Dodatek: Regresja Logistyczna

Dokumentacja techniczna

SPIS TREŚCI

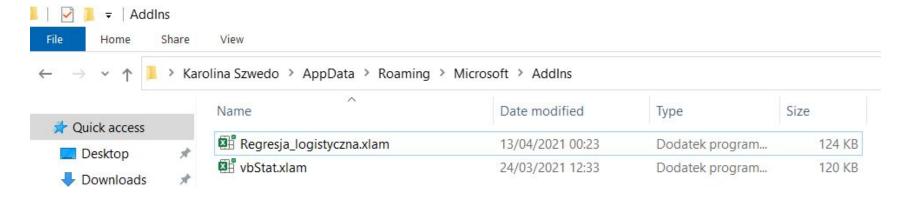
- 1. Instalacja dodatku
- 2. Dane testowe
- 3. Opis procesu przeprowadzania regresji logistycznej
 - I. Proces przeprowadzania regresji logistycznej;
 - II. Proces redukcji nieistotnych zmiennych z modelu;
 - III. Diagnostyka.
- 4. Opcje pomocy

INSTALACJA DODATKU cz. 1

Aby zainstalować plik z dodatkiem Regresja_logistyczna należy umieścić go w wybranej przez siebie lokalizacji.

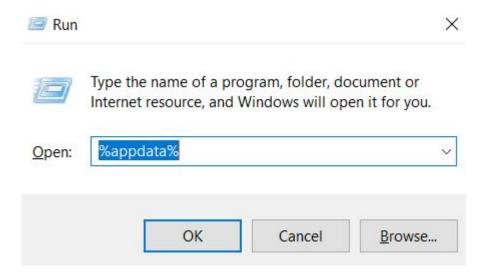
Lokalizacja preferowana przez system Windows to:

C:\Users\nazwa_użytkownika\AppData\Roaming\Microsoft\AddIns



INSTALACJA DODATKU cz.2

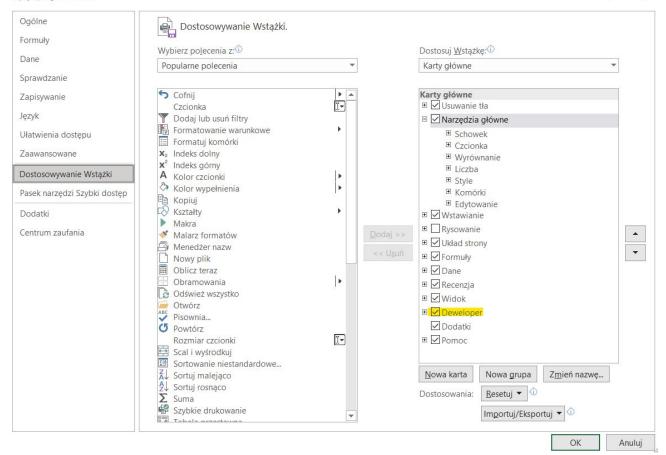
Aby dostać się do folderu AppData można skorzystać z narzędzia Uruchom lub wykorzystać skrót klawiszowy Znak Windowsa + R:



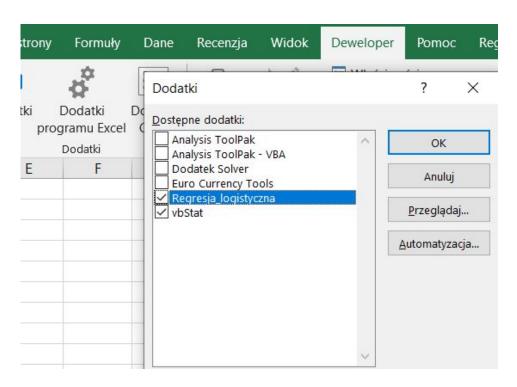
INSTALACJA DODATKU cz.3

Następnie, należy uruchomić program Microsoft Excel i uruchomić dodatek wybierając przycisk **Dodatki programu Excel** na karcie **Deweloper.** Jeśli karta Deweloper nie jest widoczna, należy ją uaktywnić wybierając kolejno: **Plik -> Opcje** -> **Dostosowywanie wstążki -> Deweloper:**

