

Uogólnione modele liniowe

Laboratorium nr 5

5.1 (Nadwyżka rozproszenia dla modelu logitowego) Zbiór **beetle.txt** zawiera dane dotyczące działania środka owadobójczego na wołki zbożowe. Grupy wołków zbożowych poddawane były różnym stężeniom środka.

- Dopasować model logitowy z $\log(\text{conc})$ jako zmienną objaśniającą.
- Narysować wykres affected/exposed względem $\log(\text{conc})$ (pozwala to ocenić, czy logitowa funkcją łącząca jest w tym przypadku sensownym wyborem). Dorysować do wykresu linię wartości prognozowanych i ocenić jej dopasowanie.
- Narysować wykres logitów empirycznych:

$$\log((\text{affected} + 0.5)/(\text{exposed} - \text{affected} + 0.5))$$

względem $\log(\text{conc})$ i dorysować do niego prostą regresyjną.

- Sprawdzić, w oparciu o wykres normalny rezyduów, czy są w zbiorze wartości odstające.
- Jednym z możliwych powodów niedopasowania, po wyeliminowaniu ewentualnych obserwacji odstających i stwierdzeniu poprawności zastosowanej funkcji łączącej, jest występowanie większego rozproszenia w danych niż te przewidywane przez model logitowy (powodami mogą być np. niejednakowe zastosowanie trucizny na różnych poziomach, niejednakowe warunki traktowania wołków po zastosowaniu środka owadobójczego – nie ma on działania natychmiastowego, niejednakowa żywotność wołków przy rozpoczęciu eksperymentu).

W modelu z nadwyżką rozproszenia zakłada się, że wariancja odpowiedzi wynosi ϕ^* (wartość wariancji dla rozkładu dwumianowego). Estymuje się ją jako $(N - p)^{-1} \sum r_i^2$ (gdzie r_i^2 to rezydua pearsonowskie).

- Oceń wielkość parametru rozproszenia ϕ . Stwierdzić, jak zmieniły się błędy standardowe po uwzględnieniu nadwyżki rozproszenia.
- Ile wynosi p -wartość testu F istotności dla zmiennej $\log(\text{conc})$ (test F stosowany jest w przypadku możliwej nadwyżki rozproszenia)?

5.2 (Nadwyżka rozproszenia dla modelu Poissona) Dane z pliku **lips.dat** dotyczą przypadków wystąpienia raka wargi w 56 hrabstwach Szkocji. Zaobserwowano następującą prawidłowość: rak wargi występuje przede wszystkim u ludzi zatrudnionych w rolnictwie, hodowli i rybołóstwie. Zmienne w zbiorze:

- obs - zaobserwowana liczba przypadków raka wargi w danym hrabstwie
- exp - oczekiwana liczba przypadków raka wargi (wielkość bazująca na wielkości populacji hrabstwa, jej wieku i proporcji liczby kobiet do liczby mężczyzn)
- affpct - procent ludności hrabstwa zatrudnionej w rolnictwie, hodowli i rybołóstwie

Interesuje nas oszacowanie wartości oczekiwanej liczby przypadków raka adjustowanej oczekiwaną liczbą przypadków raka w hrabstwie, tzn. obs/exp.

- Dopasować model poissonowski typu rate dla adjustowanej intensywności wystąpień raka za pomocą instrukcji

```
lip<-glm(obs~affpct+offset(log(exp)),family=poisson,data=...)
```

Oceń jego dopasowanie oraz istotność zmiennej affpct.

- Sporządzić wykres kwantylowy rezyduów i stwierdzić, czy w modelu występują obserwacje odstające. Jeśli tak, ocenić ich wpływ na jakość dopasowania.
- Oszacować parametr rozproszenia dla dopasowanego modelu i dopasować model poissonowski z wyestymowanym parametrem rozproszenia.

5.3 (Model ujemny-dwumianowy) Firma ATT przeprowadziła eksperyment związany z badaniem wpływu pięciu czynników na proces lutowania elementów w płytkach elektronicznych. Zmienną odpowiedzi jest skips – zlicza, ile przeskoków (cecha niepożądana) wystąpiło na płytce. Dane zawarte są w zbiorze **solder**.

- Zastosować model regresji poissonowskiej, uzależniając zmienną odpowiedzi od wszystkich pozostałych zmiennych ze zbioru. Sprawdzić dopasowanie modelu.
- Dopasować model ujemny-dwumianowy, najpierw z (arbitralnie dobranym) parametrem $k = 1$.
- Zastosować model ujemny-dwumianowy z parametrem k wyestymowanym metodą największej wiarygodności.