

Prognozowanie i symulacje (lab. 9)

Wyrównywanie wykładnicze

Wstęp

Modele wyrównywania wykładniczego oparte są koncepcji, iż na aktualną prognozę wpływ mają wszystkie poprzednie wartości szeregu czasowego, lecz wpływ ten jest coraz mniejszy dla danych coraz bardziej odległych w czasie. Szczegółowe wzory modeli wyrównywania wykładniczego są dość skomplikowane i nie będą tutaj omawiane.

W programie *STATISTICA* zaimplementowano 12 rodzajów modeli wyrównywania wykładniczego, które stanowią kombinacje modeli: **z trendem i bez trendu**, oraz **z wahaniami i bez wahań sezonowych**.

Poszczególne etapy prognozowania za pomocą modelu wykładniczego obejmują:

- 1) analizę graficzną zmienności szeregu czasowego i określenie rodzaju trendu (oraz wahań sezonowych);
- 2) poszukiwanie optymalnego modelu określonego typu „pasującego” najlepiej do danych – wyznaczanie miary błędu dopasowania modelu do danych;
- 3) wyznaczanie wartości prognozy oraz ich wizualizację graficzną.

W programie *STATISTICA* wyznacza się kilka rodzajów błędów dla modeli wyrównywania wykładniczego – są one opisane poniżej. Ponieważ różne rodzaje błędów mogą (choć nie muszą) sugerować wybór odmiennych modeli na naszych zajęciach będziemy opierać się na „ostatniej” mierze – **średnim bezwzględnym błędzie procentowym**.

- a) **Błąd średni** – jest to średnia arytmetyczna wartości reszt (błęd), przy czym znaczenie tej miary jest ograniczone przez fakt, że dodatnie i ujemne wartości błędu znoszą się nawzajem.
- b) **Średni błąd bezwzględny** – jest to dużo lepszy miernik dopasowania modelu, gdyż obliczany jest jako średnia wartości bezwzględnych reszt. Wartość średniego błędu bezwzględnego równa zero oznacza, że model jest idealnie dopasowany do danych.
- c) **Średni błąd kwadratowy** – charakteryzuje się podobnymi własnościami jak średni błąd bezwzględny, ponadto jest bardziej wrażliwy na wartości odstające.
- d) **Średni błąd procentowy** – błąd procentowy dla danego okresu czasowego wyliczany jest jako iloraz reszty (błędu) i wartości obserwowanej wyrażony w procentach. Średnia arytmetyczna procentowych błędów prognozy jest często bardziej przydatna niż błąd średni, gdyż pokazuje względne niedopasowanie modelu. Wspólna jest wada obu tych miar czyli znoszenie się nawzajem odchyleń o różnych znakach.
- e) **Średni bezwzględny błąd procentowy** – najdoskonalsza ze wszystkich wymienionych charakterystyk, jest łatwo interpretowalna i kumuluje w sobie zalety wszystkich wyszczególnionych wcześniej miar.

Przykład 1 (Dane: Transport w Polsce 1990-2023 (R))

UWAGA: Przed rozpoczęciem analizy proszę usunąć dane z lat 2020-2023.

Analizie będzie podlegać *Liczba pasażerów przewożonych koleją* – celem analizy będzie uzupełnienie prognoz stworzonych za pomocą modeli tendencji rozwojowych (por. zajęcia nr 3 i 4) wynikami modeli wyrównywania wykładniczego.

Wszystkie analizy będą prowadzone w module *SZEREGI CZASOWE*.

Za pomocą polecenia *Statystyka / Zaawansowane modele liniowe i nieliniowe / Szeregi czasowe* wywołujemy okno *Analizy szeregów czasowych*, gdzie wybieramy odpowiednią zmienną do analizy. Następnie spośród różnych metod prognozowania wybieramy *Wyrównywanie wykładnicze i prognozowanie*.

Wstępna analiza graficzna

Przełączamy się do zakładki *Przegląd szeregu* gdzie istnieje możliwość tabelarycznej i graficznej prezentacji analizowanego szeregu czasowego. Aby wykres był czytelny, określamy sposób etykietowania obserwacji (*Oznacz punkty*) wybierając odpowiednią opcję – w naszym przypadku [*Oznacz punkty*] *nazwami przypadków*. Za pomocą przycisku *Kreśl* wywołujemy wykres zmienności liczby pasażerów przewożonych koleją. Pobieźna analiza wzrokowa każe stwierdzić, iż badane zjawisko na pewno podlega wyraźnym trendom, jak również, że nie występuje w nim efekt sezonowości (o który zresztą raczej trudno przy danych rocznych).

Dobór parametrów modelu z trendem liniowym

Przełączamy się do zakładki *Więcej* i wybieramy *Model z trendem liniowym (Holta)* – modele tego typu są opisane przez dwa parametry: *alfa* i *gamma*.

Ustalamy też zakres prognozy na *Prognozuj 3 obserwacje* (będą to lata 2020-2022).

Następnie **ODZNACZAMY** domyślnie zaznaczoną opcję ☒ *Dodaj prognozy i błędy do obszaru rob.*, dzięki czemu unikniemy problemu z wyborem zmiennej przy wyznaczaniu kolejnych modeli.