Opcje programu Excel



Po uaktywnieniu karty Deweloper należy wybrać Dodatki programu Excel, a następnie zaznaczyć dodatek Regresja_logistyczna:

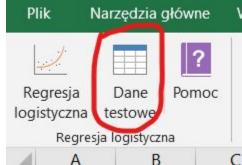


Jeżeli wszystkie kroki zostały wykonane poprawnie, dodana zostanie nowa karta:



DANE TESTOWE

- Dodatek umożliwia skorzystanie z danych testowych;
- Dane testowe są przykładowymi danymi instalowanymi razem z pakietem statystycznym STATA. Wykorzystano dane 1978 Automobile Data (auto.dta), które dotyczą obsługi technicznej samochodów osobowych i dostawczych.
- W celu dodania arkusza z danymi testowymi należy nacisnąć w przycisk Dane testowe w karcie Regresja;
- Po naciśnięciu przycisku nastąpi dodanie arkusza z danymi testowymi.



Arkusz z danymi do testowania.

4	resja tegen resja tegen resja tegen resja tegen resja tegen respective respec	Dane estowe ogistyczna B price 3291.00 3299.00 3667.00 3748.00 3799.00 3799.00	_	D trunk 17.00 9.00 7.00	E weight 2830 2110	F length	G turn	Н	I	J	K	L	М	N
2 3 4 5 6 7 8	A foreign 0 0 0 1 1 0 0 0 0	B price 3291.00 3299.00 3667.00 3748.00 3798.00	mpg 20.00 29.00 24.00 31.00	17.00 9.00 7.00	weight 2830	length	40.000		1	J	K	1	М	N
2 3 4 5 6 7 8	0 0 0 1 1 0 0	3291.00 3299.00 3667.00 3748.00 3798.00	20.00 29.00 24.00 31.00	17.00 9.00 7.00	2830		turn							I N
3 4 5 6 7 8	0 0 1 1 0	3299.00 3667.00 3748.00 3798.00	29.00 24.00 31.00	9.00 7.00		195		displacement	gear_ratio					
4 5 6 7 8	0 1 1 0 0	3667.00 3748.00 3798.00	24.00 31.00	7.00	2110	100	43	140	3.08					
5 6 7 8	1 1 0	3748.00 3798.00	31.00			163	34	231	2.93					
6 7 8	1 0 0	3798.00		0.00	2750	179	40	151	2.73					
7 8	0		25.00	9.00	2200	165	35	97	3.21					
8	0	2700 00	33.00	11.00	2050	164	36	97	3.81					
-		3/99.00	22.00	12.00	2640	168	35	121	3.08					
9		3829.00	22.00	9.00	2580	169	39	140	2.73					
	1	3895.00	26.00	10.00	1830	142	34	79	3.72					
10	0	3955.00	19.00	13.00	3430	197	43	250	2.56					
11	0	3984.00	30.00	8.00	2120	163	35	98	3.54					
12	1	3995.00	30.00	11.00	1980	154	33	86	3.73					
13	0	4010.00	18.00	17.00	3600	206	46	318	2.47					
14	0	4060.00	18.00	16.00	3330	201	44	225	3.23					
15	0	4082.00	19.00	13.00	3400	200	42	231	3.08					
16	0	4099.00	22.00	11.00	2930	186	40	121	3.58					
17	0	4172.00	24.00	7.00	2690	179	41	151	2.73					
18	0	4181.00	19.00	14.00	3370	200	43	231	3.08					
19	0	4187.00	21.00	10.00	2650	179	43	140	3.08					
20	0	4195.00	24.00	10.00	2730	180	40	151	2.73					
21	1	4296.00	21.00	16.00	2130	161	36	105	3.37					
22	0	4389.00	28.00	9.00	1800	147	33	98	3.15					
23	0	4424.00	19.00	13.00	3420	203	43	231	3.08					
24	0	4425.00	34.00	11.00	1800	157	37	86	2.97					
25	0	4453.00	26.00	10.00	2230	170	34	304	2.87					
26	0	4482.00	25.00	17.00	2200	165	36	105	3.37					
27	1	4499.00	28.00	5.00	1760	149	34	91	3.30					
28	0	4504.00	22.00	17.00	3180	193	31	200	2.73					
29	0	4516.00	18.00	15.00	3370	198	41	250	2.43					
30	1	4589.00	35.00	8.00	2020	165	32	85	3.70					
31	0	4647.00	28.00	11.00	3260	170	37	156	3.05					
22		Arkusz1	25.00	testowan	ia (+)	AFF	25	- 00	2.70					

