Sprawozdanie

Laboratorium 2

Zadanie 1

GRUPA WCY21IJ1N1

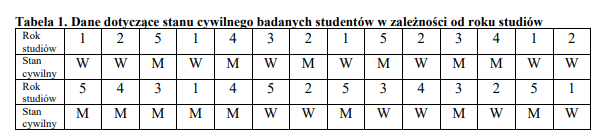
Opracowali:

Karczewski Paweł

Tarkowski Adam

# Treść zadania

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.



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)

# Scenariusz postępowania

Chcemy przetestować 2 modele logitowy i liniowy. Wybieramy więc dane trenujące i testowe losowo. Dla danych trenujących wyliczamy prawdopodobieństwo małżeństwa na określonym roku. Tworzymy oba modele. Wyświetlamy wykres i sprawdzamy, ile razy sprawdził się lepiej. Wyliczać to będziemy w taki sposób, że jeżeli wartość prawdopodobieństwa będzie większa niż 0.5 to uznajemy, że dana osoba jest w małżeństwie.

Aby stworzyć model liniowy użyjemy metody najmniejszych kwadratów, gdzie argumentem będą lata a wartościami prawdopodobieństwa bycia w małżeństwie na danym roku.

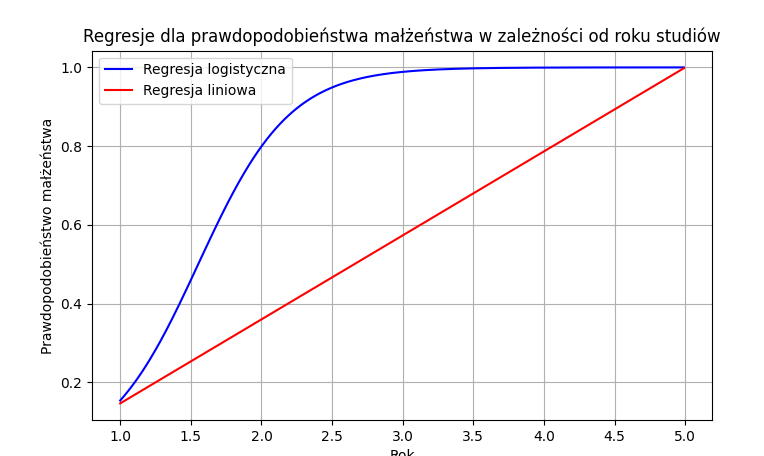
Aby stworzyć model logitowy musimy na początku przekształcić nasze dane dla każdego roku:

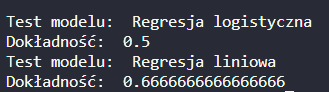
1. Obliczamy prawdopodobieństwo bycia w małżeństwie na danym roku:
2. Obliczamy wartość logitową dla każdego roku
3. Metodą najmniejszych kwadratów tworzymy model liniowy, gdzie argumentami będą lata a wartościami wartości logitowe dla danego roku.
4. Aby wyliczyć prawdopodobieństwo z danego modelu musimy zastosować dany wzór: , gdzie a i b to współczynnik i wyraz wolny modelu liniowego wygenerowanego z logitów.

# Obliczenia

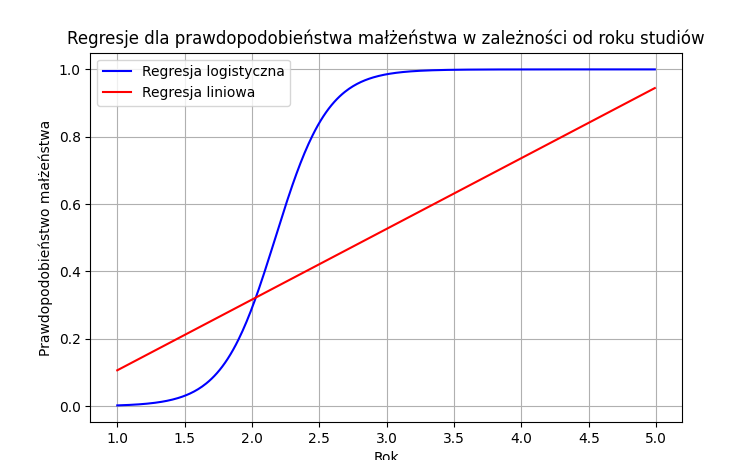
Z racji tego, że dane są losowe a wynik w zależności od wylosowania wychodzi inny to pokażemy kilka przykładów. Predykcje obiema metodami niewiele się różnią i często zwykła liniowa lepiej się sprawdzała niż logistyczna, ale są przypadki, że logistyczna miałą lepsze wyniki.

