

Prosty model regresji w prognozowaniu:

Cel analizy regresji:

Modelowanie związku pomiędzy zmienną zależną Y_t (zmn. objaśnianą) a zmiennymi niezależnymi $\{x_1, \dots, x_n\}$ (zmn. objaśniającymi)

Funkcyjna postać modelu liniowego:

$Y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n + \varepsilon$; gdzie

- b_0 – stała,
- b_1, \dots, b_n parametry równania (częstkowe współczynniki regresji)
- ε - składnik losowy

Interpretacja:

jeżeli wartość zmiennej X_i zwiększy się o 1 jednostkę, wtedy wartość Y zmieni się o wartość i -tego współczynnika korelacji częściowej, przy założeniu, że pozostałe zmienne niezależne pozostaną bez zmian.

Uwzględnienie w modelu regresji zmiennej czasowej, umożliwia zastosowanie regresji liniowej w prognozowaniu szeregów czasowych. W tym celu należy dodać zmienną czasową X , która przyjmuje wartości kolejnych liczb całkowitych $X \in \{1, \dots, n\}$, które odzwierciedlają pozycję danej obserwacji w analizowanym szeregu czasowym.

Warianty modelu regresji:

Możliwe są również różne warianty modelu regresji. W tym celu do ogólnego modelu opisanego powyżej należy dodać dodatkowe odpowiednio przekształcone zmienne objaśniające X , np.

Model podstawowy liniowy: $Y = b_0 + b_1X + \varepsilon$

Model kwadratowy: polega na dodaniu 2 zmiennych czasowych X i X^2 : $Y = b_0 + b_1X + b_2X^2 + \varepsilon$

Model wielomianowy stopnia n : polega na dodaniu n zmiennych czasowych X, X^2, X^3, \dots, X^n :

$Y = b_0 + b_1X + b_2X^2 + b_3X^3 + \dots + b_nX^n + \varepsilon$

Model hiperboliczny: polega na dodaniu zmiennej $1/X$: $Y = b_0 + b_1(1/X) + \varepsilon$


Ogólny schemat postępowania:

1. Dodanie zmiennej pomocniczej i nazwanie jej X (w prostym modelu liniowym) i wypełnienie jej kolejnymi wartościami od 1 do n . Możliwych jest wiele możliwości jak wypełnić ten punkt, np.
 - a. Wykorzystanie numerów przypadków i przypisanie zmiennej X funkcji $=V0$
 - b. Wpisanie w pierwszych dwóch komórkach zmiennej X wartości 1 i 2, a następnie po zaznaczeniu obu przeciągnąć zaznaczenie do ostatniej komórki w badanym zakresie
2. W przypadku innych wariantów regresji, należy postąpić tak jak w punkcie nr 1 oraz dodatkowo wstawić kolejne zmienne pomocnicze (ilość i rodzaj zależy od wybranego modelu) i przypisać im za pomocą funkcji wartości wynikające z odpowiednich przekształceń, na przykład: w modelu kwadratowym dodajemy zmienną X (funkcja: $=V0$) oraz X^2 (funkcja: $=X^2$), itd.
3. Na pasku narzędzi należy odszukać: Statystyka i wywołać okno *Regresja wieloraka*

4. Jako zmienną zależną, wybieramy główną zmienną, która ma być poddana analizie. Jako zmienną niezależną należy wybrać zmienną pomocniczą X (w przypadku innych modeli niż podstawowy, oprócz zmiennej X należy wybrać także pozostałe zmienne wynikające z przekształcenia zmiennej X).

5. Po zaakceptowaniu wybór odpowiednich zmiennych, kończący się wywołaniem okna z wynikami analizy regresji, należy przeanalizować podstawowe miary dopasowania regresji takie jak:

R² – miara dopasowania modelu do danych. Im bliżej 1 tym lepsze dopasowanie. Wartość podawana jest w %.

 Wyniki regresji wielorakiej: Energia odnawialna.sta

Wyniki regresji wielorakiej		
Zmn. zależ. Spain	Wielor. R = ,98425830	F = 341,1624
	R ² = ,96876441	df = 1,11
Liczba przyp. 13	Popraw. R ² = ,96592481	p = ,000000
	Błąd standardowy estymacji: ,575884847	
Wyr. wolny 7,234615385	Błąd std.: ,3388209	t(11) = 21,352 p = ,0000
	X b* = ,984	

p- prawdopodobieństwo komputerowe interpretowane w odniesieniu dopasowania modelu do danych.