OPIS PROCESU PRZEPROWADZANIA REGRESJI LOGISTYCZNEJ

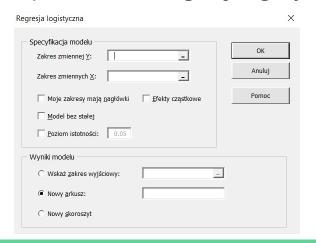
Proces przeprowadzania regresji liniowej składa się z kilku powiązanych z sobą kroków:

- Na początku zostaje przeprowadzana regresja logistyczna z wykorzystaniem algorytmu Newtona-Raphsona i Metody Największej Wiarygodności;
- 2. Potem w przypadku występowania zmiennych nieistotnych, można przeprowadzić redukcję zmiennych w modelu;
- 3. Następnie użytkownik może przeprowadzić diagnostykę na dobrze wyspecyfikowanym modelu (modelu po redukcji, jeśli zmienne w modelu były wcześniej nieistotne).

PROCES PRZEPROWADZANIA REGRESJI LOGISTYCZNEJ

 W celu przeprowadzenia regresji logistycznej należy wybrać na karcie Regresja przycisk Regresja logistyczna:

 Po naciśnięciu przycisku, powinno pokazać się następujące okno dialogowe służące do przeprowadzenia regresji logistycznej:



UWAGA: Aby przeprowadzić regresję logistyczną należy mieć włączoną opcję "Użyj separatorów systemowych".

Regresia

Pomoc

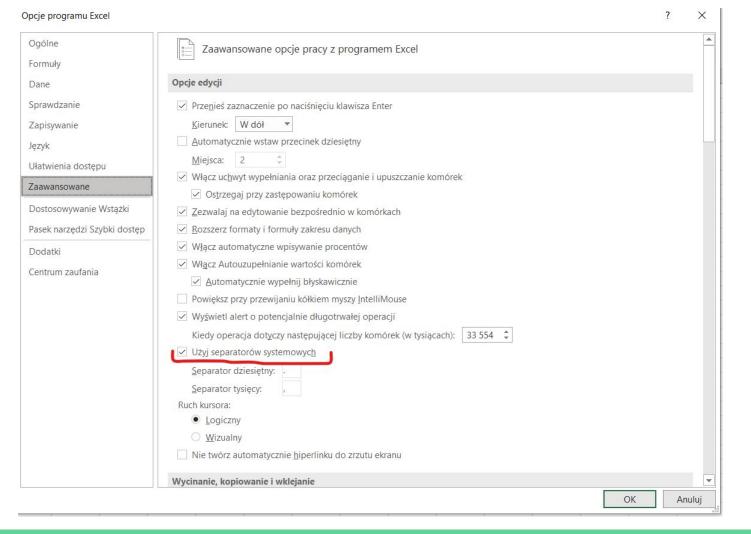
Dane

testowe

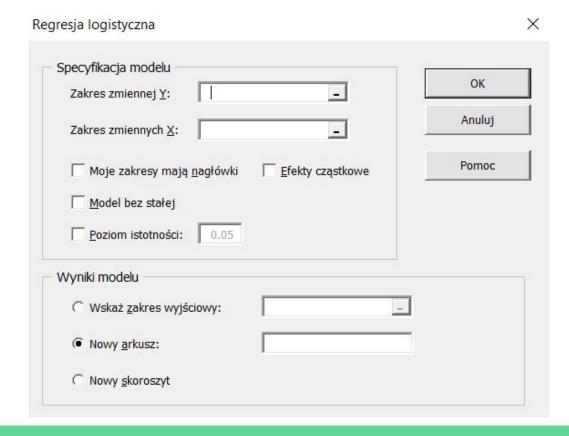
Regresja logistyczna

Aby włączyć opcję należy przejść do:

Plik -> Opcje -> Zaawansowane



PROCES PRZEPROWADZANIA REGRESJI LOGISTYCZNEJ OPIS OKNA DIALOGOWEGO 1. Zakres zmiennej Y: Należy wybrać zakres jednokolumno



- Zakres zmiennej Y: Należy wybrać zakres jednokolumnowy, tylko z wartościami 0-1. Zakres musi mieć przynajmniej 20 obserwacji;
- Zakres zmiennej X: Zakres może składać się z wielu kolumn. Liczba wierszy zakresu musi być taka sama jak liczba wierszy zakresu zmiennej Y. Zakres musi mieć przynajmniej 20 obserwacji i mieć tylko wartości liczbowe;
- Moje zakresy mają nagłówki: Jeśli ta opcja zostanie wybrana, w pierwszym wierszu zakresu zmiennej X i Y należy podać tytuły zmiennych;
- 4. **Model bez stałej:** Jeśli zostanie wybrana ta opcja, w modelu nie będzie stałej;
- Poziom istotności: Domyślnie poziom istotności wynosi 0,05 (czyli 5%). Można zmienić poziom istotności – możliwy zakres to (0;1);
- **6. Efekty cząstkowe:** Jeśli zostanie wybrana ta opcja, w wynikach regresji pojawią się obliczone efekty cząstkowe.
- Wskaż zakres wyjściowy: Po wyborze tej opcji należy wprowadzić jednokomórkowy zakres, w którym pojawią się wyniki przeprowadzonej regresji;
- Nowy arkusz: Wyniki regresji można umieścić w nowym arkuszu. W polu obok należy wprowadzić nazwę nowego arkusza;
- Nowy skoroszyt: Wyniki regresji można umieścić w nowym skoroszycie;
- OK: Po wyborze tej opcji, nastąpi sprawdzenie, czy wszystkie warunki zostały spełnione. Jeśli wszystko zostało dobrze wprowadzone, w określonym przez użytkownika miejscu pojawią się wyniki regresji logistycznej;
- 11. **Anuluj:** Wyjście z okna dialogowego. Regresja logistyczna nie zostanie wykonana;
- Pomoc: Wyświetlenie informacji o regresji w postaci pliku pdf znajdującego się w Internecie.

PROCES PRZEPROWADZANIA REGRESJI LOGISTYCZNEJ Przykładowe wyniki regresji logistycznej

REGRESJA LOGISTYCZNA	Liczba obserwacji =	74	
		LR chi2 (4) =	55.79
Liczba iteracji =	8	LR p-value =	0.0000
Logartytm funkcji wiarygodności =	-17.139712	Pseudo R2 =	0.6194

foreign	Coef.	Efekty cząstkowe	Std. Err.	Z	P> z
price	0.0009398	0.0000001	0.0003180	2.96	0.003
mpg	-0.1224238	-0.0000151	0.0949579	-1.29	0.197
trunk	-0.0312433	-0.0000038	0.1484538	-0.21	0.833
weight	-0.0068644	-0.0000008	0.0020218	-3.40	0.001
_cons	14.7596003	THE CONTRACTOR OF THE CONTRACT	5.6593421	2.61	0.009

^{*} pogrubione wartości p-value oznaczają zmienną istotną na poziomie istotności alfa = 5 %.

