

**Wojciech HYB, Joanna KALETA**

Katedra Zastosowań Matematyki  
Department of Applied Mathematics

## **Porównanie metod wyznaczania współczynników modelu matematycznego na przykładzie prognozy liczby ludności świata**

### **Comparison of methods of determination of mathematical model coefficients on the example of world population projection**

**Słowa kluczowe:** ludność świata, model matematyczny, prognoza

**Key words:** world population, mathematical model, projection

#### **Wprowadzenie**

Praca dotyczy zagadnienia prognozy liczby ludności świata, a jej celem jest zbadanie wpływu zmiany metody wyznaczania parametrów założonego modelu na prognozę. Dzięki zastosowaniu innej miary dopasowania (Hyb 2003) możliwe będzie uzyskanie znacznie mniejszych błędów predykcji.

Już od czasów Eulera próbowano dobrać idealny model demograficzny pasujący do tendencji cechującej zmianę liczby ludności w ustalonym przedziale czasowym z uwzględnieniem prognozy. Stosowano w tym celu dwa modele: deterministyczny i z trendem wzrostowym. Jednak obydwa dawały

rozbieżne prognozy, a wyniki odbiegały od rzeczywistości. W połowie XIX wieku belgijski matematyk F.F. Verhulst zaproponował krzywą logistyczną, zakładającą zmienność współczynnika wskaźnika urodzeń jako model dopasowujący się do wzrostu populacji ludzkiej zarówno na całym świecie, jak i w poszczególnych państwach.

Ze względu na trudności w otrzymaniu właściwych estymatorów parametrów tej krzywej i niepewność podawanych przez nią prognoz, szczególnie długoterminowych, okazało się, że wspomniana krzywa także nie stanowi miarodajnego modelu. Nadal jednak stosowano w demografii krzywą logistyczną, zwłaszcza że w 1985 roku opracowano metodę tzw. zadowalającej estymacji parametrów tej funkcji (Smolik 1985).

Poszukiwano również innych modeli demograficznych, a w 1996 roku S. Smolik zaproponował model postaci

$$\hat{y}_t = f(t) = a \cdot \arctg(b \cdot e^{ct}) \quad (1)$$

gdzie:

$\hat{y}_t$  – liczba ludności świata (w milionach) obliczona wzorem (1) dla roku o numerze  $t$ ,  $a$ ,  $b$ ,  $c$  – współczynniki modelu, nazywane także parametrami strukturalnymi (Smolik 1996).

Za najlepsze estymatory nieznanymi parametrów  $a$ ,  $b$ ,  $c$  dla modelu (1) przyjęto w pracy Smolika (1996) te wartości, dla których suma kwadratów odchyleń:

$$S(a, b, c) = \sum_{t=1}^n [y_t - a \cdot \arctg(b \cdot e^{ct})]^2 \quad (2)$$

osiąga wartość najmniejszą ( $y_t$  – rzeczywista liczba ludności świata w roku o numerze  $t$ ). Prowadzi to do klasycznej metody wyznaczania funkcji regresji, którą jest maksymalizacja współczynnika korelacji krzywoliniowej (Elandt 1964).

Praca Smolika (1996) jest niezwykle ciekawa z uwagi na trafność w doborze modelu opisującego numeryczny wzrost populacji świata w latach 1950–1985 wraz z prognozą od 1986 do 1994 roku (Smolik 1996, tab. 1), a także ze względu na przeprowadzony przez autora sposób estymacji parametrów funkcji regresji (1). Zaproponowana przez Smolika funkcja regresji ma postać:

$$\hat{y}_t = 19\,000 \cdot \arctg(0,127 e^{0,02t}) \quad (3)$$

gdzie  $t$  oznacza numer danego roku (przyjęto  $t = 1$  dla roku 1950).

### Zmiana miary dopasowania a prognoza liczby ludności świata

W prezentowanej pracy nadal rozważany jest model określony wzorem (1), ale jego współczynniki  $a$ ,  $b$ ,  $c$  postanowiono wyznaczyć, minimalizując procentowy średni błąd względny  $MRE \cdot 100\%$  (mean relative error) określony wzorem (Hyb 2003):

$$\begin{aligned} MRE \cdot 100\% &= \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{y_t} \cdot 100\% = \\ &= \frac{1}{n} \sum_{t=1}^n \frac{|y_t - a \cdot \arctg(b \cdot e^{ct})|}{y_t} \cdot 100\% \end{aligned} \quad (4)$$

Minimalizacja zależności (4) ze względu na parametry  $a$ ,  $b$ ,  $c$  jest możliwa tylko w sposób numeryczny. Jednakże stosowanie nowoczesnych metod i technik obliczeniowych pozwala zlokalizować dokładniej obszar poszukiwania parametrów, a następnie obliczyć ich przybliżone wartości, stosując np. metodę Newtona-Marquardta.

Aby porównać otrzymane wyniki, postanowiono wyznaczyć wartości współczynników  $a$ ,  $b$ ,  $c$ , używając danych z lat 1950–1985 i przyjmując za  $t = 1$  rok 1950 oraz za  $t = 36$  rok 1985, analogicznie jak w pracy Smolika (1996). Minimalizując procentowy średni błąd względny (4) uzyskano funkcję regresji w postaci:

$$\hat{z}_t = 9500 \cdot \arctg(0,254 e^{0,022t}) \quad (5)$$

Otrzymane wyniki dla lat 1956–1985 zestawiono w tabeli 1, w której  $y_t$  oznacza rzeczywistą liczbę ludności świata,  $\hat{y}_t$  jest obliczone wzorem (3),  $\hat{z}_t$  jest obliczone wzorem (5), natomiast  $RE_t$  oznacza błąd względny w procentach (relative terror), obliczony odpowiednio dla wartości  $\hat{y}_t$  i  $\hat{z}_t$ .

