

Regresja nieliniowa

Zagadnienia:

- Regresja linearyzowana
- Regresja nieliniowa
- Funkcja straty

MODELE LINEARYZOWANE

Modele linearyzowane to takie, które przez odpowiednią transformację zmiennych lub parametrów możemy sprowadzić do postaci liniowej. Wybrane modele i ich transformacje zebrano w poniższej tabeli:

Tabela 1. Funkcje modelowe i ich transformacje

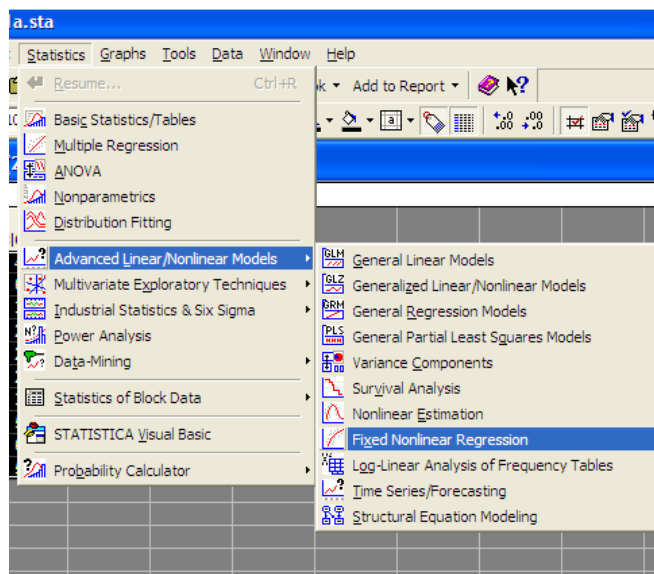
Model	Transformacja	Model po transformacji
$Y=a \cdot X^b$ funkcja potęgowa	logarytmowanie	$\log Y = \log a + b \cdot \log X$
$Y=a \cdot b^X$ funkcja wykładnicza	logarytmowanie	$\log Y = \log a + X \cdot \log b$
$Y=a+b \cdot X+c \cdot X^2$ funkcja kwadratowa	$X_1=X, X_2=X^2$	$Y=a+b \cdot X_1+c \cdot X_2$
$Y=a_0+a_1 \cdot X+a_2 \cdot X^2+a_3 \cdot X^3+\dots+a_n \cdot X^n$ funkcja wielomianowa	$X_1=X, X_2=X^2, X_3=X^3, \dots, X_n=X^n$	$Y=a_0+a_1 \cdot X_1+a_2 \cdot X_2+a_3 \cdot X_3+\dots+a_n \cdot X_n$
$Y=a+b \cdot \frac{1}{X}$ funkcja hiperboliczna	$X'=1/X$	$Y=a+b \cdot X'$
$Y=a+b \cdot \ln X$ funkcja semilogarytmiczna	$X'=\ln X$	$Y=a+b \cdot X'$

Aby zbadać czy związek dwóch zmiennych ma charakter krzywoliniowy należy rozpocząć analizę od sporządzenia wykresu rozrzutu.

