



**Projekt zaliczeniowy z przedmiotu
Statystyczna Analiza Danych**

2022/2023

Aleksandra Kilijańska

Spis treści

1. Baza danych.....	3
1.1. Utworzenie zbioru danych.....	3
1.1.1. Import danych z wybranych baz.....	3
1.1.2. Scalanie danych.....	4
1.2. Podstawowe statystyki danych.....	4
1.3. Histogramy do wykresu ramka-wąsy.....	4
1.3.1. NO2 Wodzisław Śląski.....	5
1.3.2. SO2 Wodzisław Śląski.....	6
1.3.3. NO2 i SO2 Dąbrowa Górnicza.....	6
1.3.4. Podsumowanie.....	6
1.4. Wykresy ramka-wąsy.....	7
1.5. Analiza przypadków odstających, wpływ eliminacji przypadków odstających na podstawowe statystyki.....	8
1.5.1. Wyczyszczone dane.....	8
1.5.2. Wykresy ramka-wąsy dla oczyszczonych danych.....	9
1.5.3. Podstawowe statystyki dla NO2 i SO2.....	10
2. Testy parametryczne i nieparametryczne.....	10
2.1. Test istotności różnic między grupami (założenia).....	10
2.1.1. Test dla NO2.....	10
2.2. Test istotności różnic między zmiennymi.....	10
2.2.1. Test dla NO2.....	10
2.3. Test normalności rozkładu.....	11
2.3.1. Test Shapiro-Wilka.....	11
2.3.1.1. NO2 i SO2 Wodzisław Śląski.....	11
2.3.1.2. Pozostałe zmienne.....	11
2.3.2. Test Kołmogorowa-Smirnowa.....	11
2.3.2.1. NO2 Wodzisław Śląski.....	11
2.3.2.2. Pozostałe zmienne.....	11
2.3.3. Podsumowanie.....	12
2.4. Przedział ufności dla średniej.....	12
2.5. Test na różnicę między dwoma wskaźnikami struktury.....	12
2.5.1. Test jednostronny i dwustronny.....	13
2.6. Liczność próby.....	13
3. Korelacja i regresja.....	15
3.1. Wyznaczenie i analiza współczynników Pearsona i Spearmana.....	15
3.1.1. Test istotności współczynnika korelacji.....	17
3.1.1.1. Test istotności współczynnika Pearsona dla NO2 i SO2 w Wodzisławiu.....	17
3.1.1.2. Test istotności współczynnika Spearmana dla NO2 i SO2 w Wodzisławiu.....	17
3.2. Regresja liniowa, budowa i weryfikacja.....	17
3.2.1. Regresja liniowa dla NO2 i SO2 w Wodzisławiu Śląskim.....	17
3.2.1. Regresja wieloraka.....	22
4. Analiza szeregów czasowych.....	25
4.1. Trend szeregu czasowego.....	25
4.1.1. Wygładzenie.....	26
4.2. Autokorelacja szeregu.....	26
4.2.1. Dla danych bez wygładzenia.....	26
4.2.2. Dla danych z wygładzeniem.....	27
4.3. Analiza Fouriera szeregu czasowego.....	29
4.3.1. Periodogram.....	29
4.4. Współczynnik sezonowości.....	30

1. Baza danych

Do wyboru zostały udostępnione bazy danych kilkunastu miast z pomiarami stężenia pierwiastków w powietrzu oraz meteorologicznymi. Wartości mierzone były w różnych przedziałach czasu oraz w różny sposób. W moim przypadku wybrałam Wodzisław Śląski oraz Dąbrowę Górniczą do analizy, ponieważ pomiary zaczynały się w tym samym czasie i kończyły prawie tak samo (małe straty danych przy scalaniu). Mierzone pierwiastki pokrywały się w większości oraz na pierwszy rzut oka było mniej pustych wartości niż w przypadku innych miast.

1.1. Utworzenie zbioru danych

1.1.1. Import danych z wybranych baz danych

Zaimportowanie danych dla Wodzisławia Śląskiego. Wybranie pierwszego wiersza jako nazwy zmiennych (nazwy pierwiastków oraz pomiary meteorologiczne), a pierwszej kolumny jako nazwy przypadków (daty). Do zmiennych dodany został sufix _WS, aby rozróżniać dane po scaleniu. Arkusz zawiera 14 zmiennych oraz 2917 przypadków.

Tabela 1. Zaimportowane dane dla Wodzisławia Śląskiego

E:\kili\Miasta\Wodzisław Główny\24h 2007-2014.xlsx : Sheet1											
Data	CO_WS (1g)	NO_WS (1g)	NO2_WS (1g)	Nox_WS (1g)	O3_WS (1g)	PM10_WS (1g)	SO2_WS (1g)	ciśnienie_WS (1g)	kier_WS (1g)	predk_WS (1g)	predk2_WS (1g)
2007-01-01	0,44043478		21,17391		27,347826	33,333333	20,727272				
2007-01-02	0,61652173		31,478260		23,478260	62,458333	23,454545				
2007-01-03	0,62347826		26,652173		19	60,833333	13,909090				
2007-01-04	0,38739130		20,913043		23,086956	39,416666	17,227272				
2007-01-05	0,44260869		24,521739		19,826087	42,708333	19,363636				
2007-01-06	0,38913043		30,565217		10,565217	37,083333	24,090909				
2007-01-07	0,39869565		16,217391		22,652173	29,458333	10,863636				
2007-01-08	0,61043478		29,304347		15,478260	64,875	18,363636				
2007-01-09	0,36434782		31,434782		11,434782	37,583333	20,227272				
2007-01-10	0,36130434		30,217391		13,608695	35,791666	26,869565				
2007-01-11	0,27521739		14,217391		37,130434	19,25	12,636363				
2007-01-12	0,33428571		11,095238		46,904761	32,863636	9,7				
2007-01-13	0,37043478		15,304347		32,956521	31,416666	9,136363				
2007-01-14	0,4017391		14,478260		37,347826	34,75	13,681818				
2007-01-15	0,51391304		32,739130		22,304347	55,083333	27,772727				
2007-01-16	0,77666666		35,476190		16,565217	83,5	28,190476				
2007-01-17	0,55608695		38,434782		6,8260869	66,791666					
2007-01-18	0,33652173		20,739130		30,565217	27,791666					
2007-01-19	0,25727272		7,6956521		54,434782	16,583333					
2007-01-20	0,34260869		21,260869		26,173913	32,041666					
2007-01-21	0,25409090		8,4545454		47,652173	19,375					
2007-01-22	0,37043478		20,347826		38,318181	35,291666					
2007-01-23	0,40782608		19,913043		28,739130	34,125					

Zaimportowanie danych dla Dąbrowy Górniczej. Nazwy zmiennych i przypadków tak samo jak w przypadku Wodzisławia Śląskiego. Do zmiennych dodany został sufix _DG. Arkusz zawiera 20 zmiennych oraz 2882 przypadków.

Tabela 2. Zaimportowane dane dla Dąbrowy Górniczej

E:\kili\Miasta\Dąbrowa G. 24h 2007-2014.xlsx : Sheet1											
Data	As(PM10)_DG (24h)	BaP(PM10)_DG (24h)	C6H6_DG (1g)	CO_DG (1g)	Cd(PM10)_DG (1g)	NO_DG (1g)	NO2_DG (1g)	Nox_DG (1g)	Ni(PM10)_DG (24h)	O3_DG (1g)	PM10_DG (1g)
2007-01-01				0,44521739			20,695652				45,086956
2007-01-02				0,55086956			40,260869				33,521739
2007-01-03				0,55434782			37,478260				26,956521
2007-01-04				0,50434782			35,826087				29,217391
2007-01-05				0,48521739			37,043478				29,043478
2007-01-06				0,61			42,608695				16,652173
2007-01-07				0,45391304			21,391304				38,347826
2007-01-08				0,83217391			52,478260				16,739130
2007-01-09				0,55			43				20,652173
2007-01-10				0,59304347			49				16,956521
2007-01-11				0,37521739			27,652173				44,217391
2007-01-12				0,42142857			30,190476				45,095238
2007-01-13				0,39652173			24,347826				40,956521
2007-01-14	0		7,9	0,42521739	0		28,956521		0		39,75
2007-01-15				0,60652173			55				24,95238
2007-01-16				0,87347826			59,217391				18,521739
2007-01-17				0,94478260			63,130434				11,043478
2007-01-18				0,38434782			25,043478				48,565217
2007-01-19				0,33565217			24,652173				55,608695
2007-01-20				0,45521739			28,173913				39,869565
2007-01-21				0,38863636			21,5				57,304347
2007-01-22				0,50913043			42,086956				35,545454
2007-01-23				0,48173913			23,608695				38,695652

1.1.2. Scalenie arkuszy danych

Połączenie powyższych arkuszy w jeden według nazw przypadków. W przypadku obu plików usunięte będą niezgodne przypadki jeśli wystąpi niedopasowanie. Scalony arkusz zawiera 34 zmienne i 2877 przypadków, co oznacza, że przypadków dla Wodzisławia jest o 40 mniej, a dla Dąbrowy o 5, więc nie są to duże straty przy łączeniu arkuszy.