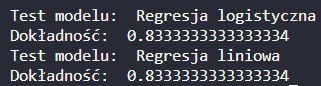
Test 1:



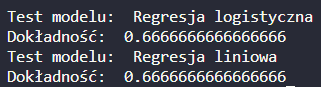
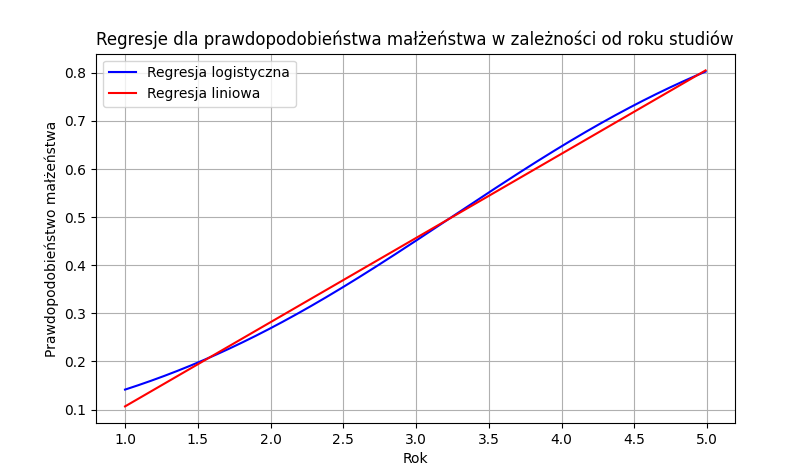


Test 2:

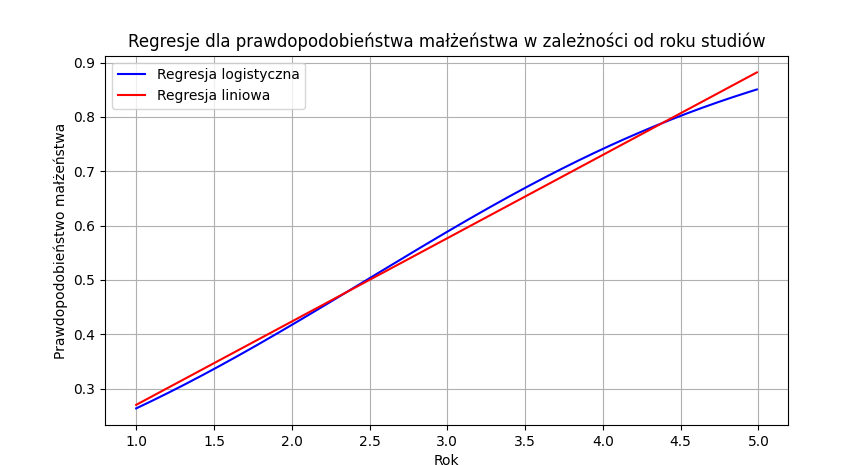


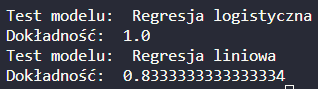


Test 3:

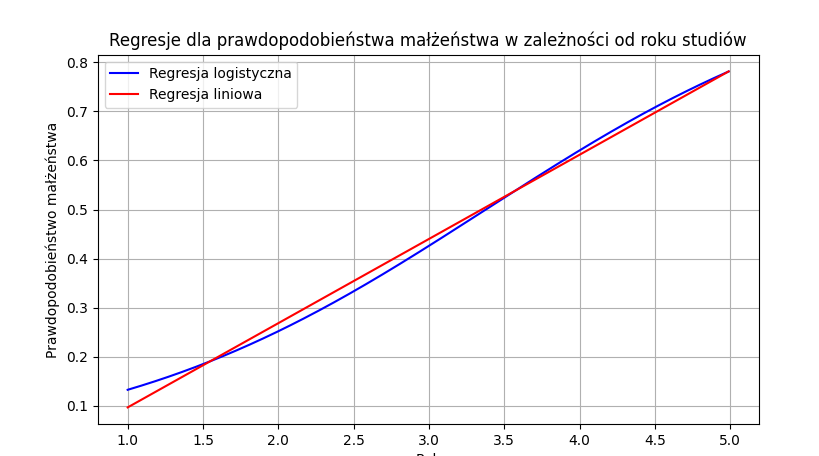


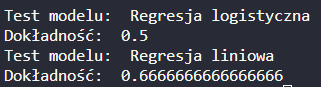
Test 4:





Test 5:





# Podsumowanie

Funkcje liniowa i logitowa dają podobne rezultaty. Widać, że funkcja liniowa często sprawdza się lepiej. Wykres funkcji logitowej jest bardzo podatny na małe zmiany w danych testowych.