

Uwzględnienie czynnika sezonowego w analizie regresji możliwe jest poprzez dodanie zmiennych wskaźnikowych odzwierciedlających poszczególne analizowane podokresy, np. 12 zmiennymi dummy w przypadku danych miesięcznych, 4 zmienne dummy przy danych kwartalnych itd. (metoda wskaźnikowa), lub oszacowanie funkcji trendu dla każdego podokresu osobno (metoda trendów jednoimiennych okresów). W przypadku metody trendów jednoimiennych okresów można wykorzystać wiele możliwości w celu oszacowania funkcji trendu osobno dla każdego okresu, np. warunków selekcji, opcji kategoryzujących, analizy w grupach, gdzie zmienną grupującą jest znacznik podgrupy, np. zmienna kodująca miesiąc, zmienna kodująca kwartał itd..

Zadanie 1. Metoda wskaźnikowa (Plik: krajowe przewozy lotnicze UE (1-12).sta)

Za pomocą metody wskaźnikowej wyznaczyć prognozę na rok 2020 dla liczby pasażerów przewożonych transportem lotniczym we Francji. W tym celu należy:

1. Dodać 13 nowych zmiennych, 1 zmienna **X** odzwierciedlająca pozycję zmiennej czasowej w analizowanym szeregu oraz 12 zmiennych dummy 0-1, wskazujące wskaźnik sezonowy dla danego miesiąca (czyli na przykład zmienna 1 powinna przyjmować wartość 1 w styczniu, a 0 w pozostałych miesiącach, zmienna 2 powinna przyjmować wartość 1 w lutym, a 0 w pozostałych miesiącach, itd.). Wypełnienie liczbami nowych 13 zmiennych można wykonać szybko wykorzystując zakładkę Dane/Wszystkie

	Nazwa	Typ	kod BD	Długość	Długa nazwa (etykieta lub formuła)
1	DATA	Podw. precyz	-999999998		
2	Rok	Podw. precyz	-999999998		
3	Miesiąc	Podw. precyz	-999999998		
4	X	Podw. precyz	-999999998		=v0
5	styczeń	Podw. precyz	-999999998		=v3=1
6	luty	Podw. precyz	-999999998		=v3=2
7	marzec	Podw. precyz	-999999998		=v3=3
8	kwiecień	Podw. precyz	-999999998		=v3=4
9	maj	Podw. precyz	-999999998		=v3=5
10	czerwiec	Podw. precyz	-999999998		=v3=6

specyfikacje. Tam można nadać nazwy zmiennym (wystarczy wpisać jedynie dwie nazwy: styczeń i luty, a następnie zaznaczyć obie komórki i przeciągnąć je w dół. Podobnie można postąpić przypisując 1 w poszczególnych miesiącach. Wystarczy dla stycznia posłużyć się funkcją (=v3=1), a następnie przeciągnąć ją przez wszystkie miesiące.

2. Na pasku narzędzi wywołać panel regresji wielorakiej (Statystyka/Regresja wieloraka) i wybrać odpowiednie zmienne: jako Y (zmienną zależną) wybrać główną zmienną analizowaną, czyli stopę bezrobocia we Francji, natomiast jako zmienne niezależne wybrać wszystkie nowoutworzone zmienne, a następnie zatwierdzić wybór klikając w OK.

Po zatwierdzeniu wyboru zmiennych na ekranie pojawi się następujący komunikat:

Ten komunikat oznacza, że zmienne niezależne dummy są współliniowe, czyli nie są niezależne  Min. tolerancja < podanej wartości, macierz źle uwarunkowana

i wartości tych zmiennych dają się obliczyć na podstawie wartości pozostałych zmiennych objaśniających, czyli zachodzi równanie **Styczeń=1-(luty+marzec+...+grudzień)**. W celu wyeliminowania tego błędu, a zarazem spełnienia warunku o niezależności zmiennych objaśniających, należy pominąć jedną ze zmiennych dummy spośród wybranych zmiennych objaśniających. Arbitralnie **pomijamy pierwszy z miesięcy (styczeń) spośród grupy zmiennych objaśniających**, zatwierdzamy wybór zmiennych, a następnie

wywołujemy wyniki analizy regresji, klikając w zakładce podstawowe na *Podsumowanie: Wyniki regresji*.
W wyniku otrzymujemy następujące zestawienie:

Podsumowanie regresji zmiennej zależnej: Francja (Krajowe przewozy lotnicze)
R= ,93790700 R²= ,87966953 Popraw. R2= ,87102303
F(12,167)=101,74 p<0,0000 Błąd std. estymacji: 94,018

	b*	Bł. std. z b*	b	Bł. std. z b	t(167)	p
N=180						
W. wolny			1908,750	26,85682	71,07132	0,000000
X	0,493206	0,026902	2,478	0,13516	18,33322	0,000000
lut	-0,126515	0,036346	-119,502	34,33082	-3,48088	0,000638
marzec	0,193280	0,036347	182,565	34,33161	5,31770	0,000000
kwiecień	0,280231	0,036348	264,696	34,33294	7,70967	0,000000
maj	0,351697	0,036350	332,200	34,33481	9,67531	0,000000
czerwiec	0,541342	0,036353	511,331	34,33720	14,89147	0,000000
lipiec	0,678802	0,036356	641,170	34,34013	18,67117	0,000000
sierpień	0,423470	0,036359	399,993	34,34359	11,64682	0,000000
wrzesień	0,282851	0,036363	267,170	34,34757	7,77843	0,000000
październik	0,314392	0,036368	296,963	34,35210	8,64468	0,000000
listopad	0,042682	0,036374	40,316	34,35715	1,17344	0,242291
grudzień	0,163582	0,036380	154,514	34,36273	4,49654	0,000013