Tabela 3. Scalone dane

Data	E:\kin\Miasta\Wodzisław Galczyńskiego 24h 2007-2014.xlsx : Sheet1									
	1	2	3	4	5	6	7	8	9	10
	CO WS (1g)	NO WS (1g)	NO2 WS (1g)	Nox WS (1g)	O3 WS (1g)	PM10 WS (1g)	SO2 WS (1g)	cisnienie WS (1g)	kier WS (1g)	predk WS (1g)
2007-01-01	0,440434783		21,173913		27,3478261	33,3333333	20,7272727			
2007-01-02	0,616521739		31,4782609		23,4782609	62,4583333	23,4545455			
2007-01-03	0,623478261		26,6521739		19	60,8333333	13,9090909			
2007-01-04	0,387391304		20,9130435		23,0869565	39,4166667	17,2272727			
2007-01-05	0,442608696		24,5217391		19,826087	42,7083333	19,3636364			
2007-01-06	0,389130435		30,5652174		10,5652174	37,0833333	24,0909091			
2007-01-07	0,398695652		16,2173913		22,6521739	29,4583333	10,8636364			
2007-01-08	0,610434783		29,3043478		15,4782609	64,875	18,3636364			
2007-01-09	0,364347826		31,4347826		11,4347826	37,5833333	20,2272727			
2007-01-10	0,361304348		30,2173913		13,6086957	35,7916667	26,8695652			
2007-01-11	0,275217391		14,2173913		37,1304348	19,25	12,6363636			
2007-01-12	0,334285714		11,0952381		46,9047619	32,8636364	9,7			
2007-01-13	0,370434783		15,3043478		32,9565217	31,4166667	9,13636364			
2007-01-14	0,40173913		14,4782609		37,3478261	34,75	13,6818182			
2007-01-15	0,513913043		32,7391304		22,3043478	55,0833333	27,7727273			
2007-01-16	0,776666667		35,4761905		16,5652174	83,5	28,1904762			
2007-01-17	0,556086957		38,4347826		6,82608696	66,7916667				
2007-01-18	0,336521739		20,7391304		30,5652174	27,7916667				
2007-01-19	0,257272727		7,69565217		54,4347826	16,5833333				
2007-01-20	0,342608696		21,2608696		26,173913	32,0416667				
2007-01-21	0,254090909		8,45454546		47,6521739	19,375				
2007-01-22	0,370434783		20,3478261		38,3181818	35,2916667				
2007-01-23	0,407826087		19,9130435		28,7391304	34,125				
2007-01-24	0,477391304		20,5652174		19,1304348	48,2083333				
2007-01-25	0,416521739		19,3913043		26,3478261	47,9166667				
2007-01-26	0,754347826		27,7826087		31,826087	79,2083333				
2007-01-27	0,400869565		16,7391304		38,5652174	34,875				
2007-01-28	0,465217391		17,826087		34,4347826	31,5833333				
2007-01-29	0,447826087		13,3043478		37,6956522	40,2916667				
2007-01-30	0,493809524		15,9545455		28,7826087	38,1666667				
2007-01-31	0,478636364		21,3636364		24,0869565	37,2083333				

1.2. Podstawowe statystyki danych

Wykonanie podstawowych statystyk, w której zawiera się Nważnych, średnia, mediana, minimum, maksimum, wariancja, odchylenie standardowe. Pomiary te wykonane zostały dla dwutlenku azotu oraz dwutlenku siarki dla obu miast.

Tabela 4. Statystyki opisowe dla SO2 i NO2 w Wodzisławiu Śląskim oraz Dąbrowie Górniczej

Zmienna	Statystyki opisowe (NOWE)						
	Nważnych	Średnia	Mediana	Minimum	Maksimum	Wariancja	Odch.std
NO2_WS (1g)	2755	23,32167	20,65217	3,666667	100,7826	144,7365	12,03065
SO2_WS (1g)	2793	17,74780	12,09091	1,272727	143,5455	277,8588	16,66910
NO2_DG (1g)	2712	28,28514	26,00000	4,130435	96,8261	175,6563	13,25354
SO2_DG (1g)	2748	14,58237	10,51087	1,238095	139,1818	162,1818	12,73506

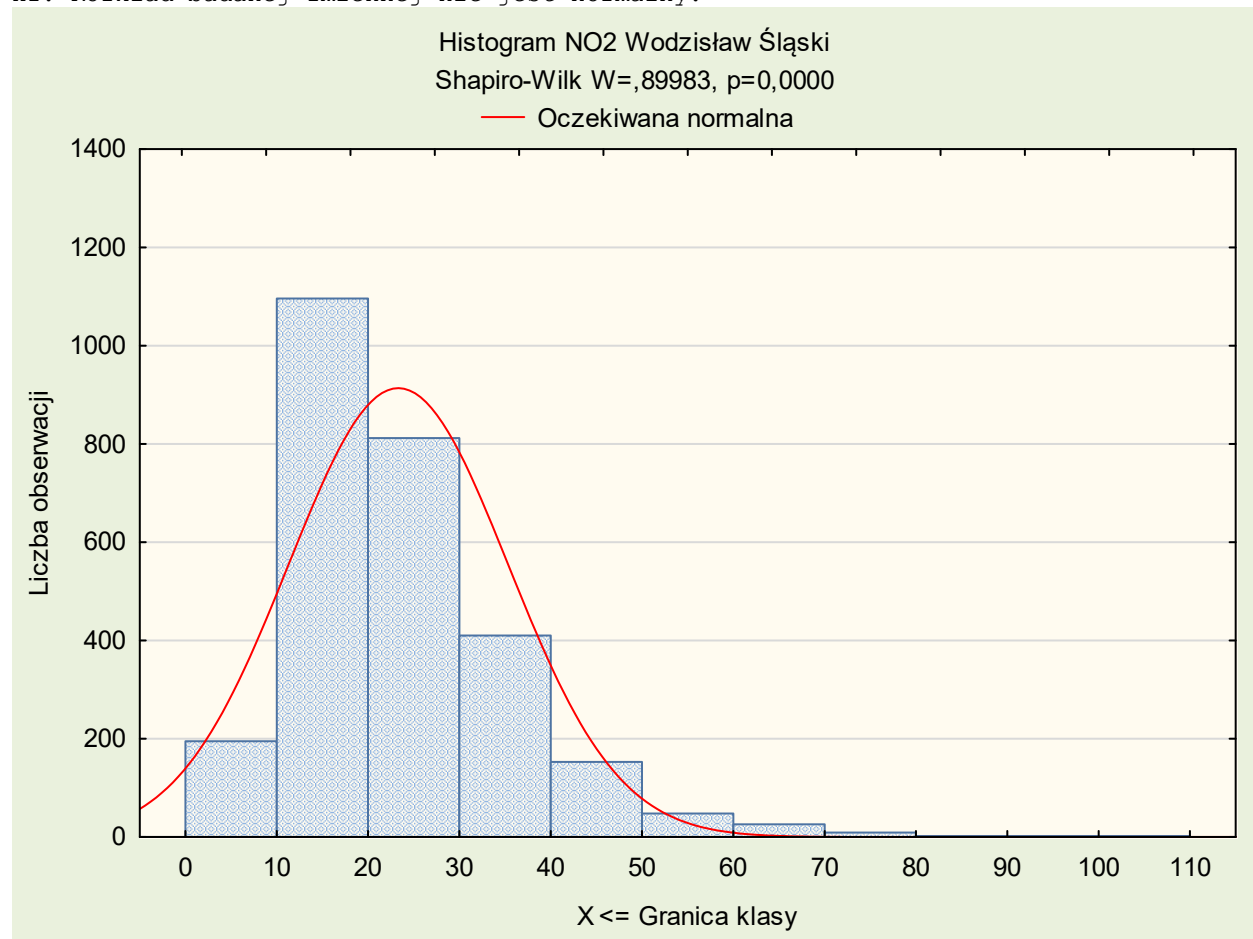
1.3. Histogramy do wykresu ramka-wąsy

Sprawdzenie normalności rozkładu dla wybranych zmiennych, aby zdecydować jaką wartość wybrać do punktu środkowego przy wykresie ramka wąsy.

1.3.1. NO2 Wodzisław Śląski

H0: Rozkład badanej zmiennej jest normalny.