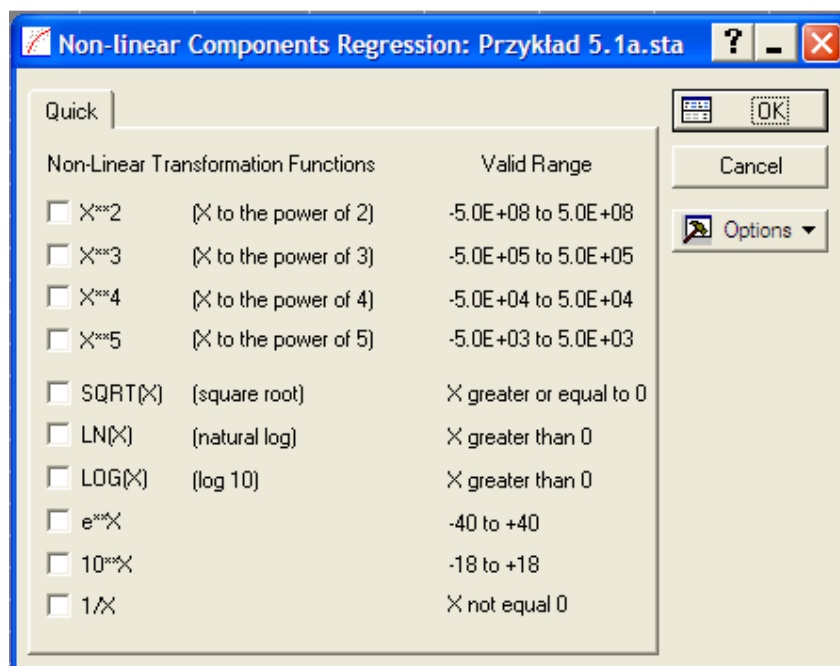
Zadanie 6 (regresja linearyzowana)

Na podstawie wykresu rozrzutu oceń charakter zależności między stężeniem dwóch parametrów biochemicznych w organizmie (PBIOCH1, PBIOCH2). W jaki kształt układają się punkty na wykresie i jaką funkcją można ten związek opisać?

Dokonaj odpowiedniej transformacji zmiennej/ych w celu sprowadzenia modelu krzywoliniowego do modelu liniowego (tabela 1). Zmienną PBIOCH2 przyjmij jako zmienną zależną. W tym celu z menu **Statystyka** wybierz **Zaawansowane modele liniowe i nieliniowe/ Linearyzowana regresja nieliniowa** (rys.1), a następnie po wyborze zmiennych należy przejść do okna **Regresja składników nieliniowych** (rys2). Po wyborze transformacji, w pamięci komputera utworzona zostanie nowa zmienna tymczasowa.



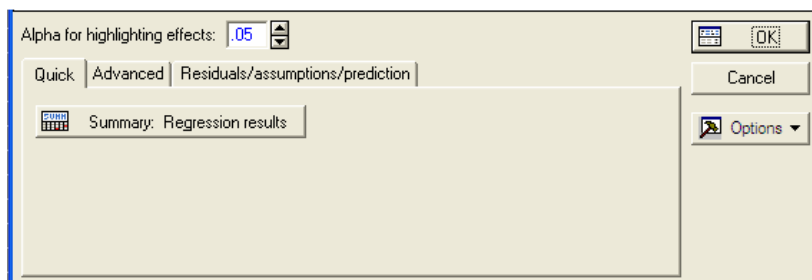
Rys1. Okno wyboru regresji linearyzowanej



Rys. 2 Okno wyboru transformacji

Dokonaj oceny otrzymanych wyników. W tym celu kliknij **Podsumowanie: Wyniki regresji** (rys3).

Dane do zadań zaczerpnięto z książki Adama Stanisza Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny



Rys.3. Podsumowanie: Wyniki regresji

- Zapisz uzyskaną postać modelu (równanie regresji) i zinterpretuj go
- Oceń istotność parametrów modelu
- ile wynosi błąd standardowy estymacji S_e ?
- jakim błędem obarczone są wyliczone współczynniki regresji b_0 i b_1 (S_{b_0} , S_{b_1})?
- na podstawie współczynnika determinacji R^2 oceń dopasowanie modelu. Model wyjaśnia% zaobserwowanej zmienności
- Ile wynoszą znormalizowane współczynniki regresji Beta? Zinterpretuj uzyskane wyniki.

Zadanie 6.1.

Dokonano 15 dniowej obserwacji zmian tyroksyny (T4) u pacjentów chorych na nadczynność tarczycy. Zbuduj model regresji i sprawdź jego poprawność.

