

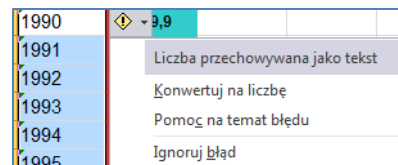
Prognozowanie i symulacje (lab. 12)

Prognoza naiwna, średnie ruchome i błędy prognozy

Przykład 1 (wyznaczanie trendów w arkuszu Excela)

W tym przykładzie sprawdzimy, jakie są możliwości graficznego dopasowania linii trendu w programie Excel. Będzie to uzupełnienie informacji z zajęć nr 4, gdzie wykorzystywaliśmy w tym celu program STATISTICA. Proszę skopiować do arkusza programu Excel dane **WYŁACZNIE z lat 1990-2019** z pliku *Transport w Polsce 1990-2024 (R)* dotyczące **Przewozów pasażerów koleją**. W tym celu zaznaczamy odpowiednią kolumnę i wykorzystujemy polecenie *Edycja / Kopiuj z nagłówkami*. Po wklejeniu usuwamy pierwszy wiersz.

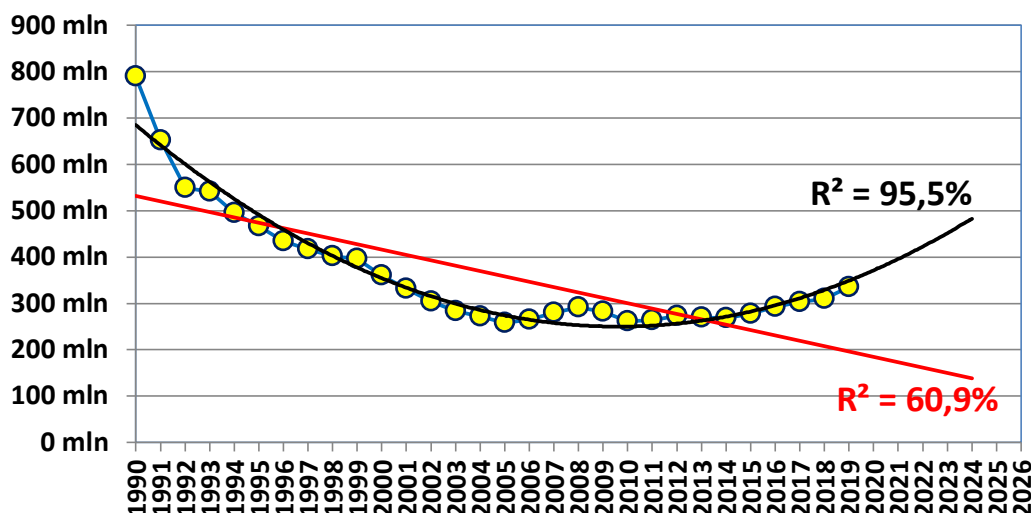
Uwaga! Po wklejeniu danych do Excela lata, które były w nazwach przypadków arkusza programu STATISTICA traktowane są czasem jako tekst – jeżeli tak jest proszę zaznaczyć komórki z latami i dokonać konwersji na postać liczbową (wskazówki na rysunku obok).



Następnie zaznaczamy zakres danych (lata i dane o liczbie pasażerów kolei) i wykonujemy wykres za pomocą poleceń *Wstawianie / Punktowy*. W zakładce *Układ / Linia trendu / Więcej opcji linii trendu* wskazujemy *Trend liniowy* oraz *Wyświetl wartość R kwadrat na wykresie*. A także określamy *Prognozę „do przodu”* na 5 okresów. Następnie w podobny sposób dodajemy linię *trendu wielomianowego stopnia 2* (także z wartością R^2). Proszę sformatować wykres tak, aby wyglądał jak poniższy wzorzec.

Uwaga! W różnych wersjach programu Excel wstawianie linii trendu może odbywać się na odmienne sposoby. Można np. wywołać menu podręczne klikając prawym przyciskiem myszki punkty na wykresie.

Przewozy pasażerów (koleje) [mln osób]



Proszę dwukrotnie kliknąć wielomianową linię – modyfikując stopień wielomianu proszę obserwować jak zmienia się R^2 dla wielomianów coraz wyższych stopni. A także jak zmienia się kształt dopasowanej funkcji, a także kierunek i wartości prognoz. Po uzupełnieniu wartości R^2 stwierdzić można, iż wraz ze wzrostem złożoności funkcji pasuje ona coraz lepiej do danych – czy jednak prognoza otrzymana za pomocą wielomianu 6-go stopnia będzie wiarygodniejsza od prognozy otrzymanej za pomocą modelu kwadratowego?

Stopień dopasowanego wielomianu	Współczynnik determinacji R^2
2	
3	
4	
5	
6	

Niewielki przyrost R^2 dla modeli wielomianowych 3-go stopnia i wyższych (o ok. 1%) nie jest na tyle duży, by uzasadnić stosowanie tak złożonych modeli. **Proszę zwrócić uwagę jak „chaotycznie” zachowują się prognozy dla wielomianów stopnia 3-go i wyższego.**

Przykład 2 (metoda naiwna)

Najprostszym „pasywnym” podejściem do prognozowania kolejnych realizacji szeregu czasowego jest tzw. **metoda naiwna**. Jej idea sprowadza się do przyjęcia, iż kolejna wartość prognozowana będzie równa ostatniej zaobserwowanej. Wyznaczenie takiej prognozy nie wymaga żadnych obliczeń, natomiast można i warto sprawdzić jej skuteczność (wyznaczyć błąd prognozy), ponieważ może stanowić ona punkt odniesienia dla innych, potencjalnie lepszych metod.