H1: Rozkład badanej zmiennej nie jest normalny.



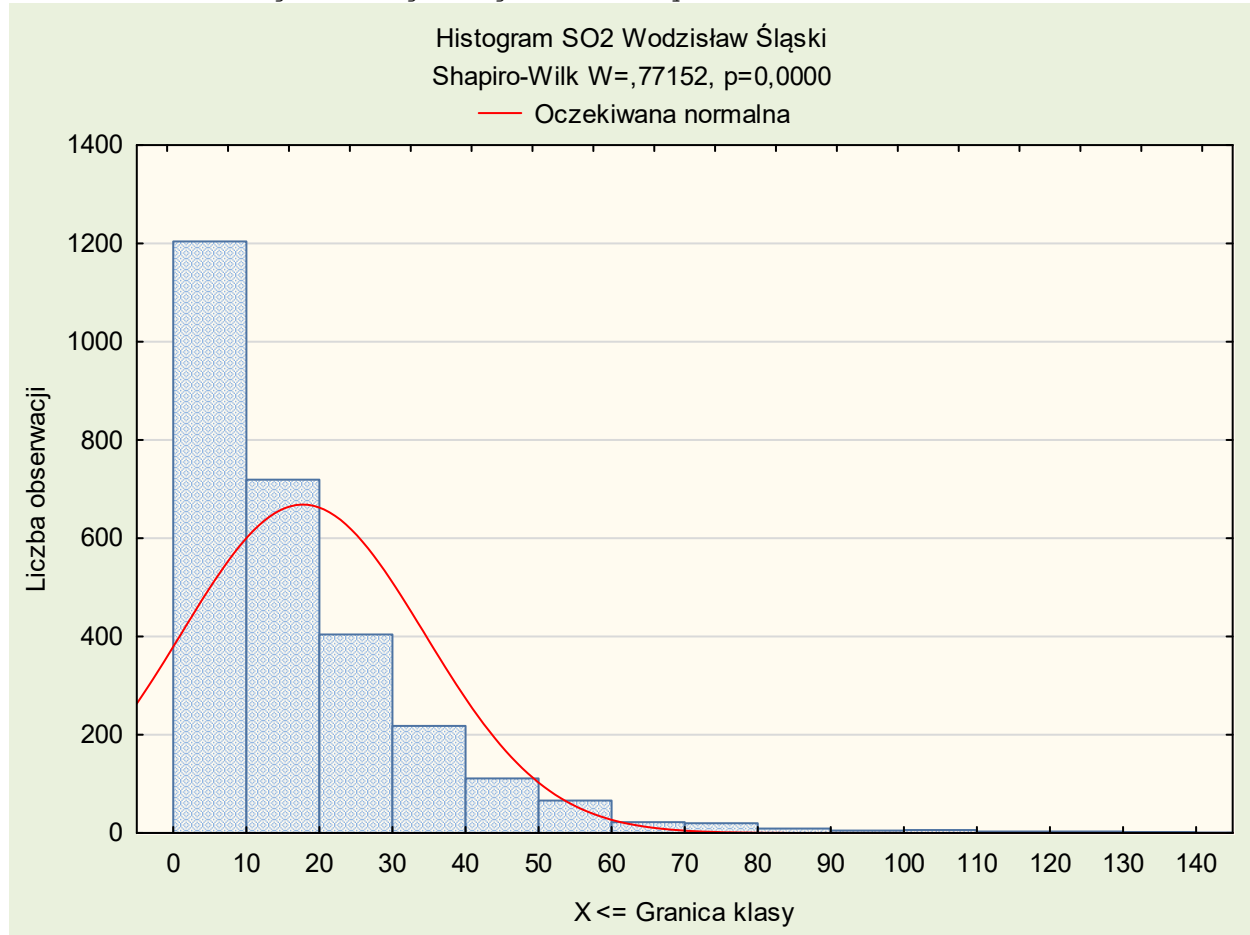
Rys. 1. Histogram z testem Shaphiro Wilka dla NO2 Wodzisław Śląski

W teście Shapiro-Wilka wartość p wynosi 0, czyli jest mniejsza od 0.05. Zatem możemy odrzucić hipotezę zerową na rzecz alternatywnej. Co oznacza, że zmienna posiada rozkład odbiegający od krzywej Gaussa.

1.3.2. SO2 Wodzisław Śląski

H0: Rozkład badanej zmiennej jest normalny.

H1: Rozkład badanej zmiennej nie jest normalny.



Rys. 2. Histogram z testem Shaphiro Wilka dla SO2 Wodzisław Śląski

W teście Shapiro-Wilka wartość p wynosi 0, czyli jest mniejsza od 0.05. Zatem możemy odrzucić hipotezę zerową na rzecz alternatywnej. Co oznacza, że zmienna posiada rozkład odbiegający od krzywej Gaussa.

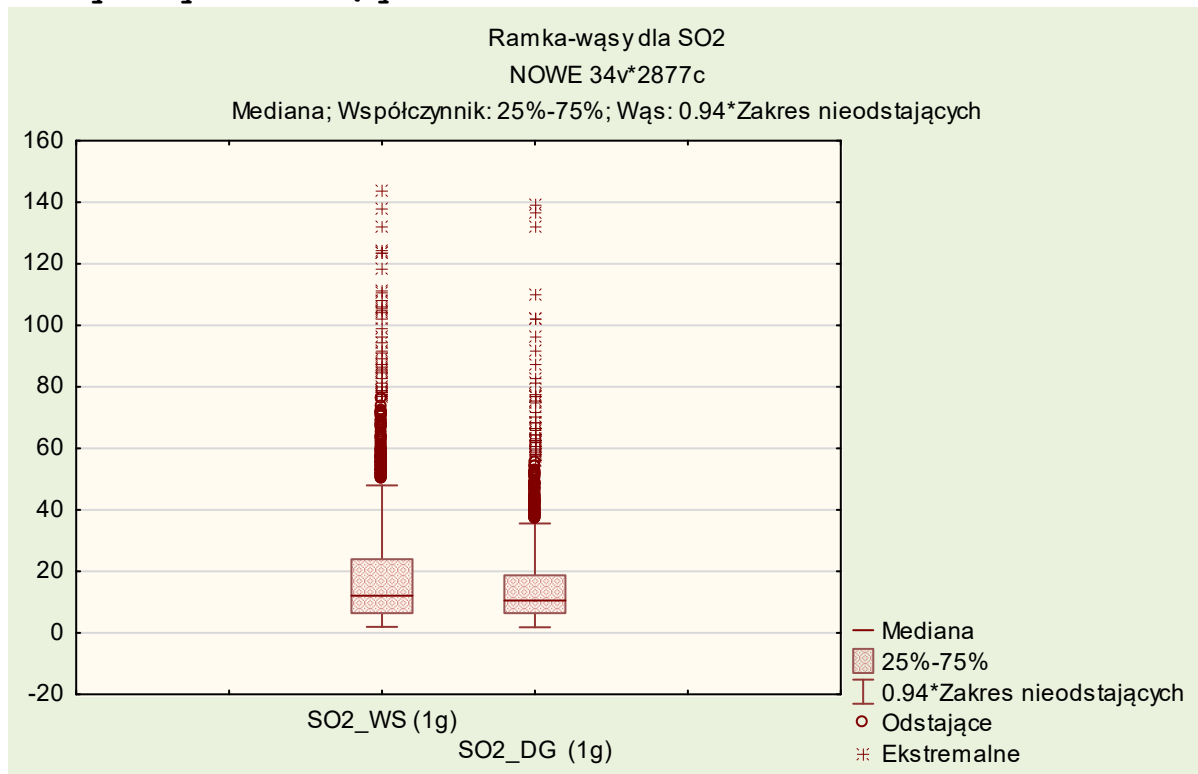
1.3.3. NO2 i SO2 Dąbrowa Górnicza

W przypadku tych samych pierwiastków dla Dąbrowy Górniczej również p wynosi 0, co oznacza że nie ma rozkładu normalnego.