Następnie przedstawiono prognozę liczby ludności świata na lata 1986–2000 dla funkcji regresji (3) i (5) oraz sprawdzono dokładność tych prognoz za pomocą błędów względnych, dysponując danymi ( $y_t$ ) uzyskanymi z ONZ (tab. 2) i przyjmując oznaczenia jak w tabeli 1.

## Podsumowanie i wnioski

Analiza wyników przedstawionych w tabeli 1 prowadzi do wniosku, że funkcje regresji określone wzorami (3) i (5) dają dobre przybliżenie rzeczywistej liczby ludności świata. Średni błąd względny estymacji dla funkcji (5) jest nieco mniejszy niż dla funkcji (3), co ma związek z inną metodą wyznaczania współczynników.

Natomiast w przypadku prognozy funkcja regresji (5), której współczynniki wyznaczone za pomocą miary dopasowania (4), gwarantuje większą dokładność niż funkcja określona wzorem (3). Miarą tej dokładności są znacznie mniejsze wartości błędów względnych w kolumnie  $RE_t$  dla wzoru (5) niż dla wzoru (3) (tab. 2). Ponadto można stwierdzić, że prognoza na lata 1986–1993 otrzymana przy zastosowaniu funkcji (5) jest dokładna. Świadczą

o tym bardzo małe wartości błędów względnych. Największy błąd względny otrzymany w 2000 roku dla funkcji (5) wynosi 2,37%, a dla funkcji (3) jest równy 5,1%. Ponieważ od 1996 roku prognoza otrzymana przy zastosowaniu funkcji (3) daje błąd względny ponad 3%, więc należy uznać ją za niewystarczającą.

Dodatkowo zamieszczono w pracy tabelę 3, z której wynika, że prognozy liczby ludności świata wyznaczone na kolejne lata funkcjami regresji (3) i (5) różnią się istotnie. Ponadto można stwierdzić, że asymptotyczna liczba ludności świata – asymptota pionowa funkcji postaci (1) – wynosi 14922,6 mln dla funkcji (5) i 29845,1 mln dla funkcji (3), a największy przyrost roczny liczby ludności świata wyniesie 103,9 mln w 2012 roku dla funkcji (5) i 190,0 mln w roku 2053 dla funkcji (3).

Otrzymane wyniki sugerują, że metoda wyznaczania współczynników funkcji regresji za pomocą miary dopasowania (4) może dawać znacznie lepsze prognozy liczby ludności świata, zwłaszcza w długim okresie czasowym. Warto zatem poszukiwać współczynników modelu matematycznego, posługując się nie tylko metodami klasycznymi.

## Literatura

- ELANDT R. 1964: Statystyka matematyczna w zastosowaniu do doświadczeń rolniczego. PWN, Warszawa.
- HYB M. 2003: Analiza statystyczna wybranych charakterystyk konsolidacyjnych gruntów organicznych. *Przegląd Naukowy Inżynieria i Kształtowanie Środowiska* 2 (XVII): 56–65.
- SMOLIK S. 1985: Wyznaczanie parametrów krzywej logistycznej. *Przegląd Statystyczny*, XXXII, 4: 365–373.

SMOLIK S. 1996: Long-term Projection of Numerical Growth of Population. II Krajowa Konferencja Zastosowań Matematyki w Biologii i Medycynie.

## Summary

**Comparison of methods of determination of mathematical models coefficients on the example of world population projection.** Analysis of the results obtained in this paper indicates, in some sense, predominance of the method determining regression function coefficients (fit measure (4)) over the method using fit measure (2). Thanks to fit measure (4) one can obtain more exact projections, particularly long-term projections.

### Authors' address:

Wojciech Hyb, Joanna Kaleta  
Szkoła Główna Gospodarstwa Wiejskiego  
Katedra Zastosowań Matematyki  
02-787 Warszawa, ul. Nowoursynowska 159  
Poland  
e-mail: hyb@alpha.sggw.waw.pl  
kaleta@alpha.sggw.waw.pl

### **Streszczenie**

Wojciech Hyb, Joanna Kaleta  
Katedra Zastosowań Matematyki  
Warsaw Agricultural University –  
SGGW

02-787 Warszawa, ul. Nowoursy-  
nowska 166  
e-mail: [hyb@alpha.sggw.waw.pl](mailto:hyb@alpha.sggw.waw.pl)  
[kaleta@alpha.sggw.waw.pl](mailto:kaleta@alpha.sggw.waw.pl)

**Porównanie metod wyznaczania współczynników modelu matematycznego na przy-  
kładzie prognozy liczby ludności świata**

W pracy wyznaczono współczynniki pewnego modelu matematycznego opisującego wzrost populacji ludzkiej metodą minimalizacji średniego błędu względnego. Otrzymane wyniki, szczególnie dla prognozy, okazały się znacznie lepsze niż uzyskane wcześniej przez dla analogicznego modelu, którego współczynniki wyznaczono klasyczną metodą najmniejszych kwadratów. Przedstawiono także hipotetyczny rozwój ludności świata na lata 2005-2100. W konkluzji stwierdzono, że należy poszukiwać nowych metod wyznaczania współczynników funkcji regresji.