Poniższy przykład oparty jest na danych z lat 1990-2019 z pliku *Transport w Polsce 1990-2024 (R)*. Obliczenia zostaną wykonane za pomocą arkusza kalkulacyjnego Excel.

Prognozowanie i symulacje (lab. 12)

Prognoza naiwna, średnie ruchome i błędy prognozy

- 1) Proszę skopiować dane z lat **1990-2019** dotyczące *Przewozów pasażerów koleją* do arkusza programu *Excel* (dane powinny znaleźć się w kolumnie B – pierwsza wartość w komórce B2, w kolumnie A powinny znaleźć się informacje o latach, z których pochodzą dane).
- 2) W kolumnie C proszę wyznaczyć prognozę metodą naiwną – nie tylko na 2020 rok, ale także tzw. prognozy wygasłe dla lat wcześniejszych, które posłużą do policzenia błędu prognozy (w komórce C3 wpisujemy formułę =B2 i przeciągamy aż do roku 2020).
- 3) W kolumnie D proszę wyznaczyć (dla każdej obserwacji z lat 1991-2019) bezwzględny błąd procentowy prognozy (ang. *APE* – *absolute percentage error*) według wzoru:

$$APE = \frac{|y_t - \hat{y}_t|}{y_t} \cdot 100\% \quad (\hat{y}_t \text{ oznacza prognozę, } y_t \text{ – to wartość rzeczywista})$$

- 4) Proszę wyznaczyć bezwzględny procentowy błąd prognoz dla każdego roku z przedziału 1991-2019, a następnie wartość średnią, czyli średni bezwzględny błąd procentowy (ang. *MAPE* – *mean absolute percentage error*) dla prognozowanej liczby pasażerów przewożonych koleją za pomocą modelu naiwnego.

Wyniki dla przewozów pasażerów koleją proszę umieścić w poniższej tabeli, a następnie wykorzystując wprowadzone formuły i wklejając **tylko** nowe dane, proszę wyznaczyć prognozy dla przewozów pasażerów transportem samochodowym.

Wyniki prognozy metodą naiwną	Przewozy pasażerów (mln osób)	
	Transport kolejowy	Transport samochodowy
Prognoza na 2020 rok		
Średni bezwzględny procentowy błąd prognozy (MAPE)		

Przykład 3 (metoda średniej ruchomej)

Nieco bardziej złożoną metodą jest prognozowanie za pomocą tzw. średniej ruchomej. Średnia ruchoma jest wyznaczana nie z całego szeregu czasowego, ale dla pewnego jego (najczęściej dość krótkiego) fragmentu kolejnych wyrazów. W naszym przypadku dla danego okresu czasowego wyznaczać będziemy średnią obserwacji poprzedzających. Analiza ponownie dotyczy tego samego pliku i tej samej zmiennej co w przykładzie 1.

- 1) Proszę skopiować szereg czasowy dotyczący *Przewozów pasażerów koleją* do arkusza programu *Excel* (dane powinny znaleźć się w kolumnie B).
- 2) W kolumnie C proszę wyznaczyć średnią ruchomą **5-okresową** dla każdego roku z poprzednich obserwacji, począwszy od 1995 (dla tego roku należy wziąć średnią z lat 1990-1994). Dla wcześniejszych lat wyznaczenie średniej ruchomej jest niemożliwe. Oczywiście średnią wyznaczamy także dla roku **2020** – jest to prognoza na ten okres.
- 3) W kolumnie D proszę wyznaczyć wartości błędu bezwzględnego procentowego dla każdej obserwacji (APE) i wyliczyć z nich wartość średnią (czyli MAPE).

Wyniki prognozy za pomocą 5-okresowej średniej ruchomej	Przewozy pasażerów (mln osób)	
	Transport kolejowy	Transport samochodowy
Prognoza na 2020 rok		
MAPE		

Przykład 4 (ważona średnia ruchoma)

Logicznie uzasadnione wydaje się, iż na przyszłą (prognozowaną) wartość szeregu czasowego większy wpływ wywierają obserwacje niedawne. Proszę dokonać prognozy danych dotyczących funkcjonowania transportu w Polsce na rok 2020, stosując metodę średniej ruchomej **5-okresowej** z wagami **10, 10, 10, 20 i 50**.

Proszę uzupełnić wyniki dla zmiennych zawartych w tabeli.

Uwaga! W programie *Excel* nie ma formuły na średnią ważoną. Trzeba wykorzystać formułę **SUMA.ILOCZYNÓW** lub po prostu wpisać formułę ręcznie.

Wyniki prognozy metodą średniej 5-okresowej z wagami 10, 10, 10, 20 i 50	Przewozy pasażerów (mln osób)	
	Transport kolejowy	Transport samochodowy
Prognoza na 2020 rok		
MAPE		

Zaprezentowane w przykładzie 2, 3 i 4 metody prognozowania należą do tzw. metod adaptacyjnych. W przeciwieństwie do stosowanej w przykładzie 1. i na wcześniejszych zajęciach analizy regresji (trendów) nie wyznaczamy tu jednego równania dla całego zakresu danych, a wartość prognozy na kolejne okresy uzależniamy od kolejnych pojawiających się obserwacji.

Do metod adaptacyjnych należą też modele wyrównywania wykładniczego omówione na laboratoriach nr 9.