R2 – współczynnik determinacji: % dopasowania modelu do danych

p – prawdopodobieństwo komputerowe, informuje o statystycznej istotności współczynników regresji. Jeżeli wartość $p < 0.05$, wówczas dana zmienna jest statystycznie istotna dla objaśnienia Y. Jeżeli wartość $p > 0.05$, wówczas zmienna nie jest statystycznie istotna i powinna zostać pominięta w dalszej analizie. W tym przypadku, jedynie zmienna listopad posiada wartość $p > 0.05$ i powinna zostać pominięta. W tym celu należy wrócić do wyboru zmiennych, odznaczyć listopad z listy zmiennych objaśniających a następnie ponownie wywołać okno regresji.

Kolumna b – oszacowane współczynniki regresji. Na ich podstawie można odtworzyć dopasowany wzór funkcji regresji opisujący zależność między zmienną zależną a zmiennymi objaśniającymi. Analizuje się je przy założeniu, że pozostałe wartości współczynników =0, np.

X – z każdym miesiącem liczba osób przewożonych transportem lotniczym wzrasta o 2 478 osób

Luty – w lutym przewozi się średnio o 119 502 mniej osób niż w styczniu

- Po ustaleniu ostatecznej wersji modelu, przechodzimy do zakładki Reszty/założenia/predykcja i ustalamy przedział ufności prognozy na 90% (Alfa=0.1) . Przedział ufności prognozy pozwala na oszacowanie prognozy przedziałowej. Im szerszy, tym mniej dokładna prognoza. Następnie wywołujemy okno Predykcja zmiennej zależnej.

- W celu oszacowania prognozy na styczeń 2020, wystarczy w polu X wpisać wartość 181 (jest to miejsce stycznia 2020 w analizowanym szeregu czasowym, odczytanym z kolumny X).

	2019	t	t/s	u	u
Lip-2019	2019	8	176	0	0
Sie-2019	2019	9	177	0	0
Wrz-2019	2019	10	178	0	0
Paź-2019	2019	11	179	0	0
Gru-2019	2019	12	180	0	0
Sty-2020	2020	1	181	1	0
Lut-2020	2020	2	182	0	1
Mar-2020	2020	3	183	0	0
Kwi-2020	2020	4	184	0	0
Mai-2020	2020	5	185	0	0

W celu oszacowania prognozy na luty 2020, w polu X wpisujemy pozycję lutego 2020 odczytana z kolumny X (poz. 182), a ponieważ szacowaną wartością jest luty, należy przypisać lutemu wartość 1 (pozostałe miesiące mają przypisaną wartość 0).

Proszę wypełnić poniższą tabelę wartościami prognozy przewozów lotniczych w 2020 roku we Francji.

	Prognoza punktowa na 2020 rok	Prognoza przedziałowa
Styczeń		
Luty		
Marzec		
Kwiecień		
Maj		
Czerwiec		
Lipiec		
Sierpień		
Wrzesień		
Październik		
Listopad		
Grudzień		

Zadanie 2. Metoda trendów jednoimiennych okresów (Plik: krajowe przewozy lotnicze UE (1-12).sta)

Wykonać prognozę przewozów lotniczych dla Polski na wszystkie miesiące roku 2020 i 2021 wykorzystując analizę regresji w grupach.

Jeżeli w analizowanym szeregu czasowym występuje trend oraz wahania sezonowe można zastosować metodę trendów jednoimiennych okresów, która polega na oszacowaniu funkcji trendu osobo dla każdego podokresu (np. miesięcy w przypadku danych miesięcznych, kwartałów w przypadku danych kwartalnych, itd.). Jeżeli trend jest taki sam dla różnych podokresów, wówczas otrzymane wyniki są takie same jak przy zastosowaniu metody wskaźnikowej. Metodę jednoimiennych okresów można wykonać wykorzystując analizę w grupach, warunki selekcji (ekstrahując pojedyncze okresy z analizowanego szeregu), czy inne metody kategoryzujące.

