

METODY EKSPLOKACJI DANYCH
Laboratorium. Modele logitowe. Regresja logistyczna

Zadanie 1

W wybranej losowo grupie studentów jednolitych studiów magisterskich z warszawskich uczelni badano ich stan cywilny w zależności od roku studiów. Wyniki obserwacji zebrano w tabeli 1. Przyjęto przez M oznaczać osoby będące w związku małżeńskim, natomiast przez W – osoby stanu wolnego.

Tabela 1. Dane dotyczące stanu cywilnego badanych studentów w zależności od roku studiów

Rok studiów	1	2	5	1	4	3	2	1	5	2	3	4	1	2
Stan cywilny	W	W	M	W	M	W	M	W	M	W	M	M	W	W
Rok studiów	5	4	3	1	4	5	2	5	3	4	3	2	5	1
Stan cywilny	M	M	M	M	M	W	W	M	W	W	M	W	M	W

W zadaniu proszę:

1. Wyznaczyć zależność stanu cywilnego badanych studentów do roku studiów, zakładając najpierw liniowy model prawdopodobieństwa, a następnie model logitowy.
2. Na podstawie opracowanych modeli i przeprowadzonych obliczeń sformułować własne wnioski.
3. Wyniki analizy proszę zawrzeć w postaci sprawozdania, do którego proszę dodać jako załączniki wszystkie pliki z obliczeniami (obliczenia można przeprowadzić w dowolnie wybranym narzędziu)

Zadanie 2 (dodatkowe)

Pewien duży urząd pracy (w Polsce) zlecił sporządzenie prognozy szans znalezienia pracy przez osobę bezrobotną w ciągu roku od momentu zarejestrowania się w urzędzie. Wyniki obserwacji zebrano w tabeli 2.

Tabela 2. Dane dotyczące bezrobotnych

Numer grupy i	Liczba badanych bezrobotnych n_i	Wiek bezrobotnych (lata)	Średni staż pracy (lata)	Liczba bezrobotnych, którzy znaleźli pracę m_i
1	1000	19-22	1	100
2	1500	23-26	3	160
3	900	27-30	5	110
4	800	31-34	7	130
5	1000	35-38	9	180
6	800	39-42	20	200
7	400	43-46	20	110
8	200	47-50	25	60
9	100	51-54	28	33
10	400	55-58	27	13

W zadaniu proszę:

1. Zbudować model logitowy dla prawdopodobieństwa znalezienia pracy w zależności od wieku rejestrującego się oraz jego stażu pracy.
2. Następnie proszę zbudować liniowy model prawdopodobieństwa znalezienia pracy.
3. Proszę porównać opracowane modele.
4. Wyniki analizy proszę zawrzeć w postaci sprawozdania, do którego proszę dodać jako załączniki wszystkie pliki z obliczeniami (obliczenia można przeprowadzić w dowolnie wybranym narzędziu)

Zadanie 3

Przez cały okres eksploatacji pewnego systemu operacyjnego (OS) zbierano dane dotyczące liczby błędów krytycznych wykrytych w tym czasie w oprogramowaniu. Zebrane obserwacje w układzie miesięcznym przedstawiono w tabeli 3.

Tabela 3. Liczba błędów krytycznych oprogramowania badanego OS w układzie miesięcznym

Nr miesiąca	Liczba błędów	Nr miesiąca	Liczba błędów	Nr miesiąca	Liczba błędów	Nr miesiąca	Liczba błędów	Nr miesiąca	Liczba błędów
1	1	21	9	41	3	61	2	81	8
2	0	22	0	42	2	62	19	82	0
3	0	23	0	43	1	63	7	83	6
4	0	24	3	44	6	64	2	84	5
5	0	25	1	45	3	65	5	85	10
6	0	26	1	46	0	66	1	86	0
7	0	27	0	47	1	67	12	87	5
8	0	28	1	48	1	68	4	88	1
9	0	29	0	49	0	69	6	89	2
10	0	30	2	50	0	70	4	90	1
11	0	31	10	51	14	71	7	91	2
12	0	32	0	52	1	72	2	92	1
13	0	33	16	53	4	73	2	93	0
14	0	34	0	54	1	74	3		
15	0	35	2	55	1	75	8		
16	0	36	2	56	7	76	4		
17	0	37	1	57	14	77	6		
18	0	38	1	58	6	78	3		
19	0	39	1	59	0	79	3		
20	1	40	0	60	1	80	5		

W zadaniu proszę:

1. Wyznaczyć zależność sumarycznej liczby błędów w okresie eksploatacji oprogramowania badanego systemu operacyjnego.
2. Na podstawie opracowanego modelu i przeprowadzonych obliczeń sformułować własne wnioski.
3. Wyniki analizy proszę zawrzeć w postaci sprawozdania, do którego proszę dodać jako załączniki wszystkie pliki z obliczeniami (obliczenia można przeprowadzić w dowolnie wybranym narzędziu)