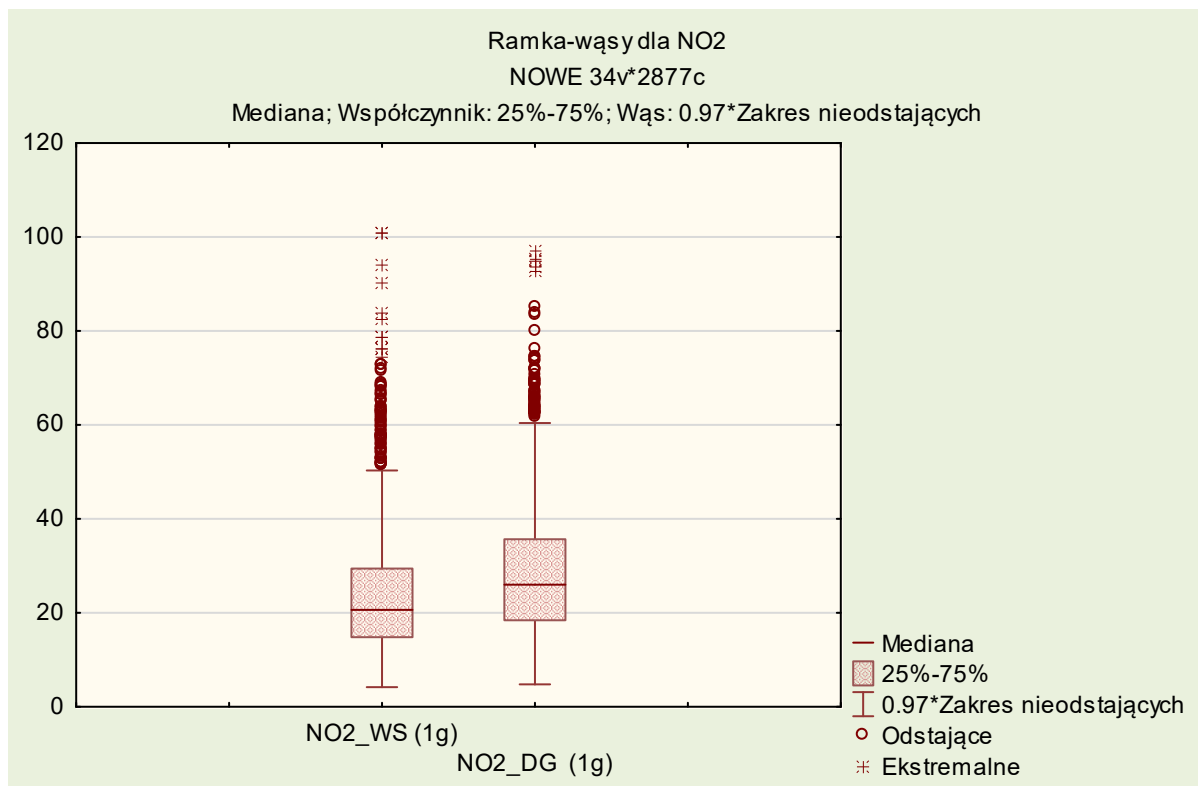
1.3.4. Podsumowanie

W żadnym z powyższych histogramów poziom istotności alfa wynoszącego 0,05 nie został przekroczony, co oznacza że nigdzie nie ma rozkładu normalnego. Możemy również zobaczyć to po samym wykresie. Zatem do wykresu ramka-wąsy jako punkt środkowy brana jest mediana.

1.4. Wykresy ramka-wąsy



Rys. 3. Wykres ramka-wąsy dla SO₂ Wodzisław Śląski i Dąbrowa Górnicza



Rys. 3. Wykres ramka-wąsy dla NO₂ Wodzisław Śląski i Dąbrowa Górnicza

1.5. Analiza przypadków odstających, wpływ eliminacji przypadków odstających na podstawowe statystyki

Zamiana typu zmiennej dla predk2_WS oraz opad_atm_DG(1g) na podwójnej precyzji, ponieważ zostały uznane przez program jako tekst.

Zmienianie wartości ekstremalnych na wartości graniczne za pomocą testu Tukeya dwustronnego. Nie wymaga on rozkładu normalnego, którego przeanalizowane wyżej zmienne nie posiadają (jak i inne zmienne). Jako parametr została wybrana 3, ponieważ według klasycznego wykresu ramka-wasy od tego momentu zaczynają się wartości ekstremalne. Wartości zostały zastąpione granicznymi, ponieważ w przypadku pomiarów stężenia pierwiastków w powietrzu oraz pomiarów meteorologicznych mogą nastąpić wszelkie anomalie, spowodowane przykładowo jakimś wybuchem.