Interpretacja wyników

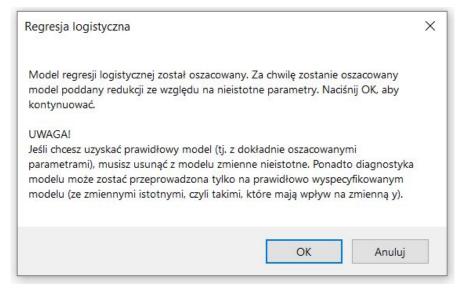
Na podstawie statystyki LR chi2 (4) oraz odpowiadającego jej p-value możemy stwierdzić, że zmienne w modelu są łącznie istotne. Pseudo R2 to R2 McFaddena, które nie ma interpretacji.

Interpretacja współczynników

W przypadku współczynników możemy interpretować jedynie znak, aby zinterpretować zmienne ilościowo należy wykorzystać efekty cząstkowe.

PROCES REDUKCJI NIEISTOTNYCH ZMIENNYCH Z MODELU

- Również po wybraniu **OK**, jeśli chociaż jedna zmienna jest istotna, powinien pojawić się poniższy komunikat:
- Należy wybrać OK w celu przeprowadzenia redukcji modelu ze względu na nieistotne parametry. W przeciwnym razie model będzie źle wyspecyfikowany, a przez to nie będzie można przeprowadzić na nim testów diagnostycznych



PROCES REDUKCJI NIEISTOTNYCH ZMIENNYCH Z MODELU: Przykładowe wyniki regresji po redukcji nieistotnych zmiennych

REGRESJA LOGISTYCZNA	Liczba obserwacji = 74
(PO ZREDUKOWANIU)	LR chi2 (2) = 54.11
Liczba iteracji = 7	LR p-value = 0.0000
Logartytm funkcji wiarygodności = -17.976341	Pseudo R2 = 0.6008

foreign	Coef.	Efekty cząstkowe	Std. Err.	Z	P> z
price	0.0009296	0.0000001	0.0002999	3.10	0.002
weight	-0.0058785	-0.0000009	0.0016986	-3.46	0.001
_cons	9.0004726		2.6275769	3.43	0.001

^{*} pogrubione wartości p-value oznaczają zmienną istotną na poziomie istotności alfa = 5 %.

Interpretacja wyników

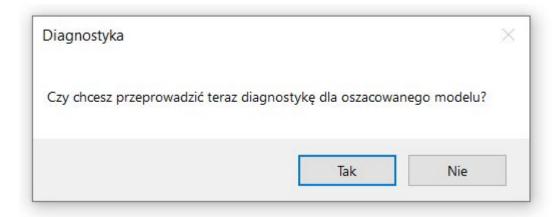
Na podstawie statystyki LR chi2 (2) oraz odpowiadającego jej p-value możemy stwierdzić, że zmienne w modelu są łącznie istotne. Pseudo R2 to R2 McFaddena, które nie ma interpretacji.

Interpretacja współczynników

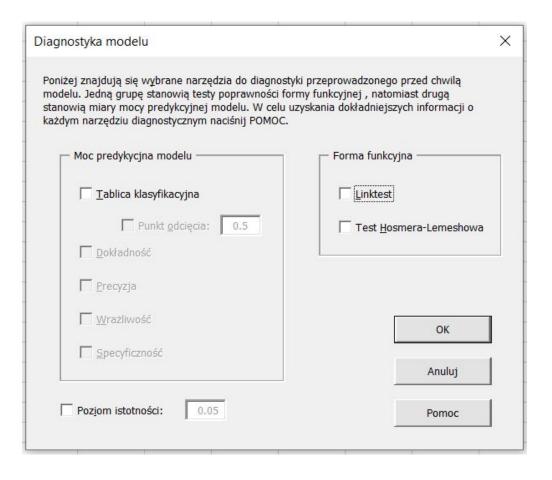
W przypadku współczynników możemy interpretować jedynie znak, aby zinterpretować zmienne ilościowo należy wykorzystać efekty cząstkowe.

