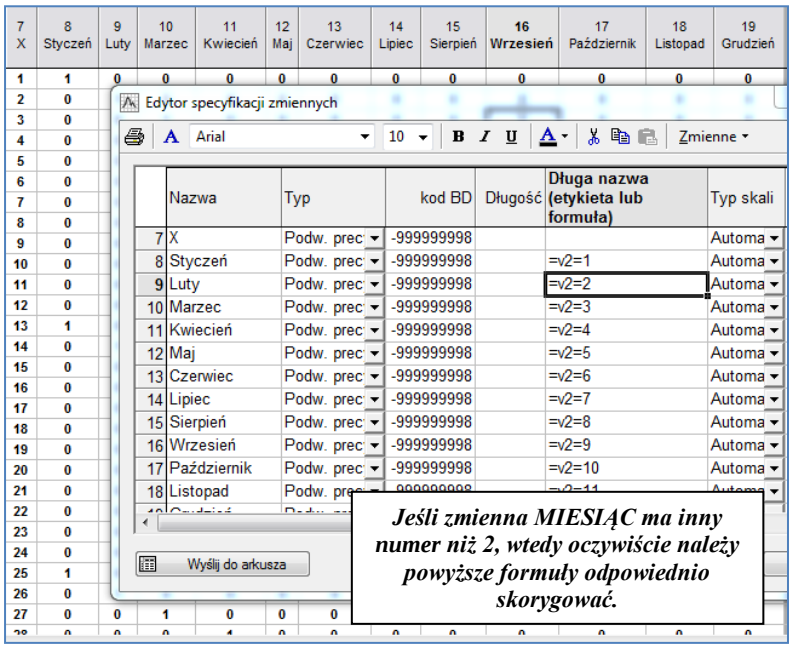
Prognozowanie i symulacje (lab. 7)
Modele regresji ze zmiennymi wskaźnikowymi

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

UWAGA: Proszę zapisać plik pod inną nazwą i usunąć dane z 2024 i 2025 r.

Celem analizy będzie dokonanie prognozy **liczby wypadków** na **wszystkie miesiące roku 2024 i 2025 r.**, 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).



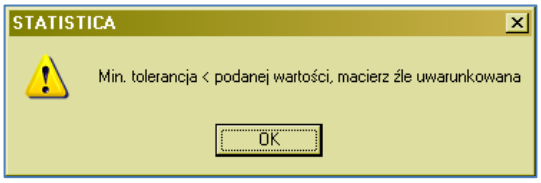
Jeśli zmienna MIESIĄC ma inny numer niż 2, wtedy oczywiście należy powyższe formuły odpowiednio skorygować.

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

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 umożliwia 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***.

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= ,95897969 R^2= ,91964205 Popraw. R2= ,91641699
F(12,299)=285,15 p<0,0000 Błąd std. estymacji: 331,05

N=312	b*	Bł. std. z b*	b	Bł. std. z b	t(299)	p
W. wolny			4313	72,14	59,8	0,0000
X	-0,944	0,01641	-11	0,21	-51,5	0,0000
Luty	-0,068	0,02220	-281	91,82	-3,1	0,0024
Marzec	0,015	0,02220	64	91,82	0,7	0,4878
Kwiecień	0,085	0,02220	351	91,82	3,8	0,0002
Maj	0,229	0,02220	947	91,82	10,3	0,0000
Czerwiec	0,292	0,02220	1207	91,82	13,1	0,0000
Lipiec	0,307	0,02220	1271	91,83	13,8	0,0000

Model jest dopasowany w 91,96% 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. 11.***
2) ***Dla lutego – w tym miesiącu jest średnio rzecz biorąc 281 wypadków MNIEJ niż w styczniu.***
3) ***Dla kwietnia – o 351 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.