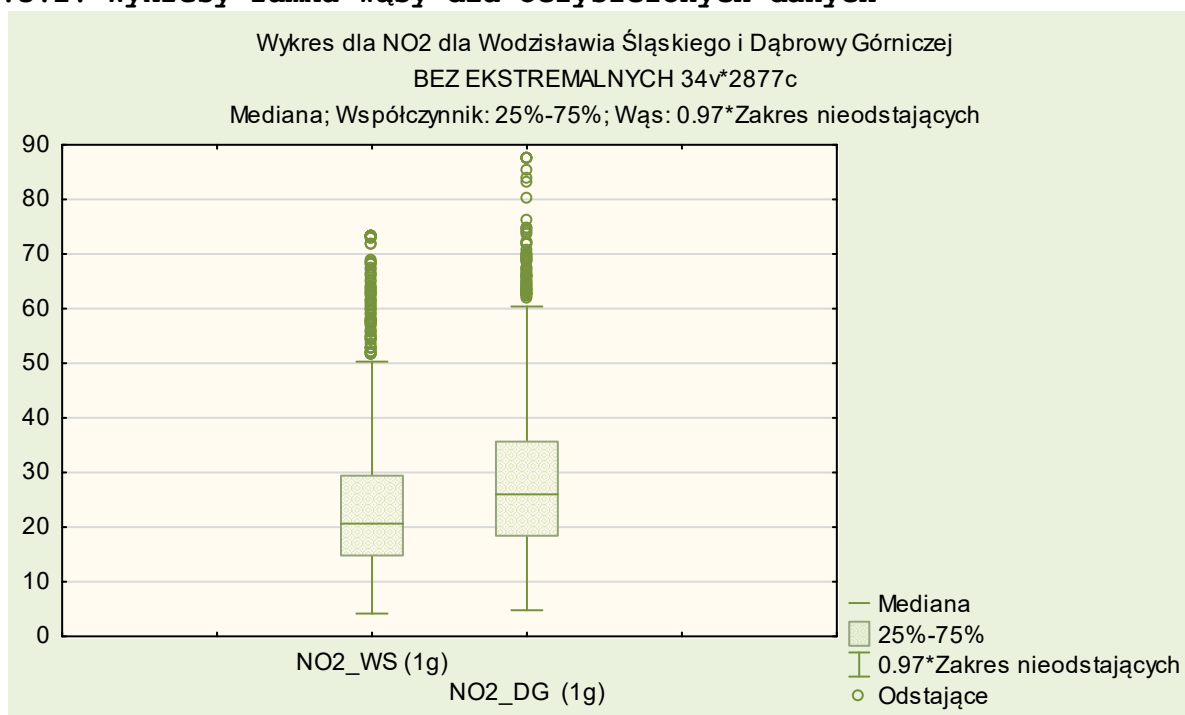
1.5.1. Wyczyszczone dane

Tabela 5. Dane z zamienionymi wartościami ekstremalnymi

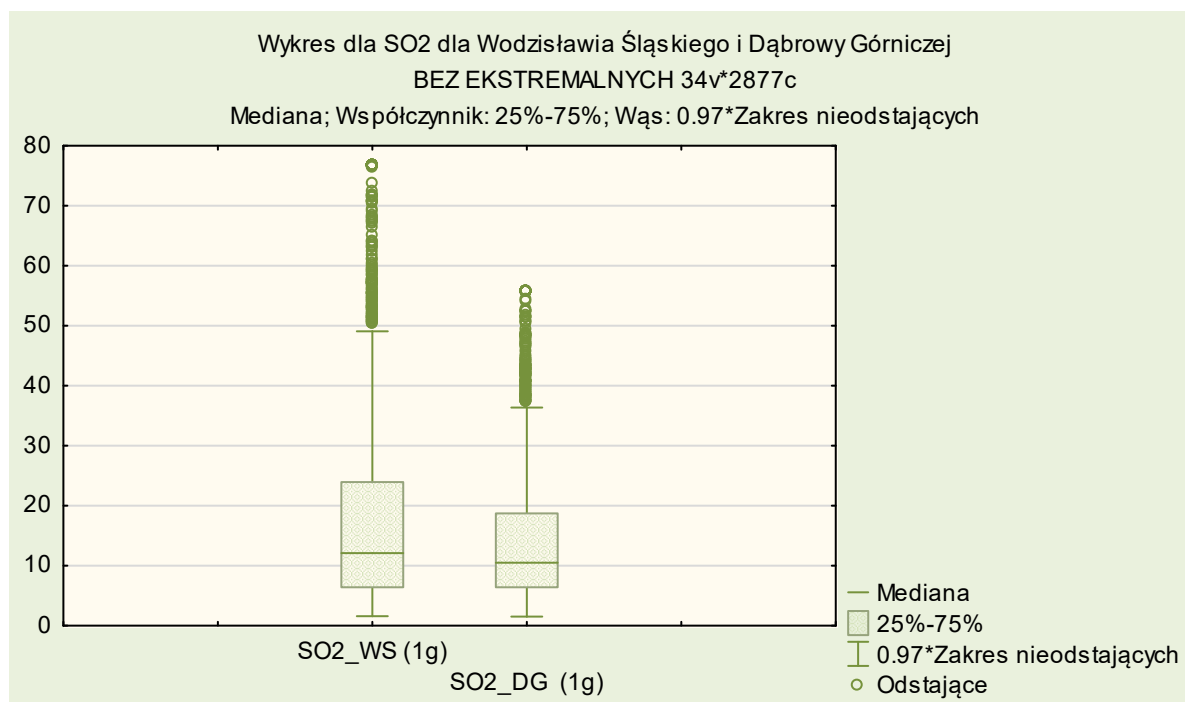
Data	E:\kili\Miasta\Wodzisław Galczyńskiego 24h 2007-2014.xlsx : Sheet1									
	1	2	3	4	5	6	7	8	9	10
	CO WS (1g)	NO WS (1g)	NO2 WS (1g)	Nox WS (1g)	O3 WS (1g)	PM10 WS (1g)	SO2 WS (1g)	cisnienie WS (1g)	kier WS (1g)	predk WS (1g)
2007-01-01	0,440434783		21,173913		27,3478261	33,3333333	20,7272727			
2007-01-02	0,616521739		31,4782609		23,4782609	62,4583333	23,4545455			
2007-01-03	0,623478261		26,6521739		19	60,8333333	13,9090909			
2007-01-04	0,387391304		20,9130435		23,0869565	39,4166667	17,2272727			
2007-01-05	0,442608696		24,5217391		19,826087	42,7083333	19,3636364			
2007-01-06	0,389130435		30,5652174		10,5652174	37,0833333	24,0909091			
2007-01-07	0,398695652		16,2173913		22,6521739	29,4583333	10,8636364			
2007-01-08	0,610434783		29,3043478		15,4782609	64,875	18,3636364			
2007-01-09	0,364347826		31,4347826		11,4347826	37,5833333	20,2272727			
2007-01-10	0,361304348		30,2173913		13,6086957	35,7916667	26,8695652			
2007-01-11	0,275217391		14,2173913		37,1304348	19,25	12,6363636			
2007-01-12	0,334285714		11,0952381		46,9047619	32,8636364	9,7			
2007-01-13	0,370434783		15,3043478		32,9565217	31,4166667	9,13636364			
2007-01-14	0,40173913		14,4782609		37,3478261	34,75	13,6818182			
2007-01-15	0,513913043		32,7391304		22,3043478	55,0833333	27,7727273			
2007-01-16	0,776666667		35,4761905		16,5652174	83,5	28,1904762			
2007-01-17	0,556086957		38,4347826		6,82608696	66,7916667				
2007-01-18	0,336521739		20,7391304		30,5652174	27,7916667				
2007-01-19	0,257272727		7,69565217		54,4347826	16,5833333				
2007-01-20	0,342608696		21,2608696		26,173913	32,0416667				
2007-01-21	0,254090909		8,45454546		47,6521739	19,375				
2007-01-22	0,370434783		20,3478261		38,3181818	35,2916667				
2007-01-23	0,407826087		19,9130435		28,7391304	34,125				
2007-01-24	0,477391304		20,5652174		19,1304348	48,2083333				
2007-01-25	0,416521739		19,3913043		26,3478261	47,9166667				
2007-01-26	0,754347826		27,7826087		31,826087	79,2083333				
2007-01-27	0,400869565		16,7391304		38,5652174	34,875				
2007-01-28	0,465217391		17,826087		34,4347826	31,5833333				
2007-01-29	0,447826087		13,3043478		37,6956522	40,2916667				
2007-01-30	0,493809524		15,9545455		28,7826087	38,1666667				
2007-01-31	0,478636364		21,3636364		24,0869565	37,2083333				

Przykładowo jedna z największych wartości NO2 dla Wodzisławia Śląskiego z 100,695652 zamieniła się na 73,260869565.

1.5.2. Wykresy ramka-wąsy dla oczyszczonych danych



Rys. 5. Wykres ramka-wąsy dla NO₂ Wodzisław Śląski i Dąbrowa Górnicza - wyczyszczone dane



Rys. 6. Wykres ramka-wąsy dla SO₂ Wodzisław Śląski i Dąbrowa Górnicza - wyczyszczone dane

Na powyższych wykresach widać brak zmiennych w porównaniu do poprzednich wykresów ramka-wąsy.

1.5.3. Podstawowe statystyki dla NO2 i SO2

Tabela 6. Statystyki opisowe dla SO₂ i NO₂ w Wodzisławiu Śląskim oraz Dąbrowie Górniczej - wyczyszczone dane

	Statystyki opisowe (BEZ EKSTREMALNYCH)						
Zmienna	Nważnych	Średnia	Mediana	Minimum	Maksimum	Wariancja	Odch.std
NO2_WS (1g)	2755	23,27433	20,65217	3,666667	73,26087	139,0946	11,79384
SO2_WS (1g)	2793	17,46814	12,09091	1,272727	76,59091	234,3353	15,30801
NO2_DG (1g)	2712	28,27174	26,00000	4,130435	87,39130	173,9708	13,18980
SO2_DG (1g)	2748	14,27106	10,51087	1,238095	55,68182	123,2768	11,10301

W porównaniu ze statystykami przeprowadzonymi dla danych przed pozbyciem się wartości ekstremalnych średnia trochę zmalała dla każdego pierwiastka, mediana została taka sama jak i minimum, z kolei maksimum zmalało znacznie. Zmalała również wariancja i odchylenie standardowe.

2. Testy parametryczne i nieparametryczne

2.1. Test istotności różnic między grupami (założenia)

Utworzenie 3 nowych zmiennych:

- 1) data wzięta z nazw przypadków (w przypadku pobierania miesiąca z przypadków robiło się to błędnie przez zły format),
- 2) wzięcie numeru miesiąca z daty,
- 3) przypisanie do odpowiednich miesięcy pory roku, dla grudnia, stycznia i lutego jest to zima, dla marca, kwietnia, maja - wiosna, dla czerwca, lipca, sierpnia - lato, dla września, października, listopada - zima.

2.1.1. Test dla NO2

Do testu t Studenta prób niezależnych wybrałam dwie zmienne zależne - NO₂ dla Wodzisławia i Dąbrowy oraz zmienną grupującą - pory roku, a z niej zimę i lato.

HO: Średnia stężenie NO₂ w Wodzisławiu-Śląskim i Dąbrowie Górniczej są takie same zimą i latem.

H1: Średnia stężenie różni się.

Tabela 7. Test istotności różnic między grupami dla NO₂ (Wodzisław Śląski i Dąbrowa Górnicza)

	Testy t; Grupująca:pora roku: =Match(V36;1;"Zima";2;"Zima";3;"Wiosna";4;"Wiosna";5;"Wiosna";6;"Lato";7;"Lato";8;"Lato";9;"Jesień";10;"Jesień";11;"Jesień";12;"Zima") (BEZ EKSTREMALN Grupa 1: Zima Grupa 2: Lato											
Zmienna	Srednia Zima	Srednia Lato	t	df	p	Nważnych Zima	Nważnych Lato	Odc.h.std Lato	Odc.h.std Lato	Iloraz F	p	
NO2_WS (1g)	32,68392	14,37617	31,76399	1334	0,00	681	655	13,99112	4,746421	3,822795	0,00	
NO2_DG (1g)	34,68582	22,67213	17,31792	1344	0,00	651	695	15,76066	8,979010	3,081001	0,00	

Odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej w obu przypadkach, ponieważ poziom istotności p jest mniejszy od 0,05. Oznacza to, że średnie stężenia NO_2 w wybranych dwóch miastach różnią się statystycznie istotnie, czyli jest to wynik badania, a nie przypadku.

2.2. Test istotności różnic między zmiennymi

2.2.1. Test dla NO2

Do testu t Studenta dla prób niezależnych, względem zmiennych jako listy
wybrałam NO2 dla Wodzisławia Śląskiego jak i Dąbrowy Górniczej.

H0: Średnia stężenia NO₂ w Wodzisławiu-Śląskim i Dąbrowie Górniczej są takie same.

H1: Średnia stężenie różni się.