DIAGNOSTYKA

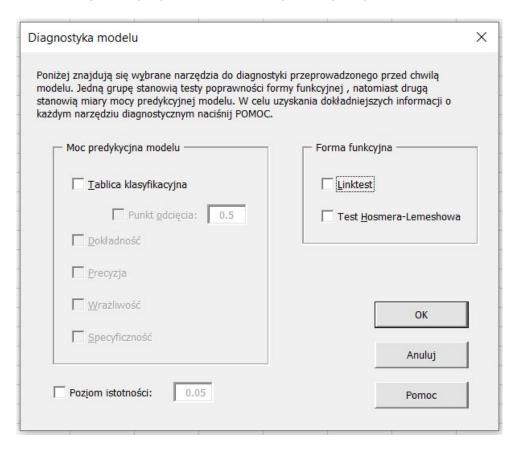
Po przeprowadzeniu redukcji (lub w przypadku, kiedy już od początku wszystkie zmienne były istotne) pojawi się następujący komunikat:



Po naciśnięciu TAK pojawi się poniższe okno dialogowe:



DIAGNOSTYKA: OPIS OKNA DIALOGOWEGO



- Forma funkcyjna: Do wyboru są dwa testy diagnostyczne (Linktest i Test Hosmera-Lemeshowa). Możesz wybrać zero, jeden lub oba;
- 2. Moc predykcyjna modelu: Po wyborze tablicy klasyfikacyjnej pojawią się do wyboru dodatkowe opcje: Dokładność, Precyzja, Wrażliwość i Specyficzność. Po zaznaczeniu opcji Punkt odcięcia możliwy jest wybór wartości punktu odcięcia z zakresu (0;1). Domyślnie punkt odcięcia wynosi 0,5 (czyli 50%).;
- Poziom istotności: Domyślnie poziom istotności wynosi 0,05 (czyli 5%). Można zmienić poziom istotności – możliwy zakres to (0;1);
- OK: Po wyborze tej opcji, nastąpi sprawdzenie, czy wszystkie warunki zostały spełnione. Jeśli wszystko zostało dobrze wprowadzone, pojawią się wyniki wybranych testów diagnostycznych;
- 5. **Anuluj:** Wyjście z okna dialogowego. Wybierz tę opcję, jeśli nie chcesz wykonywać testów diagnostycznych;
- 6. **Pomoc:** Wyświetlenie pomocy do testów diagnostycznych w postaci pliku pdf znajdującego się w Internecie.

DIAGNOSTYKA: PRZYKŁADOWE WYNIKI cz. 1

LINKTEST

	Coef.	Std. Err.	Z	P> z
_hat	1.1604280	0.5049115	2.30	0.022
	-0.2200453		-1.08	0.282
_cons	0.3174269	0.5130970	0.62	0.536

TEST HOSMERA-LEMESHOWA

Statystyka 11.80633

P-value **0.160056**

Interpretacja

Forma funkcyjna modelu jest poprawna.

Interpretacja

Zmienna _hatsq jest nieistotna w związku z czym forma funkcyjna modelu jest prawidłowa.

DIAGNOSTYKA: PRZYKŁADOWE WYNIKI cz. 2

TABLICA KLASYFIKACYJNA

	Prawdziwe wartości			
Predykcje	1	0		
1	20	6		
0	0	46		

Dokładność: 92% Precyzja: 77%

Wrażliwość: 100%

Specyficzność: 88%

Interpetacja wyniku

Model w 92% poprawnie zaklasyfikował obserwacje.

Model w 77% poprawnie zaklasyfikował sukcesy (jedynki).

Model w 100% poprawnie zaklasyfikował sukcesy (jedynki) tam gdzie one występowały.

Model w 88% poprawnie zaklasyfikował porażki (zera) tam gdzie one występowały.

OPCJE POMOCY

Na każdym z formularzy znajduje się przycisk **POMOC.** W przypadku formularza regresji wyświetlają się informacje o regresji, natomiast w przypadku formularza diagnostyki wyświetlają się informacje dotyczące testów diagnostycznych.

Aby wyświetlić niniejszy plik należy nacisnąć przycisk Pomoc na karcie Regresja.

