

Uogólnione modele liniowe

Laboratorium nr 4

- 4.1 W przypadku danych (zgrupowanych) ze zbioru **bliss** porównać wartości dopasowane uzyskane z zastosowania regresji logistycznej, probitowej i funkcji łączącej „complementary log-log”, zarówno w skali π , jak i predyktora liniowego η . Dla skali π narysować wykresy odpowiednich krzywych w zakresie od -2 do 8 (zakres danych - zmiennej *conc* - jest od 0 do 4). Następnie dla wszystkich (trzech) par funkcji łączących narysować na jednym rysunku wykresy ilorazów π_a/π_b i $(1 - \pi_a)/(1 - \pi_b)$, gdzie π_a, π_b - dopasowane wartości π wynikające z użycia odpowiednich funkcji łączących.
- 4.2 Zbiór **discoveries** (opis - patrz: `?discoveries`) zawiera trajektorie szeregu czasowego z liczbą wielkich odkryć od 1860 do 1959 roku. Celem ćwiczenia jest stwierdzenie, czy średnia liczba odkryć w roku jest stała.
- (a) Narysować wykres zależności liczby odkryć od czasu (*discoveries* są obiektem typu *time series* (*ts*), dlatego instrukcja `plot(discoveries)` daje na osi x zmienną o wartościach rzeczywistych).
 - (b) Zakładając, że liczba odkryć w roku ma rozkład Poissona i postulując model poissonowski:
 - i. przeprowadzić test hipotezy o stałości średniej liczby odkryć postulując najprostszy możliwy model, sprawdzając uprzednio jego dopasowanie
 - ii. metoda alternatywna: podobnie jak w postępowaniu z danymi ze zbioru **kypnosis**, dopasować do liczby odkryć trend kwadratowy względem czasu i stwierdzić, czy współczynniki odpowiadające członowi liniowemu i kwadratowemu są istotne.
- 4.3 Zbiór **gala** zawiera informacje o liczbie gatunków żółwi znalezionych na każdej z 30 wysp należących do archipelagu Galapagos oraz o liczbie gatunków stale występujących na danej wyspie (endemicznych). Dodatkowo zbiór zawiera pięć zmiennych geograficznych, które opisują każdą z wysp.
- (a) Dopasować model liniowy **species~.** (oprócz zmiennej *Endemics*). Sporządzić wykres rezyduów (jako funkcji od wartości dopasowanych) i zauważyć wyraźną heteroskedastyczność (niestałość wariancji).
 - (b) Znaleźć (metodą Boxa–Coxa, funkcja `boxcox` w bibliotece **MASS** z opcją `plotit=T` i wybranym odpowiednio zakresem parametru λ) przekształcenie zmiennej *Species* poprawiające problem z poprzedniego punktu. Na podstawie analizy Boxa-Coxa wybrać najbardziej naturalną wartość λ . Dopasować nowy model i sporządzić wykres jego rezyduów.
 - (c) Dopasować model poissonowski. Stwierdzić, czy jest dopasowany.
 - (d) Obliczyć procent dewiacji objaśnianej przez model poissonowski i porównać go z wartością R^2 w modelu liniowym.
 - (e) Sprawdzić, czy ewentualne duże wartości odstające są przyczyną problemu ze słabym dopasowaniem modelu poissonowskiego.
 - (f) Sprawdzić, czy spełnione jest założenie dotyczące rozkładu Poissona o równości średniej i wariancji. W tym celu narysować wykres $(y - \hat{\mu})^2$ jako funkcji od $\hat{\mu}$.