Tabela 8. Test istotności różnic między zmiennymi dla NO₂ (Wodzisław Śląski i Dąbrowa Górnicza)

		Testy dla prób niezależnych (BEZ EKSTREMALNYCH)										
		Uwaga: Zmienne traktowane są jako niezależne próby.										
Grupa 1 wz. Grupy 2		Średnia Grupa 1	Średnia Grupa 2	t	df	p	Nważnych Grupa 1	Nważnych Grupa 2	Odch.std Grupa 1	Odch.std Grupa 2	iloraz F Wariancie	p Wariancie
NO2 WS (1a) vs. NO2 DG (1a)		23.27433	28.27174	-14.7728	5465	0.00	2755	2712	11.79384	13.18980	1.250737	0.000000

Odrzucamy hipotezę zerową na rzecz hipotezy alternatywnej, ponieważ poziom istotności p jest mniejszy od 0,05. Oznacza to, że średnie stężenie NO₂ w wybranych dwóch miastach różni się statystycznie istotnie, czyli jest to wynik badania, a nie przypadku.

2.3. Test normalności rozkładu

2.3.1. Test Shapiro-Wilka

2.3.1.1. NO₂ i SO₂ Wodzisław Śląski

Wyżej został przeprowadzony ten test dla pierwiastków.

2.3.1.2. Pozostałe zmienne

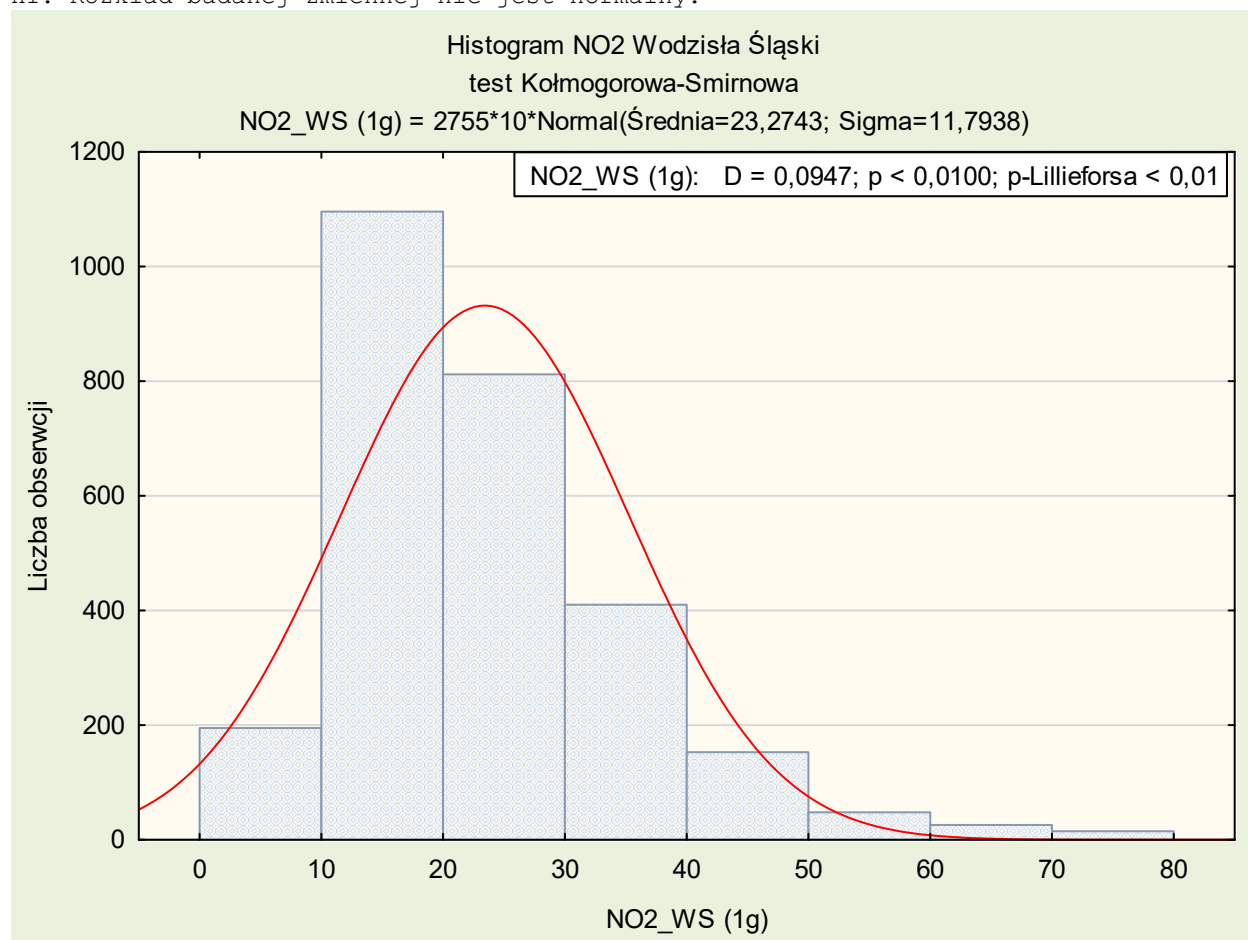
Po sprawdzeniu wszystkich zmiennych w arkuszu, żadna z nich nie posiada rozkładu normalnego.

2.3.2. Test Kołmogorowa-Smirnowa

2.3.2.1. NO₂ Wodzisław Śląski

H₀: Rozkład badanej zmiennej jest normalny.

H₁: Rozkład badanej zmiennej nie jest normalny.



Rys. 7. Histogram z testem Kołmogorowa-Smirnowa dla NO₂ Wodzisław Śląski

W teście Kołmogorowa-Smirnowa wartość p jest mniejsza od 0,01. Jest to więc wartość istotności statystyczna niższa od 0,05. Zatem możemy odrzucić hipotezę zerową na rzecz alternatywnej. Co oznacza, że zmienna posiada rozkład odbiegający od krzywej Gaussa.

2.3.2.2. Pozostałe zmienne

Przeprowadziłam test Kołmogorowa-Smirnowa dla wszystkich zmiennych i żadna z nich nie posiada rozkładu normalnego.

2.3.3. Podsumowanie

Z przeprowadzonych obu testów wyszło, iż żadna zmienna nie posiada rozkładu normalnego.

2.4. Przedział ufności dla średniej

Dzięki przedziałowi ufności na podstawie próby można stwierdzić, że dana wartość w populacji będzie w tym zakresie z wybraną dokładnością. Przedział ufności zależy od **poziomu ufności** $1-\alpha$, gdzie α jest **poziomem istotności**, czyli maksymalnym prawdopodobieństwem, że popełni się błąd pierwszego rodzaju (prawdopodobieństwo odrzucenia hipotezy zerowej, pomimo faktu, iż była prawdziwa). Najczęściej poziom istotności wynosi 0,05, 0,02, 0,01 lub 0,1. Przedział ufności zależy również od liczebności próby oraz odchylenia standardowego. Im większe odchylenie, tym szerszy jest przedział ufności.

W poniższym przykładzie poziom ufności wynosi 95%.

Tabela 9. Przedział ufności dla wybranych pierwiastków

Zmienna	Statystyki opisowe (BEZ EKSTREMALNYCH)						
	Nważnych	Średnia	Ufność -95,000%	Ufność 95,000%	Minimum	Maksimum	Odch.std
CO_WS (1g)	2821	0,53338	0,51955	0,54721	0,117826	1,93826	0,37473
CO_DG (1g)	2748	0,53009	0,51895	0,54124	0,131304	1,56048	0,29791
NO2_WS (1g)	2755	23,27433	22,83375	23,71492	3,666667	73,26087	11,79384
NO2_DG (1g)	2712	28,27174	27,77511	28,76837	4,130435	87,39130	13,18980
SO2_WS (1g)	2793	17,46814	16,90018	18,03610	1,272727	76,59091	15,30801
SO2_DG (1g)	2748	14,27106	13,85575	14,68637	1,238095	55,68182	11,10301

Czyli, średnia populacji dla danego pierwiastka będzie w tym przedziale z 95% pewnością. Dla NO2 w Wodzisławiu Śląskim przedział ufności średniej wynosi [22,83375, 23,71492]. Jak wspomniałam wcześniej, przy małym odchyleniu przedział ufności jest mniejszy, co można zauważyć porównując przedział dla CO w porównaniu z NO2 i SO2.

2.5. Test na różnicę między dwoma wskaźnikami struktury

Sprawdzenie, czy liczba dni, kiedy stężenie NO2 przekracza normę 40, jest taka sama w Wodzisławiu Śląskim i Dąbrowie Górniczej za pomocą testu na różnicę między dwoma wskaźnikami struktury. Do tego tworzę 2 nowe zmienne, gdzie znajdują się 0, kiedy norma nie jest przekroczona i 1 w przeciwnym przypadku dla obu miast. Utworzenie podstawowych statystyk potrzebnych do testu.