Zadanie 7 (regresja linearyzowana)

W pewnym eksperymencie medycznym badano aktywność cholinesterazy (CHE) w surowicy krwi w zależności od wielkości dawki stosowanego leku (LEK) u 18 pacjentów chorych na nadczynność tarczycy. Dokonaj wykresu rozrzutu. Za zmienną zależną przyjmij CHE. Oceń, który model (która postać funkcji regresji) jest najlepiej dopasowany:

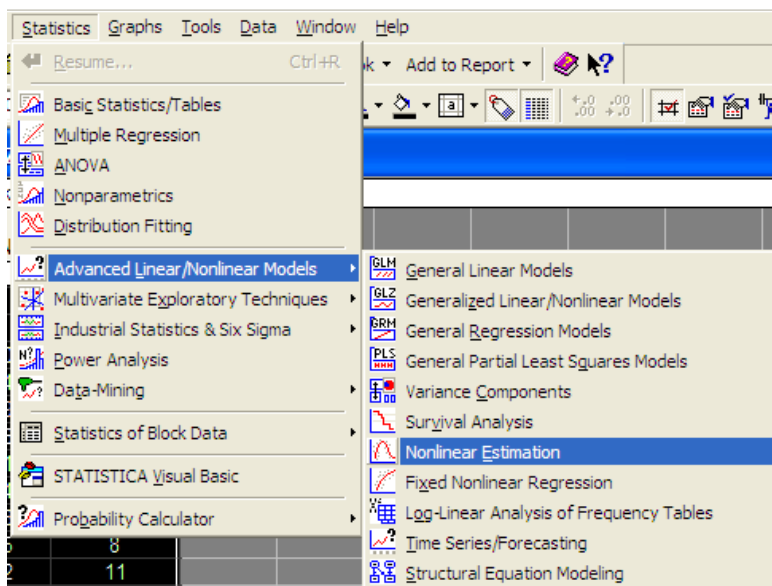
- funkcja potęgowa $Y=a \cdot X^b$;
- funkcja wykładnicza $Y=a \cdot b^X$;
- funkcja liniowa $Y=a \cdot X+b$

MODELE NIELINIOWE

Nie wszystkie modele możemy sprowadzić do modelu liniowego. W takim przypadku sami musimy zdefiniować funkcję estymowaną oraz funkcję straty. Dalej będziemy rozważać modele regresji postaci $y=F(x_1, x_2, x_3, \dots, x_k)$ gdzie F jest dowolną funkcją. Funkcja straty ma za zadanie szacować wielkość odchylenia wartości przewidywanych (PRED) od wartości obserwowanych (OBS). Minimalizacja funkcji straty jest procedurą estymacji współczynników równania regresji. Postać funkcji straty zależy od metody szacowania parametrów. Najczęściej wykorzystywana jest metoda najmniejszych kwadratów. W pakiecie Statistica definiujemy ją jako $L=(OBS-PRED)^2$

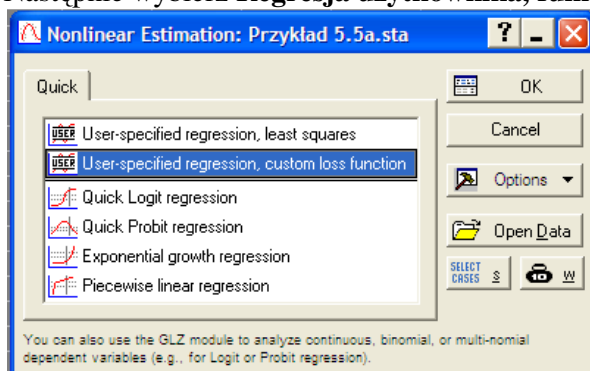
Zadanie 8 (regresja nieliniowa)

Na podstawie danych zebranych dla 15 pacjentów zależność dni spędzonych w szpitalu, a szansą powrotu do zdrowia opisano zależnością eksponentialną $PROGNOZA=a_1 \cdot \exp(a_2 \cdot DNI)$. Do oceny parametrów modelu wykorzystaj moduł Estymacja nieliniowa z opcji **Zaawansowane modele liniowe i nieliniowe** (rys. 4) .