Aby dobrać parametry najlepiej opisujące rozważane dane postępujemy w następujący sposób:

- 1) W zakładce *POSZUKIWANIE SIECIOWE* naciskamy przycisk *Wykonaj poszukiwanie sieciowe* (program bada pewne kombinacje parametrów modelu: α i γ). Szukamy takiej kombinacji parametrów, które minimalizują **Średni bezwzględny błąd procentowy** (dla tego modelu powinna to być kombinacja: **$\alpha = 0,9$ i $\gamma = 0,8$**).

Prognozowanie i symulacje (lab. 9)

Wyrównywanie wykładnicze

UWAGA! Wartości parametrów *alfa* i *gamma* (lub innych specyficznych dla danego modelu) mogą być dla kolejnych modeli zupełnie inne – **każdorazowo wykonujemy od początku Poszukiwanie sieciowe**.

- 2) Znalezione za pomocą poszukiwania sieciowego parametry, które są wyliczone z dokładnością do 0,1, wprowadzamy jako **Wstępne wartości parametrów** w zakładce **AUTOMATYCZNE POSZUKIWANIE**. Ustawiamy **Wskaźnik braku dopasowania** na **Średni bezwzględny błąd procentowy** i wykonujemy **Estymację automatyczną**.
- 3) W skrócie wyników pojawiają się wyniki analizy w postaci: **dwóch nowych arkuszy i wykresu**. W arkuszach znaleźć można wartość prognozy i ostateczne miary błędów zaś na wykresie wizualizację oryginalnych wartości i prognoz. Odpowiednie wyniki dla modelu liniowego proszę umieścić w tabeli poniżej – część wyników już wpisano dla sprawdzenia poprawności obliczeń.

Proszę w analogiczny sposób, uzyskać prognozy (wraz z miarą błędu) dla **wyrównywania wykładniczego z trendem wykładniczym i gasnącym**. Wznawiamy w tym celu analizę (Ctrl + R) i jeżeli jest to kontynuacja wcześniejszych obliczeń wystarczy, że zmienimy w zakładce **Więcej** rodzaj modelu i powtórzmy czynności z p. 1-3. Pozostałe ustawienia powinny być zapamiętane. Za każdym razem trzeba od początku powtarzać proces **sieciowego i automatycznego poszukiwania parametrów**. Wyniki proszę zamieścić w poniższej tabeli.

Tab. 1. Prognozy wykonane za pomocą modeli wyrównywania wykładniczego z trendem dla liczby przewozów pasażerskich kolejną w latach 2020-2022

Model z trendem	Średni bezwzględny błąd procentowy	Prognozowane LICZBY PASAŻERÓW (w mln)		
		2020	2021	2022
liniowym	4,41%	352,2	369,1
wykładniczym
gasnącym

Na podstawie poziomu błędu proszę wskazać, która prognoza powinna zostać uznana za najlepszą (ze statystycznego punktu widzenia):

Proszę znaleźć faktyczną wielkość (Y) przewozów kolejną w latach 2020-22 (utk.gov.pl/pl/raporty-i-analizy/analizy-i-monitoring) i porównać z wyznaczoną prognozą (Y_p).

Proszę wyznaczyć błąd procentowy prognozy za pomocą wzoru: $Błąd\ procentowy = (Y - Y_p) / Y * 100\%$

Przykład 2 (Dane: Wypadki w Polsce od 1998 roku (M))

Stosując sześć rodzajów modeli wyrównywania wykładniczego z różnymi rodzajami trendów i wahaniami sezonowymi, proszę dokonać prognoz **Liczby wypadków na wszystkie miesiące do końca 2026 roku**.

Ponieważ będziemy prognozować 24 wartości (2 lata × 12 miesięcy), dobrze byłoby, aby były one opisane w arkuszu wyników. W tym celu w arkuszu danych dodajemy (na końcu) 24 przypadki i wypełniamy je wartościami zmiennej **Data**, przeciągając dwie ostatnie obserwacje. W zakładce **Przegląd szeregu** nie zapominamy o ustaleniu etykietowania przypadków za pomocą zmiennej **Data**. Dalej postępujemy dokładnie według punktów opisanych w przykładzie 1. Na podstawie uzyskanych wyników proszę uzupełnić poniższą tabelę.

Model z trendem	Wahania sezonowe	Średni bezwzględny błąd procentowy	Prognozowana LICZBY WYPADKÓW				
			styczeń 2025	luty 2025	październik 2025	maj 2026	październik 2026
liniowym	addytywne	%					
wykładniczym		%					
gasnącym		%					
liniowym	multiplikatywne	%					
wykładniczym		%					
gasnącym		%					

Na podstawie poziomu błędu proszę wskazać, która prognoza powinna zostać uznana za najlepszą (ze statystycznego punktu widzenia):

Na stronie <https://obserwatoriumbrd.pl/statystyki/>, po wybraniu polecenia **NAJNOWSZE DANE**, można znaleźć dane o wypadkach za część miesięcy z 2025 r. Proszę znaleźć faktyczną liczbę wypadków w styczniu i lutym 2025 – który model dał wyniki najbliższe prawdziwym wartościom dla stycznia, a który dla lutego

CIEKAWOSTKA: Modele wyrównywania wykładniczego są uważane za jedno z najskuteczniejszych narzędzi prognozowania zjawisk ekonomicznych (i nie tylko). Warto wiedzieć, że idea wyrównywania wykładniczego powstała podczas II wojny światowej, a metodę tę wykorzystywała amerykańska marynarka wojenna w „Bitwie o Atlantyk” do prognozowania położenia niemieckich okrętów podwodnych podczas ścigania ich przez niszczyciele.