{n^2}$

5. Tabela 5 zawiera uzyskane losowo wielkości sprzedaży czterech produktów w trzech grupach wiekowych w dwóch różnych regionach.

Tabela 5:

		produkt			
		A	B	C	D
region 1	20 < wiek < 30	75	28	36	28
	30 < wiek < 40	30	14	17	12
	40 < wiek < 50	40	51	87	18
region 2	20 < wiek < 30	42	28	22	13
	30 < wiek < 40	21	81	14	14
	40 < wiek < 50	18	16	42	55

(a) Zbadać, czy wielkość sprzedaży jest zależna od grup wiekowych i regionów.

(b) Zbadać warunkową niezależność sprzedaży i regionów pod warunkiem grup wiekowych.

(c) Zbadać warunkową niezależność sprzedaży i grup wiekowych pod warunkiem regionów

(d) Wybrać optymalny model.

Przyjąć poziom istotności $\alpha = 0.05$.

6. Metoda iteracyjnego oszacowania proporcjonalnego (ang. **I**terative **P**roportional **F**itting, skrót **IPN**)¹ jest przeznaczona do znajdowania estymatorów $\hat{\mu}_{ijk}$.

¹ Została opracowana przez Deminga i Stephana w 1940

Algorytm - Rozdz. 8.7.2 z książki Agresti (2002), str. 343-344. Zastosować algorytm **IPN** dla danych z Tabeli 8.15 (patrz str. 345, Agresti (2002)) i sprawdzić, że wartości $\mu_{ij}^{(t)}$ zbiegają do wartości w nawiasach.

TABLE 8.15 Marginal Standardization of Attitudes toward Abortion by Years of Schooling

Schooling	Attitude toward Legalized Abortion			Total
	Generally Disapprove	Middle Position	Generally Approve	
Less than high school	209 (49.4)	101 (32.0)	237 (18.6)	(100)
High school	151 (32.8)	126 (36.6)	426 (30.6)	(100)
More than high school	16 (17.8)	21 (31.3)	138 (50.9)	(100)
Total	(100)	(100)	(100)	

Source: Smith (1976).

TABLE 8.3 Alcohol, Cigarette, and Marijuana Use for High School Seniors

Alcohol Use	Cigarette Use	Marijuana Use	
		Yes	No
Yes	Yes	911	538
	No	44	456
No	Yes	3	43
	No	2	279

Source: Data courtesy of Harry Khamis, Wright State University.

7. Napisać algorytm **IPN** dla modeli a) (X, YZ) oraz b) (XZ, YZ) .
8. Używając pakietu *R*, przeprowadzić analizy opisane w Rozdziale 8 książki Agresti (2002) dla danych ankietowych dotyczących używek (A-Alcohol, C-Cigarette, M-Marijuana) z Tabeli 8.3 (patrz str. 323, Agresti (2002)).

Wyznaczyć warunkowe i brzegowe ilorazy szans (patrz Tabela 8.5, Agresti (2002)), testy zgodności G^2 i X^2 dla wszystkich modeli logarytmiczno-liniowych (patrz Tabela 8.6, Agresti (2002)). Za pomocą kryterium Akaike'a wybrać najlepiej dopasowany model oraz zbadać różnice dewiancji między modelami hierarchicznymi. Wyznaczyć Przedziały ufności Walda dla warunkowego ilorazu szans AC , AM i CM w najlepiej dopasowanym modelu.