

Zajęcia 2. Obserwacje nietypowe i wpływowe. Regresja nieliniowa

Ćwiczenia

W podstawowym pakiecie R znajdują się dane `stackloss`. Dane zawierają 21 obserwacji 4 zmiennych dotyczących utleniania amoniaku do kwasu azotowego i jego wpływu na wydajność roślin.

<code>Air.Flow</code>	Przepływ powietrza
<code>Water.Temp</code>	Temperatura wody krążącej w cewkach roślin.
<code>Acid.Conc.</code>	Koncentracja kwasu
<code>stack.loss</code>	Zmienna zależna; miara ogólnej efektywności rośliny

Dla tych danych wykonaj następujące polecenia:

- i) Określ model liniowy zależności między zmienną `stack.loss` a wielkością przepływu powietrza.
- ii) Czy któraś z obserwacji może być uznana za obserwację o dużej dźwigni?
- iii) Jeśli tak, przedstaw na wykresie prostą regresji dopasowaną do wszystkich obserwacji oraz prostą dopasowaną po usunięciu tego punktu.
- iv) Przedstaw podsumowania obu modeli. Czy usunięcie punktu o dużej dźwigni spowodowało zmianę w dopasowaniu modelu? Podaj dokładność dopasowania dla obu modeli.
- v) Dla modelu dopasowanego do pełnego zbioru danych wyznacz reszty studentyzowane
- vi) Wyznacz obserwacje odstające korzystając z wektora reszt studentyzowanych.
- vii) Wyznacz obserwacje odstające korzystając z testu `outlierTest`.
- viii) Przedstaw i skomentuj wykres wpływów, zaznacz 3 obserwacje o największych odległościach Cooka.
- ix) Zdecyduj które z zaznaczonych obserwacji są obserwacjami wpływowymi.

x) Zweryfikuj dokonany wybór za pomocą testu `influence.measures`

xi) Usuń obserwacje wpływowe i ponownie dopasuj model.

Zadania

Zadanie 3. Dla danych `stackloss` z powyższego ćwiczenia zdecyduj które ze zmiennych `Water.Temp`, `Acid.Conc.` powinny się znaleźć w modelu ze zmienną zależną `stackloss` (czy jedna z nich (która?) czy też obie). Dla tak wybranego modelu przeprowadź analizę taką jak w ćwiczeniu powyżej (punkty ii)-xi))

Zadanie 4. Plik danych:

`DANE_Predkosc_reakcji.txt`

zawiera 7 obserwacji. Dane w nim zawarte opisują predkość reakcji chemicznej dla różnych wartości koncentracji składowych. Chcemy dopasować model regresji opisujący związek między zmiennymi `Predkosc` i `Koncentracja`.

i) Przedstaw na wykresie punktowym zależność pomiędzy zmienną `Predkosc` i zmienną `Koncentracja`. Czy zależność ma charakter liniowy?

ii) Zastosuj transformację logarytmiczną dla zmiennej `Predkosc`, dla obu zmiennych `Predkosc` i `Koncentracja` oraz tylko dla zmiennej `Koncentracja`. Czy któraś z transformacji prowadzi do uzyskania liniowego charakteru zależności? Przedstaw odpowiednie wykresy punktowe danych. Przedstaw wykresy reszt dla tych trzech modeli. Czy można zauważyć jakiś trend w układzie reszt czy też układają się one w sposób losowy? Co to oznacza dla dopasowanego modelu?

iii) Dopasuj model wielomianowy stosując wielomian stopnia 2 dla zmiennej `Koncentracja`, tzn model

$$\text{Predkosc} = \beta_0 + \beta_1 * \text{Koncentracja} + \beta_2 * (\text{Koncentracja})^2$$

iv) Przedstaw dopasowaną krzywą na wykresie wraz z danymi. Podsumuj wyniki dopasowanego modelu. Jaki jest współczynnik dopasowania tego modelu? Czy jest to model właściwy do opisu danych?

v) Dopasuj model wielomianowy stosując wielomian stopnia 6 dla zmiennej (Koncentracja), tzn model

$$\text{Predkosc} = \beta_0 + \beta_1 * \text{Koncentracja} + \beta_2 * (\text{Koncentracja})^2 + \beta_3 * (\text{Koncentracja})^3 + \beta_4 * (\text{Koncentracja})^4 + \beta_5 * (\text{Koncentracja})^5 + \beta_6 * (\text{Koncentracja})^6$$

vi) Przedstaw dopasowaną krzywą na wykresie wraz z danymi. Podsumuj wyniki dopasowanego modelu. Jaki jest współczynnik dopasowania tego modelu? Czy jest to model właściwy do opisu danych? Czy jest on dobry dla przewidywania wartości zmiennej Predkosc dla nowych obserwacji?