

# Prognozowanie i symulacje (lab. 7)

## Modele regresji ze zmiennymi wskaźnikowymi

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

**UWAGA: Proszę zapisać plik pod inną nazwą i usunąć dane z 2023 r.**

Celem analizy będzie dokonanie prognozy **liczby wypadków** na **wszystkie miesiące roku 2023 i 2024**, za pomocą modelu trendu liniowego z wahaniami sezonowymi (miesięcznymi).

Aby wykorzystać w tym celu **Analizę regresji wielorakiej** należy w wyjściowym zbiorze danych dodać zmienną określającą numer kolejnej obserwacji (X) oraz 12 zmiennych wskaźnikowych, za pomocą których kodowany będzie efekt sezonowości (każdego miesiąca).

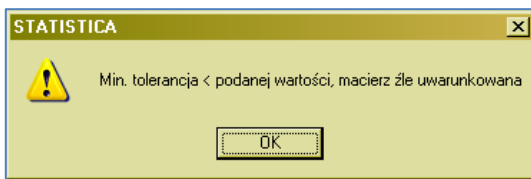
Najpierw należy rozbić zmienną **DATA** na zmienne **Rok** i **Miesiąc**, wykorzystując w tym celu **Operacje na danych**.

Wartości zmiennej X zawierającej numer zmiennej czasowej oraz dwunastu zmiennych **Styczeń**, ..., **Grudzień** określających efekt sezonowości dla danego miesiąca wyznaczamy za pomocą formuł arkusza danych wykorzystując do tego polecenie **Zmienne / Wszystkie specyfikacje**, co umożliwi sprawną edycję nazw i formuł dla kilku zmiennych naraz.

Po wprowadzeniu zmiennych pomocniczych przechodzimy do okna wyboru zmiennych modułu **Regresja wieloraka**, wybierając na liście zmiennych zależnych **Liczbę wypadków** a na liście zmiennych niezależnych zmienną X oraz **Styczeń**, ..., **Grudzień**.

Nazwa	Typ	kod BD	Długość	Długa nazwa (etykieta lub formuła)	Typ skali
7 X	Podw. prec.	-999999998			Automa
8 Styczeń	Podw. prec.	-999999998		=v2=1	Automa
9 Luty	Podw. prec.	-999999998		=v2=2	Automa
10 Marzec	Podw. prec.	-999999998		=v2=3	Automa
11 Kwiecień	Podw. prec.	-999999998		=v2=4	Automa
12 Maj	Podw. prec.	-999999998		=v2=5	Automa
13 Czerwiec	Podw. prec.	-999999998		=v2=6	Automa
14 Lipiec	Podw. prec.	-999999998		=v2=7	Automa
15 Sierpień	Podw. prec.	-999999998		=v2=8	Automa
16 Wrzesień	Podw. prec.	-999999998		=v2=9	Automa
17 Październik	Podw. prec.	-999999998		=v2=10	Automa
18 Listopad	Podw. prec.	-999999998		=v2=11	Automa

Jednak próba przejścia do okna wyników powoduje wyświetlenie następującego komunikatu:



Ten „techniczny” i niezbyt jasny komunikat, oznacza że: **niektóre zmienne niezależne w modelu zawierają informację, którą można wywnioskować z wartości innych zmiennych, czyli nie są niezależne.** I faktycznie, proszę sprawdzić, że zachodzi np. równanie: **Styczeń = 1 – (Luty + ... + Grudzień)**

Analiza zmiennych wskaźnikowych **Styczeń**, ..., **Grudzień** pozwala stwierdzić, iż na podstawie informacji o wartościach dowolnych 11-tu spośród nich można określić wartości 12-ej zmiennej, czyli nie są one niezależne. Pomijamy na liście zmiennych niezależnych **Styczeń** (arbitralnie ustalamy ten miesiąc, jako punkt odniesienia) i teraz już bez problemu przechodzimy do okna wynikowego.

Przechodzimy do zakładki **Podstawowe** i za pomocą przycisku **Podsumowanie: wyniki regresji** wywołujemy tabelę z podstawowymi wynikami analizy, których fragment został zaprezentowany i omówiony poniżej.

	R= ,95577026 R^2= ,91349679 Popraw. R2= ,90987993 F(12,287)=252,57 p<0,000 Błąd std. estymacji: 335,28					
	b*	Bł. std. z b*	b	Bł. std. z b	t(287)	p
N=300						
W. wolny			4297,149	74,48928	57,6881	0,000000
X	-0,830970	0,017375	-10,699	0,22370	-47,8261	0,000000
luty	-0,069045	0,023507	-278,541	94,83201	-2,9372	0,003580
marzec	0,017281	0,023507	69,757	94,83280	0,7356	0,462585
kwiecień	0,081826	0,023508	362,376	94,83412	3,8212	0,000163
maj	0,238311	0,023508	961,395	94,83597	10,1374	0,000000
czerwiec	0,303766	0,023509	1225,153	94,83834	12,9215	0,000000
lipiec	0,320567	0,023509	1293,232	94,84125	13,6358	0,000000

Model jest dopasowany w 91,3% co jest wynikiem dobrym, pozwalającym na wykorzystanie go do prognozowania.