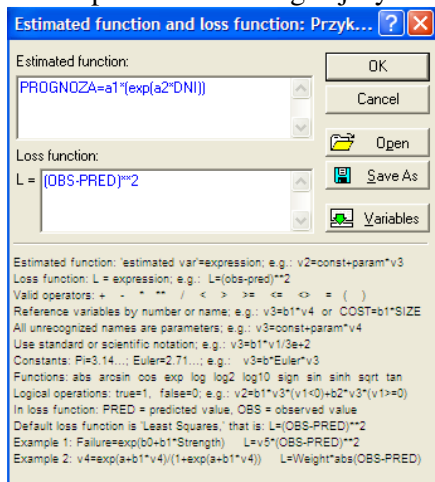
Rys.4 Okno wyboru estymacji nieliniowej

Następnie wybierz **Regresja użytkownika, funkcja starty użytkownika** (rys5)



Rys.5 Okno wyboru modelu regresji

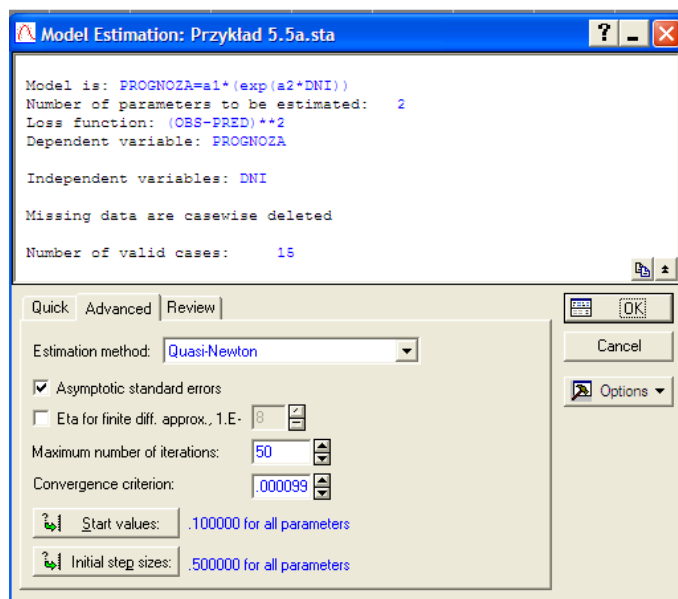
Oraz wpisz równanie regresji rys.6



Rys.6 Okno specyfikacji funkcji użytkownika i funkcji straty

Dane do zadań zaczerpnięto z książki Adama Stanisza Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny

W celu odnalezienia minimum funkcji starty wybierz **Metodę estymacji** quasi-Netwona oraz zaznacz opcje **Asymptotyczne błędy standardowe** (rys7), uzyskane błędy standardowe będą pomocne przy ocenie istotności otrzymanych ocen.



Rys.7 Okno wyboru procedury estymacji

Na podstawie uzyskanych wyników odpowiedź na następujące pytania i wykonaj poniższe polecenia:

- ile wynoszą oceny a_1 i a_2 ? Zapisz otrzymane równanie regresji
- ile wynoszą błędy standardowe uzyskanych ocen a_1 i a_2 ?
- czy parametry modelu są istotne (p dla t-Studenta)?
- ile % zmienności wskaźnika prognozy może być wyjaśnione na podstawie liczby dni spędzonych w szpitalu (inaczej, jaka część zmiennej zależnej jest wyjaśniona przez model)?
- na podstawie analizy **Obserwowane, przewidywane, reszty** oraz wykresu **Dopasow. Funkcja 2W i wart. Obserw.**, oceń czy wystąpiły obserwacje odstające i dopasowanie modelu (wzrokowo)
- sprawdź czy reszty modelu mają rozkład normalny (sporządź **Wykres normalności reszt**)

Zadanie 9

Opisano zależność między poziomem ciśnienia skurczowego krwi od stężenia glukozy i aldosteronu za pomocą modelu:

$$\text{Ciśn} = a_1 * \text{Glukoza}^{a_2} * \text{Aldo}^{a_3}$$

Sprawdź dopasowanie modelu.