6. W zakładce Podstawowe wywołujemy okno Podsumowanie: wyniki regresji. W tym oknie istotne są dwie kolumny: **b** oraz **p**. W kolumnie b zawarte są informacje o wartościach współczynników regresji, natomiast kolumna p informuje o statystycznej istotności poszczególnych współczynników. Współczynnik jest statystycznie istotny jeśli wartość p jest mniejsza od założonego poziomu istotności. Zwyczajowo poziom istotności wyznacza się na poziomie 5%. Dla ułatwienia wartości statystycznie istotne (o ile nic nie zmieniono w ustawieniach programu) zaznaczane są na czerwono.

Podsumowanie regresji zmiennej zależnej: Spain (Energia odnawialna.sta)						
R= ,98425830 R ² = ,96876441 Popraw. R ² = ,96592481						
F(1,11)=341,16 p<,00000 Błąd std. estymacji: ,57588						
	b*	Bł. std. z b*	b	Bł. std. z b	t(11)	p
N=13						
W. wolny			7,234615	0,338821	21,35233	0,000000
X	0,984258	0,053288	0,788462	0,042687	18,47058	0,000000

W przypadku gdy któraś zmienna miałaby współczynnik statystycznie nieistotny ($p > 0,05$), należałoby taką zmienną usunąć z analizy i wykonać analizę regresji od nowa.

7. W celu wykonania prognozy należy przywołać okno edycji (ctr+R) i przejść do zakładki Reszty, założenia, predykcja. Wybierając odpowiednią wartość alfa, można wyznaczyć przedział ufności dla wartości prognozowanej. Alfa=0.05, oznacza wyznaczenie 95% PU. Alfa=0.1 oznacza wyznaczenie 90% PU. Następnie należy wybrać przycisk Predykcja zmiennej zależnej.

Podstawowe | Więcej | Reszty, założenia, predykcja

Wykonaj analizę reszt

Statystyki opisowe

Generator kodów

Wartości przewidywane:

Predykcja zmiennej zależnej

Oblicz granice ufności Alfa: 0,05

Oblicz granice predykcji

W oknie, które się pojawi, jako wartość X należy wpisać numer zmiennej X dla prognozowanej wartości, na przykład: jeśli analizowany szereg miał 100 obserwacji, to do prognozy na kolejny okres będzie przypisana wartość 101 i tą należy wpisać w oknie jako wartość X. W oknie wynikowym wartość prognozowaną (prognozę punktową) odczytuje się z pozycji: Przewidyw., natomiast przynależny przedział ufności (prognozę przedziałową) z wierszy: $\pm 95,0\%GU$.

Zmienna	Wagi b	Wartość	Wagi b *Wartość
X	0,788462	14,00000	11,03846
W. wolny			7,23462
Przewidyw.			18,27308
-95,0%GU			17,52734
+95,0%GU			19,01882

Zadanie 1. Plik: energia odnawialna.sta

Celem jest wykonanie prognozy punktowej oraz 90% prognozy przedziałowej ilości energii odnawialnej w Szwecji na lata 2017-2020 za pomocą modelu regresji prostej i modelu kwadratowego. Dodatkowo należy podać i zinterpretować wartości R^2 i p modelu.

Rok	Model liniowy		Model kwadratowy	
	prognoza punktowa	prognoza przedziałowa	prognoza punktowa	prognoza przedziałowa
2017				
2018				
2019				
2020				
R^2				
p				

Zadanie 2. Plik: energia odnawialna.sta

Celem jest wykonanie prognozy punktowej oraz 95% prognozy przedziałowej ilości energii odnawialnej w Polsce na lata 2017-2020 za pomocą modelu wielomianowego 3 stopnia. Dodatkowo należy podać i zinterpretować wartości R^2 i p modelu.

Rok	Wielomian trzeciego stopnia	
	prognoza punktowa	prognoza przedziałowa
2017		
2018		
2019		
2020		
R^2		
p		