Aby wykonać prognozę przewozów lotniczych stosując analizę regresji w grupach należy:

1. Wywołać okno regresji wielorakiej (Statystyka/regresja wieloraka), a następnie jako zmienną zależną Y wybrać przewozy lotnicze w Polsce, jako zmienną niezależną X wybrać Rok.
2. Po zatwierdzeniu wyboru zmiennych i przejściu do okna z wynikami regresji wybrać opcję **Grupami**, gdzie jako zmienną grupującą zaznaczyć **Miesiąc**. Dodatkowo odznaczyć opcję: wykonaj również analizę bez grupowania oraz zaznaczyć opcję: Zbierz wyniki w jednej tabeli.
3. W zakładce Reszty, założenia, predykcja ustawić szerokość przedziału ufności na 90% (alfa=0,1) oraz wybrać okres na który będzie obliczana prognoza (ROK=2020).
4. W celu ułatwienia sobie wykorzystania otrzymanych wyników, w sposób automatyczny można dokonać kilku przekształceń, tak aby zostały przedstawione w kompaktowym, zbiorczym formacie (wszystkie poniższe przekształcenia wykonujemy w otrzymanej tabeli wynikowej):
 - a. Usuujemy dwie „niepotrzebne” kolumny: wagi b oraz wartość.
 - b. Dodajemy nową zmienną pomocniczą Kolumna. Będzie ona wypełniona kolejnymi numerami od 1 do 5, reprezentującymi powtarzające się wyniki dla każdego otrzymanego podokresu. Można ro wykonać manualnie lub skorzystać z formuły obliczającą resztę z dzielenia przez 5 w postaci: $=\text{modulo}(v0-1;5)+1$
 - c. Na pasku narzędzi wybrać Dane/Sterta i jako zmienne wybrać:
 - i. Zmienne kod. Kolumny: Kolumna
 - ii. Zmienne z wartościami: wagi b
 - iii. Zmienne kod. Wiersze: miesiąc
 - d. Z otrzymanej tabeli usunąć zmienną V2 oraz V3, a następnie:
 - i. Zmienić określenia miesięcy z liczbowych na tekstowe
 - ii. Wybrać Dane/Wszystkie specyfikacje i zmienić typ wyświetlania zmiennej nr 2,3 i 4 z Podw.precyz. na tekstową.
 - iii. Dodać nową zmienną: Prognoza (90% PU) i wypełnić ją wartościami wynikającymi z połączenia informacji zawartych w Zmn 3,4,5 za pomocą formuły: $=v2+”(„,+ v3 + ”-„, + v4 + „)”$
 - iv. Do otrzymanej tabeli dodać zmienną R2. Aby je otrzymać należy przywołać okno edycji i w zakładce podstawowe wywołać Podsumowanie: wyniki regresji. W Skoroszycie odnaleźć: Stat.podsum.Zmn.zal. i posortować malejąco po nazwach przypadków.

1 Miesiąc	2 Prognoza punktowa	3 Dolna granica 90% PU	4 Górna granica 90% PU	5 Prognoza punktowa (wraz z 90% PU)	6 R2
styczeń	137,6	126,7	148,5	137,6 (126,7-148,5)	84,9%
luty	139,1	126,5	151,6	139,1 (126,5-151,6)	82,4%
marzec	164,2	149,7	178,6	164,2 (149,7-178,6)	83,5%
kwiecień	174,8	148,9	200,7	174,8 (148,9-200,7)	62,8%
maj	196,5	162,8	230,3	196,5 (162,8-230,3)	54,5%
czerwiec	200,3	163,5	237,1	200,3 (163,5-237,1)	48,4%
lipiec	208,3	173,0	243,7	208,3 (173,0-243,7)	54,5%
sierpień	197,9	179,1	216,7	197,9 (179,1-216,7)	80,4%
wrzesień	199,0	182,1	215,9	199,0 (182,1-215,9)	82,1%
październik	195,0	179,6	210,4	195,0 (179,6-210,4)	85,7%
listopad	172,4	160,2	184,6	172,4 (160,2-184,6)	89,6%
grudzień	165,2	152,5	177,9	165,2 (152,5-177,9)	88,0%

5. Podobną analizę wykonać dla roku 2021.

Zadanie 3. Plik (wypadki w Polsce od 1998 roku M.sta)

1. Wykonać prognozę punktową wraz z 95% przedziałem ufności liczby wypadków w Polsce dla roku 2020 za pomocą 3 modeli:
 - a. prostego modelu regresji
 - b. metody wskaźnikowej
 - c. metody jednoimiennych okresów.
2. Ze strony policja.pl/Statystyka → Ruch drogowy z raportu „Wypadki drogowe w 2020 roku” odczytać rzeczywistą liczbę wypadków w kolejnych miesiącach roku 2020 i na tej podstawie obliczyć błąd APE
3. Stwierdzić który model był najlepszy w prognozowaniu miesięcznej liczby wypadków w 2020 r w Polsce

Miesiąc	prognoza punktowa			Wypadki w Polsce 2020 Raport policji	APE		
	model liniowy	metoda wskaźnikowa	metoda trendów jednoimiennych okresów		model liniowy	metoda wskaźnikowa	metoda trendów jednoimiennych okresów
styczeń							
luty							
marzec							
kwiecień							
maj							
czerwiec							
lipiec							
sierpień							
wrzesień							
październik							
listopad							
grudzień							