Tabela 10. Statystyki podstawowe potrzebne do testu na różnicę między dwoma wskaźnikami struktury

Zmienna	Statystyki opisowe (BEZ EKSTREMALNYCH)				
	Nważnych	Średnia	Minimum	Maksimum	Odch.std
w normie NO2_WS	2755	0,087840	0,00	1,000000	0,283114
w normie NO2_DG	2712	0,176254	0,00	1,000000	0,381106

H0: Liczba dni jest taka sama.

H1: Liczba dni różni się.

Rys. 8. Test na różnicę między dwoma wskaźnikami struktury

Inne testy istotności: BEZ EKSTREMALNYCH

☐ Wyślij lub drukuj wyniki do okna raportu dla każdego obliczenia Anuluj

Różnica między dwoma współczynnikami korelacji

r1: 0,00 N1: 10
r2: 0,00 N2: 10 p= 1,0000 ☐ Jednostronny ☒ Dwustronny Oblicz

Różnica między dwiema średnimi (rozkład normalny)

Śr.1: 0 Odch.std.1: 1 N1: 10
Śr.2: 0 Odch.std.2: 1 N2: 10 p= 1,0000 ☐ Jednostronny ☒ Dwustronny Oblicz

☐ Średnia z pomiarów 1 a średnia z populacji 2

Różnica między dwoma wskaźnikami struktury

% 1: ,08784 N1: 2755
% 2: ,176254 N2: 2712 p= 0,0000 ☐ Jednostronny ☒ Dwustronny Oblicz

W %1 oraz %2 wpisuję odpowiednie średnie, a N1 i N2 odpowiednie wartości z kolumny Nważnych. Wybranie testu dwustronnego. Wartość p jest niższa od 0,05, więc odrzucam hipotezę zerową na rzecz alternatywnej, co oznacza że liczba dni kiedy norma jest przekroczona różni się.

2.5.1. Test jednostronny i dwustronny

Test jednostronny to test statystyczny, dla którego skrajne wartości obserwowanej zmiennej znajdują się po lewej lub prawej stronie jej rozkładu (włas na wykresie ramka-wąsy jest tylko po jednej stronie).
 $p_1 > p_2$ lub $p_1 < p_2$

Test dwustronny to rodzaj testu statystycznego, dla którego skrajne wartości obserwowanej zmiennej znajdują się po obydwu stronach jej rozkładu (wąsy są po obu stronach wykresu).
 $p_1 = p_2$ lub $p_1 \neq p_2$

2.6. Liczność próby

Dodanie nowej zmiennej do danych, która jest połączeniem NO2 dla Wodzisławia Śląskiego i Dąbrowy Górniczej, aby wyliczyć odchylenie standardowe potrzebne do obliczenia liczności próby.

Tabela 11. Odchylenie standardowe potrzebne do liczności próby

Zmienna	Statystyki opisowe (BEZ EKSTREMALNYCH)	
	Nważnych	Odch.std
NO2_WS i DG	5467	12,75191

Policzenie średniej dla NO2 w obu miastach.

Tabela 12. Średnie potrzebne do liczności próby

Zmienna	Statystyki opisowe (BEZ EKSTREMALNYCH)	
	Nważnych	Średnia
NO2_WS (1g)	2755	23,27433
NO2_DG (1g)	2712	28,27174

Wykonanie testu t dla prób niezależnych, licznosc próby. Mi1 i Mi2 są to średnie dla Wodzisławia jak i Dąbrowy. Przy α zostawiałam wartość domyślną, sigma - odchylenie standardowe połączonych zmiennych, moc docelowa domyślnie, czyli 0,90.

Tabela 13. Licznosc próby

	Licznosc próby (BEZ EKSTREMALNYCH) Dwie średnie, test t, próby niezależne H0: Mi1 = Mi2
	Wartość
Średnia populacyjna Mi1	23,2743
Średnia populacyjna Mi2	28,2717
Odch. std. w populacji (Sigma)	12,7519
Efekt standaryzowany (Es)	-0,3919
Prawdop. bł. I rodzaju (Alfa)	0,0500
Wartość krytyczna t	1,9687
Moc docelowa	0,9000
Moc dla wymaganej licznosci próby N	0,9004
Wymagane N (w grupie)	138,0000

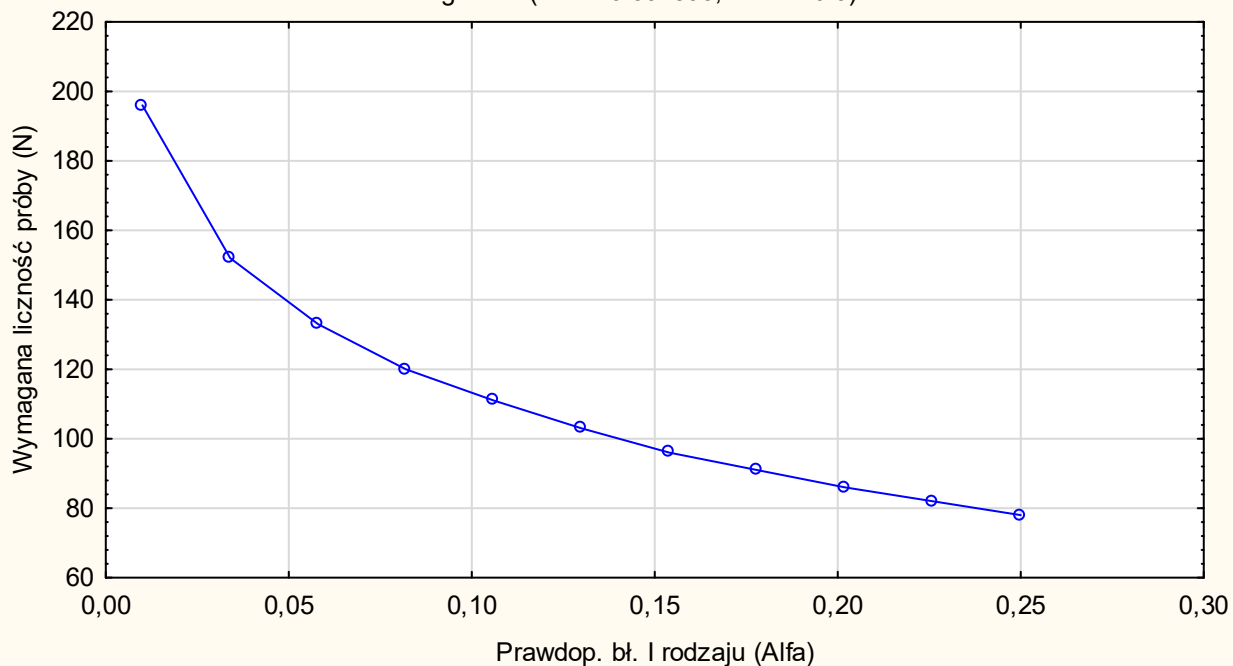
Minimalny rozmiar próby wynosi 138 na poziomie istotności równym 0,05.

Im niższy poziom istotności, tym trudniej jest odrzucić hipotezę zerową, kiedy ta jest prawdziwa i tym większa jest licznosc próby.

Test t dla prób niezależnych: Obliczanie licznosci próby

Dwie średnie, test t, próby niezależne (H0: Mi1 = Mi2)

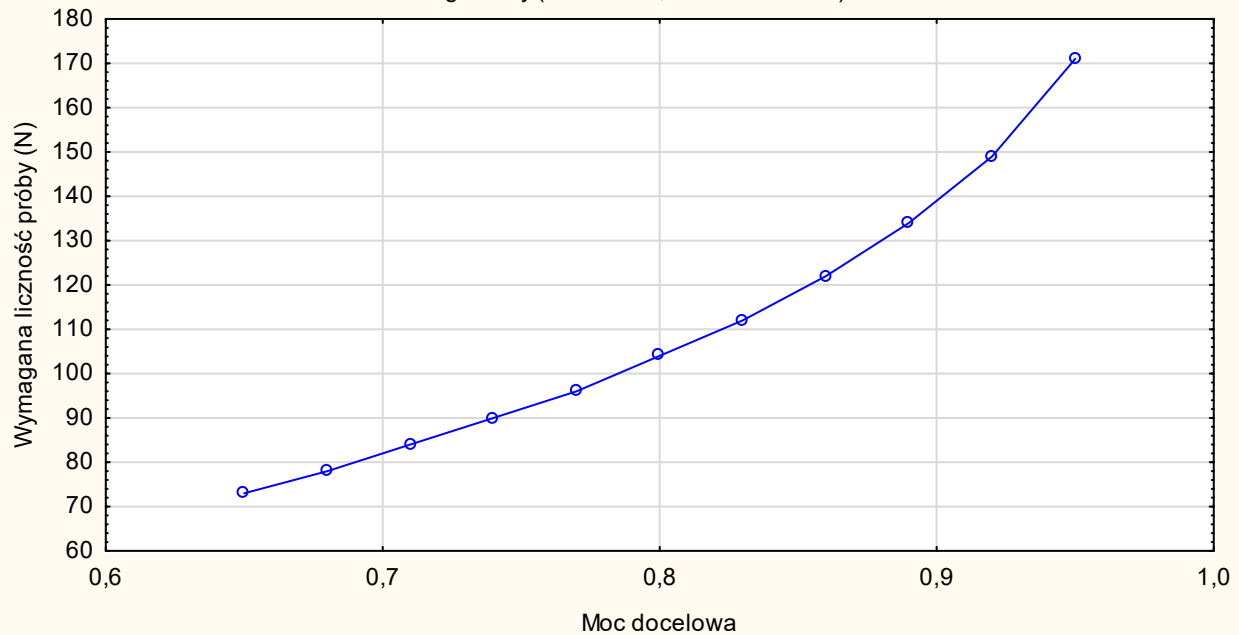
N wzgl. Alfa (Es = -0.391895, Moc = 0.9)