- Współczynniki modelu są w kolumnie „b”
- 1) Przy X – z każdym miesiącem liczba wypadków spada o ok. 10-11.
  - 2) Dla lutego – w tym miesiącu jest średnio rzecz biorąc 279 wypadków mniej niż w styczniu.
  - 3) Dla kwietnia – o 362 wypadków więcej niż w styczniu.
  - 4) I podobnie dla innych miesięcy.

Wszystkie zmienne poza „MARZEC” są istotne statystycznie (p < 0,05). Wracamy więc do okna analizy, cofamy się do miejsca wyboru zmiennych i odznaczamy zmienną „MARZEC”, ponownie wywołując okno podsumowania.

## Prognozowanie i symulacje (lab. 7)

### Modele regresji ze zmiennymi wskaźnikowymi

Wznawiamy analizę i cofamy się za pomocą przycisku *Anuluj* do momentu, kiedy istnieje możliwość wyboru zmiennych do analizy. Po wyeliminowaniu zmiennej *MARZEC* z listy zmiennych niezależnych (**oczywiście cały czas pomijamy także zmienną *STYCZEŃ***) otrzymujemy ostateczną postać modelu:

*W kolumnie b znajdują się współczynniki modelu.*

*Współczynnik przy X pozwala stwierdzić, iż z każdym miesiącem liczba wypadków spada średnio rzecz biorąc o 11, czyli w okresie rocznym o ok. 130.*

*Współczynniki przy zmiennych LUTY, KWIECIEŃ, ..., GRUDZIEŃ pozwalają określić efekt sezonowości (wartości są odnoszone do liczby wypadków ze stycznia, dlatego ich suma w przeciwieństwie do addytywnych wskaźników sezonowości nie wynosi 0)*

Podsumowanie regresji zmiennej zależnej: Liczba wypadków (Wypadki w F)						
R= ,95568494 R <sup>2</sup> = ,91333371 Popraw. R <sup>2</sup> = ,91002353						
F(11,288)=275,92 p<0,0000 Błąd std. estymacji: 335,01						
N=300	b*	Bł. std. z b*	b	Bł. std. z b	t(288)	p
W. wolny			4332	57,53	75,30	0,0000
X	-0,83	0,02	-11	0,22	-47,86	0,0000
luty	-0,08	0,02	-313	82,06	-3,82	0,0002
kwiecień	0,08	0,02	327	82,06	3,99	0,0001
maj	0,23	0,02	927	82,06	11,29	0,0000
czerwiec	0,30	0,02	1191	82,07	14,51	0,0000
lipiec	0,31	0,02	1258	82,07	15,33	0,0000
sierpień	0,32	0,02	1285	82,07	15,66	0,0000
wrzesień	0,29	0,02	1164	82,08	14,18	0,0000
październik	0,32	0,02	1303	82,08	15,87	0,0000
listopad	0,18	0,02	730	82,09	8,89	0,0000
grudzień	0,19	0,02	777	82,09	9,46	0,0000

W kolejnym etapie przechodzimy do prognozowania **liczby wypadków** na poszczególne miesiące 2023 i 2024 roku, za pomocą narzędzie prognozowania dostępnego w oknie *Reszty, Założenia, Predykcja*. Ustalamy ufność prognozy na poziomie 90%.

**Przykładowy** sposób ustalenia wartości zmiennych niezależnych dla prognozy **na luty 2024** podano poniżej.



Obliczanie wartości (Wypadki w Polsce zmiennej: Liczba wypadków)			
Zmienna	Wagi b	Wartość	Wagi b * Wartość
X	-10,698	314,0000	-3359,14
luty	-313,420	1,0000	-313,42
kwiecień	327,496	0,0000	0,00
maj	926,514	0,0000	0,00
czerwiec	1190,572	0,0000	0,00
lipiec	1258,350	0,0000	0,00
sierpień	1285,367	0,0000	0,00
wrzesień	1163,505	0,0000	0,00
październik	1302,843	0,0000	0,00
listopad	730,101	0,0000	0,00
grudzień	776,599	0,0000	0,00
W. wolny			4331,91
Przewidyw.			<b>659</b>
-90,0%GU			<b>533</b>
+90,0%GU			<b>786</b>

Proszę uzupełnić poniższą tabelę, podając prognozę punktową i przedziałową **liczby wypadków** na 2023 i 2024 rok.

Prognoza <b>liczby wypadków</b>	2023		2024	
	Prognoza	Prognoza przedziałowa (90% p.u.)	Prognoza	Prognoza przedziałowa (90% p.u.)
<i>styczeń</i>				
<i>luty</i>			659	533-786
<i>marzec</i>				
<i>kwiecień</i>				
<i>maj</i>				
<i>czerwiec</i>				
<i>lipiec</i>				
<i>sierpień</i>				
<i>wrzesień</i>				
<i>październik</i>				
<i>listopad</i>				
<i>grudzień</i>				