Prognozowanie i symulacje (lab. 8)
Trendy jednoimiennych okresów

Wprowadzenie

W przypadku występowania w szeregu czasowym wyraźnych wahań sezonowych wraz z trendem, do sporządzenia prognozy można zastosować ***metodę trendów jednoimiennych okresów***. Pod tą nazwą kryje się dość prosta koncepcja analizy, która polega na oddzielnym oszacowaniu modeli trendów dla każdego okresu cyklu sezonowości. Jeżeli trend jest taki sam dla różnych okresów (miesięcy) wtedy wyniki otrzymane za pomocą tej metody będą takie same jak dla metody wskaźnikowej omawianej na zajęciach nr 7.

Jeżeli jednak trendy są dla różnych miesięcy różne – na przykład wzrost w miesiącach zimowych, spadek w letnich, wtedy metoda trendów jednoimiennych okresów da lepsze wyniki.
Z technicznego punktu widzenia dopasujemy po prostu 12 modeli trendów oddzielnie dla danych ze stycznia, oddzielnie dla danych z lutego, etc. Zagadnienie sprowadza się więc do podzielenia szeregu danych miesięcznych na 12 plików danych i przeprowadzenia prognozy dla każdego z nich niezależnie.
W przypadku pracy z programem *STATISTICA* analizę taką można zrealizować poprzez:

- zastosowanie warunków selekcji;
- wykorzystanie wykresów w postaci skategoryzowanej (względem miesięcy);
- wykorzystanie analizy grupami.

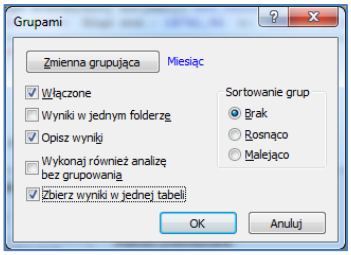
Poniżej przedstawiono propozycję przeprowadzenia analizy regresji grupami oraz zaprezentowano niebanalne (choć proste!) operacje na arkuszu danych *STATISTICA*, które pozwalają na uzyskanie prognozy za pomocą metody trendów jednoimiennych okresów dla 12-tu miesięcy kolejnego roku w ciągu co najwyżej kilku minut.

Przykład 1 (*Wypadki w Polsce od 1998 roku (M)*)

Celem analizy jest sporządzenie prognozy ***Liczby wypadków drogowych*** na wszystkie miesiące **2025** roku za pomocą metody trendów jednoimiennych okresów, **na podstawie danych do 2024 r.**

Wykorzystana zostanie ***Analiza regresji*** z opcją przeprowadzania ***Analizy grupami***.

1. W oknie wyboru zmiennych *Analizy regresji* jako zmienną zależną wskazujemy ***Liczbę wypadków*** zaś jako zmienną niezależną zmienną ***Rok*** (w tym przykładzie, nieco inaczej niż dotychczas posłużymy się po prostu tą zmienną, zamiast wprowadzać pomocniczą zmienną *X* zawierającą numery kolejnych okresów).
2. Wprowadzamy warunek selekcji: ***Rok <= 2024***
3. Przechodzimy do okna wyników ***Analizy regresji*** i wybieramy opcję ***Grupami***, po czym jako zmienną grupującą wskazujemy *Miesiąc* i zaznaczamy dodatkowo *Zbierz wyniki w jednej tabeli* i inne opcje tak jak na rysunku obok.
4. W zakładce ***Reszty, założenie, predykcja*** w oknie ***Predykcji…*** wskazujemy jako wartość zmiennej ***Rok*** oczywiście wartość ***2025***.



5. Dalsze operacje przeprowadzamy na arkuszu wynikowym, w którym mamy wszystkie informacje o prognozach punktowych i przedziałowych, tyle, że w dość niewygodnym do prezentowania wyników układzie.
6. Z arkusza wynikowego usuwamy kolumnę nr 2 i 3 (zawierają one pośrednie obliczenia, a nie wyniki prognoz).
7. Dodajemy na końcu arkusza wynikowego jedną pomocniczą kolumnę, nazywamy ją *Wynik* i wypełniamy ją sekwencyjnie wartościami 1, 2, 3, 4 i 5 (będą to numery kolumn w przekształconym arkuszu wyników). Najprościej wykorzystać w tym celu formułę odwołującą się do numerów przypadków i reszt z dzielenia przez 5 w postaci: **=modulo(v0-1;5)+1**
8. Aby przekształcić arkusz wynikowy wykorzystujemy polecenie ***Dane / Rozrzuć po zmiennych…*** i na liście zmiennych wskazujemy:
   - zmienną *Wynik* jako kodującą *Kolumny*;
   - zmienną *Miesiąc* jako kodującą *Wiersze*;
   - zmienną *Wagi b* jako *Zmienną z wartościami*.
9. Po zatwierdzeniu uzyskujemy arkusz w przekształconej postaci, skąd usuwamy kolumny nr 2 i 3, uzyskując ostatecznie zestawienie miesięcznych prognoz na 2025 rok w takiej postaci jak na rysunku obok.
10. Formatujemy wyniki (ponieważ prognozujemy liczbę wypadków zaokrąglamy liczby do całości), nazywamy odpowiednio kolumny i kończymy obliczenia. Tabelę w takiej postaci można wkleić do raportu w programie *Word* czy *PowerPoint*.