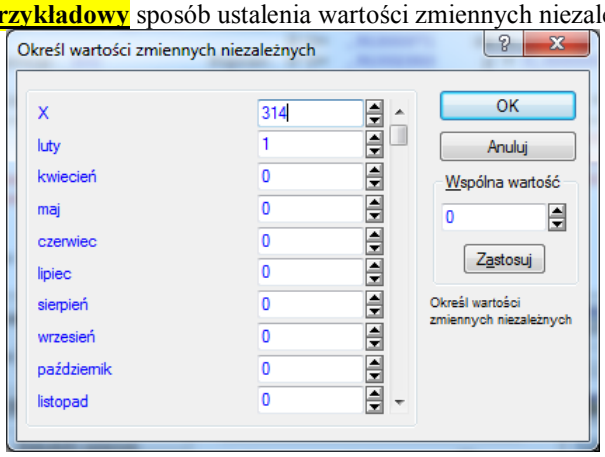
Współczynnik przy X pozwala stwierdzić, iż z każdym miesiącem liczba wypadków spada średnio rzecz biorąc o 11, czyli w okresie rocznym o ok. 130.

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 (Wypa
R= ,95891207 R^2= ,91951236 Popraw. R2= ,91656114
F(11,300)=311,57 p<0,0000 Błąd std. estymacji: 330,77

N=312	b*	Bł. std. z b*	b	Bł. std. z b	t(300)	p
W. wolny			4345,2	55,715	77,99	0,0000
X	-0,8444	0,0164	-10,7	0,208	-51,52	0,0000
Luty	-0,0755	0,0192	-312,4	79,448	-3,93	0,0001
Kwiecień	0,0771	0,0192	319,0	79,449	4,01	0,0001
Maj	0,2211	0,0192	914,7	79,451	11,51	0,0000
Czerwiec	0,2842	0,0192	1175,5	79,452	14,80	0,0000
Lipiec	0,2996	0,0192	1239,3	79,455	15,60	0,0000
Sierpień	0,3062	0,0192	1266,7	79,458	15,94	0,0000
Wrzesień	0,2793	0,0192	1155,2	79,461	14,54	0,0000
Październik	0,3093	0,0192	1279,4	79,466	16,10	0,0000
Listopad	0,1735	0,0192	717,5	79,470	9,03	0,0000
Grudzień	0,1835	0,0192	759,1	79,475	9,55	0,0000

W kolejnym etapie przechodzimy do prognozowania **liczby wypadków** na poszczególne miesiące **2024 i 2025 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 Polsc
zmiennej: Liczba wypadków

Zmienna	Wagi b	Wartość	Wagi b *Wartość
X	-10,719	314,0000	-3365,64
Luty	-312,442	1,0000	-312,44
Kwiecień	318,956	0,0000	0,00
Maj	914,675	0,0000	0,00
Czerwiec	1175,547	0,0000	0,00
Lipiec	1239,343	0,0000	0,00
Sierpień	1266,677	0,0000	0,00
Wrzesień	1155,165	0,0000	0,00
Październik	1279,422	0,0000	0,00
Listopad	717,487	0,0000	0,00
Grudzień	759,051	0,0000	0,00
W. wolny			4345,21
Przewidyw.			667
-90,0%GU			547
+90,0%GU			788

Proszę uzupełnić poniższą tabelę, podając prognozę punktową i przedziałową **liczby wypadków** na 2024 i 2025 rok.

Prognoza *liczby wypadków*	**2024**		**2025**	
	Prognoza	**Prognoza przedziałowa (90% p.u.)**	**Prognoza**	**Prognoza przedziałowa (90% p.u.)**
styczeń				
luty	**667**	**547-788**		
marzec				
kwiecień				
maj				
czerwiec				
lipiec				
sierpień				
wrzesień				
październik				
listopad				
grudzień				

ZADANIE DOMOWE: Posługując się prognozami na 2024 r. z perspektywy 2023 r. oraz faktyczną liczbą wypadków, która była podana w pierwotnym arkuszu danych, proszę sporządzić tabelę w Excelu, z wyliczonymi błędami prognozy dla wszystkich miesięcy 2024 i 2025 r.
Procentowy błąd prognozy wyznaczamy ze wzoru:

$$\text{Błąd procentowy prognozy} = (Y_t - Y_t^*)/\, Y_t \cdot 100\%$$