1 Miesiąc	2 Prognoza na 2025 r.	3 Lewy koniec 95% przedziału ufności	4 Prawy koniec 95% przedziału ufności
1	1090	703	1478
2	1037	803	1270
3	1185	950	1419
4	1251	986	1516
5	1619	1389	1849
6	1897	1688	2105
7	1897	1753	2042
8	1855	1681	2028
9	1913	1786	2041
10	1732	1539	1924
11	1412	1150	1675
12	1570	1155	1985

Funkcja modulo(a; b) pozwala wyznaczyć resztę z dzielenia liczby a przez b. Na przykład:
Modulo(2; 5) = 2 modulo(7; 5) = 2 modulo(12; 5) = 2
Odpowiednio wykorzystana funkcja modulo pozwala wypełniać kolumny cyklicznie wartościami z pewnego zakresu.

11. W uzyskanym arkuszu możemy dokonać jeszcze pewnych ciekawych modyfikacji – na przykład prognozę przedziałową można zapisać w jednej kolumnie w nawiasie po prognozie punktowej (taki zapis często stosuje się w profesjonalnych opracowań zawierających wyniki prognoz.
    Proszę wykonać następujące czynności:
    - dodajemy nową zmienną na końcu arkusza (będzie to zmienna numer 5);
    - nazywamy nową zmienną *Prognoza punktowa (wraz z 95% p.u.)*;
    - zmieniamy typ **każdej** ze zmiennych 2, 3 i 4 na *tekstowy*;
    - wartości zmiennej 5 wyliczamy za pomocą formuły: **=v2+" (" +v3+"-"+v4+")"** i uzyskujemy w ostatniej kolumnie zapis, który jest wygodny do prezentacji prognozy wraz z jej oszacowaniem w raporcie statystycznym.

1 Miesiąc	2 Prognoza na 2025 r.	3 Lewy koniec 95% przedziału ufności	4 Prawy koniec 95% przedziału ufności	5 Prognoza (95% p.u.)
1	1 090	703	1 478	1 090 (703-1 478)
2	1 037	803	1 270	1 037 (803-1 270)
3	1 185	950	1 419	1 185 (950-1 419)
4	1 251	986	1 516	1 251 (986-1 516)
5	1 619	38	1 849	1 619 (1 389-1 849)
6	1 897	1	2 105	1 897 (1 688-2 105)
7	1 897	75	2 042	1 897 (1 753-2 042)
8	1 855	1 681	2 028	1 855 (1 681-2 028)
9	1 913	1 786	2 041	1 913 (1 786-2 041)
10	1 732	1 539	1 924	1 732 (1 539-1 924)
11	1 412	1 150	1 675	1 412 (1 150-1 675)
12	1 570	1 155	1 985	1 570 (1 155-1 985)

Jeżeli dodatkowo chcielibyśmy podać równania modeli trendów dla każdego miesiąca wystarczy w tym celu wznowić analizę (Ctrl + *R*), przejść do zakładki ***Podstawowe*** i wywołać wyniki przyciskiem ***Podsumowanie: wyniki regresji***. W skoroszycie wyników pojawiają się dwa arkusze:

- W pierwszym z wyświetlonych arkuszy znajdziemy wartości współczynnika determinacji dla każdego miesiąca. **Aby zgrupować wartości współczynnika determinacji (w arkuszu pod nazwą *Wielorakie R2*) dla poszczególnych miesięcy obok siebie, wystarczy dane posortować względem nazwy przypadków. Współczynnik determinacji proszę wyświetlić w postaci procentowej z jednym miejscem po przecinku.**
- W drugim arkuszu znajdują się wartości współczynników regresji liniowej (*B*) dla każdego miesiąca.

Przykład 2 (*Krajowe przewozy lotnicze w UE (1-12)*)

Proszę sporządzić prognozy za pomocą metody trendów jednoimiennych okresów dla ***Liczby pasażerów w Norwegii*** na wszystkie miesiące roku **2020** i **2021**. W programie *Word* proszę przygotować tabelę według poniższego wzorca, a następnie wkleić wyniki prognoz i wartości współczynnika determinacji.

Miesiąc	**Prognoza liczby pasażerów w Norwegii** (prognoza punktowa wraz z 90% przedziałem ufności)		**Współczynnik determinacji modelu liniowego (%)**
	Rok 2020	**Rok 2021**	
styczeń			
luty			
marzec			
kwiecień			
maj			
czerwiec			
lipiec			
sierpień			
wrzesień			
październik			
listopad			
grudzień			

Przykład 3 (*Wypadki w Polsce od 1998 roku (M)*)

Pokażemy jeszcze jak można otrzymać równania trendów dla poszczególnych miesięcy za pomocą graficznych narzędzi analizy danych. W tym celu posłużymy się wykresem rozrzutu (choć oczywiście moglibyśmy też zastosować wykres liniowy). Wybieramy polecenie ***Wykres rozrzutu*** i na pierwszej liście zmiennych wskazujemy zmienną ***Rok***, a na drugiej ***Liczbę wypadków***. W zakładce ***Więcej*** włączamy: ***Dopasowanie liniowe*** oraz ***Współczynnik determinacji***.

Aby otrzymać wykresy dla każdego miesiąca włączamy jeszcze opcję ***Grupami*** wskazując jako zmienną grupującą ***Miesiąc***. Odznaczamy opcję ***Wykonaj analizę bez grupowania***, poza tym nie wprowadzamy żadnych zmian.

Po naciśnięciu przycisku *OK* otrzymujemy 12 wykresów, pokazujących trendy dla jednoimiennych okresów – dla każdego miesiąca oddzielnie. Jak to już było widać wcześniej, przewidywalnością i systematycznym spadkiem liczby wypadków cechują się przede wszystkim miesiące letnie (współczynnik determinacji jest dla nich najwyższy, nawet ponad 90% – dla lipca i września).