Rys. 9. Wymagana licznosc próby w zaleznosci od prawdopodobieństwa błędów I rodzaju

Im większa moc docelowa, tym większa szansa na odrzucenie hipotezy zerowej, gdy ta jest fałszywa.

Test t dla prób niezależnych: Obliczanie liczności próby
Dwie średnie, test t, próby niezależne ($H_0: \mu_1 = \mu_2$)
N wzgl. Moc (Alfa = 0.05, Es = -0.391895)



Rys. 10. Wymagana licznosc próby w zaleznosci od mocy testu

Suma tych dwóch prawdopodobieństw jest równa 1, dlatego jeśli chcemy zwiększyć moc docelową, musimy zwiększyć prawdopodobieństwo błędu pierwszego rodzaju, i na odwrót.

3. Korelacja i regresja

3.1. Wyznaczenie i analiza współczynników Pearsona i Spearmana

Obliczenie współczynników Pearsona wszystkich zmiennych. Określa on poziom zależności liniowej między zmiennymi.

Tabela 14. Współczynnik Pearsona dla wszystkich zmiennych

Zmienna	Korelacje (BEZ EKSTREMALNYCH) Oznaczone wsp. korelacji są istotne z p < .05000										
	CO WS (1g)	NO WS (1g)	NO2 WS (1g)	Nox WS (1g)	O3 WS (1g)	PM10 WS (1g)	SO2 WS (1g)	cisnienie WS (1g)	WS (1g)	kier WS (1g)	predk WS (1g)
CO WS (1g)	1,000000	0,802948	0,822972	0,884922	-0,550012	0,907346	0,829706	0,125045	-0,219984	-0,11108	-0,1110
NO WS (1g)	0,802948	1,000000	0,761116	0,922740	-0,644117	0,790546	0,676591	0,185112	-0,204557	-0,13747	-0,1374
NO2 WS (1g)	0,822972	0,761116	1,000000	0,932300	-0,558865	0,846804	0,863336	0,134732	-0,233559	0,04661	0,0466
Nox WS (1g)	0,884922	0,922740	0,932300	1,000000	-0,633171	0,896978	0,820812	0,146786	-0,231532	-0,05671	-0,0567
O3 WS (1g)	-0,550012	-0,644117	-0,558865	-0,633171	1,000000	-0,486251	-0,485476	-0,103311	0,052597	0,06747	0,0674
PM10 WS (1g)	0,907346	0,790546	0,846804	0,896978	-0,486251	1,000000	0,837068	0,199395	-0,218422	-0,04451	-0,0445
SO2 WS (1g)	0,829706	0,676591	0,863336	0,820812	-0,485476	0,837068	1,000000	0,134460	-0,198810	0,08111	0,0811
cisnienie WS (1g)	0,125045	0,185112	0,134732	0,146786	-0,103311	0,199395	0,134460	1,000000	-0,042410	-0,21544	-0,2154
kier WS (1g)	-0,219984	-0,204557	-0,233559	-0,231532	0,052597	-0,218422	-0,198810	-0,042410	1,000000	0,40524	0,4052
predk WS (1g)	-0,11108	-0,13747	0,046636	-0,056760	0,067464	-0,044560	0,081196	-0,215494	0,405246	1,00000	1,0000
predk2 WS (1g)	-0,306510	-0,309156	-0,294212	-0,301432	0,029630	-0,255996	-0,137135	-0,000538	0,176516	-0,05801	-0,0580
prom WS (1g)	-0,429520	-0,398982	-0,413066	-0,422196	0,781290	-0,360556	-0,441031	0,010194	-0,069797	-0,12801	-0,1280
temp WS (1g)	-0,598930	-0,446786	-0,591502	-0,547226	0,611267	-0,552113	-0,714016	-0,079482	-0,043054	-0,32261	-0,3226
wilg WS (1g)	0,227605	0,309223	0,210192	0,265713	-0,712626	0,172535	0,157576	-0,058054	0,135117	-0,07301	-0,0730
As(PM10) DG (24g)	0,371944	0,325061	0,409792	0,396935	-0,366166	0,384233	0,500063	-0,134562	-0,040641	0,20311	0,2031
BaP(PM10) DG (24g)	0,619158	0,451348	0,629736	0,563614	-0,528181	0,578990	0,705764	0,001325	0,003344	0,28601	0,2860
Cd(H6) DG (1g)	0,768854	0,663704	0,703767	0,745986	-0,566900	0,738676	0,713533	0,149231	-0,137594	-0,10211	-0,1021
CO DG (1g)	0,804250	0,667712	0,768296	0,776715	-0,543595	0,750865	0,748836	0,133912	-0,142515	-0,11201	-0,1120
Cd(PM10) DG (24g)	0,282443	0,229020	0,252511	0,300705	-0,212065	0,294993	0,354662	0,122855	0,018307	0,26031	0,2603
NO DG (1g)	0,553255	0,616793	0,573887	0,646614	-0,352781	0,548535	0,424505	0,177445	-0,033247	-0,13131	-0,1313
NO2 DG (1g)	0,598541	0,527064	0,723275	0,686383	-0,308397	0,592333	0,580090	0,081355	0,013962	0,07411	0,0741
Nox DG (1g)	0,619557	0,646772	0,674786	0,712371	-0,368490	0,593991	0,509155	0,162005	-0,022333	-0,06221	-0,0622
Ni(PM10) DG (24g)	0,186583	0,117135	0,151335	0,133915	-0,145297	0,188232	0,177525	0,122460	-0,017897	0,22041	0,2204
O3 DG (1g)	-0,546500	-0,605299	-0,565071	-0,610243	0,867165	-0,461545	-0,463147	-0,085930	0,059511	0,10351	0,1035
PM10 DG (24g)	0,710535	0,584473	0,731623	0,720093	-0,360498	0,762765	0,705811	0,162936	-0,156890	-0,09261	-0,0926
Pb(PM10) DG (24g)	0,382735	0,293627	0,359215	0,383545	-0,257781	0,363334	0,472434	0,010255	0,026123	0,36761	0,3676
SO2 DG (1g)	0,625546	0,442355	0,646872	0,598762	-0,410244	0,586105	0,742377	0,073160	0,049637	0,20761	0,2076
cisnienie DG (1g)	0,148885	0,176100	0,188341	0,190905	-0,121687	0,263967	0,147461	0,982524	-0,039455	-0,11931	-0,1193
kier DG (1g)	-0,224710	-0,209872	-0,260377	-0,247561	0,038722	-0,230792	-0,226713	-0,046895	0,785883	0,29911	0,2991
opad atm DG (1g)	-0,197369	-0,172033	-0,257334	-0,215725	0,029440	-0,252055	-0,230985	-0,250202	0,261443	0,44071	0,4407
predk DG (1g)	0,036530	0,063784	0,134604	0,099081	-0,051632	0,051685	0,188455	-0,132755	0,171216	0,54861	0,5486
prom DG (1g)	-0,410071	-0,389333	-0,401925	-0,410474	0,775445	-0,341855	-0,435420	0,034054	-0,092819	-0,15351	-0,1535
temp DG (1g)	-0,546192	-0,438377	-0,550326	-0,513875	0,604876	-0,541461	-0,698951	-0,104405	-0,013805	-0,26351	-0,2635

Jeśli współczynnik jest czerwony, oznacza to że jest istotny statystycznie.

Kolorowa wersja powyższej macierzy korelacji.

Tabela 15. Współczynnik Pearsona dla wszystkich zmiennych - wersja kolorowa

Zmienna	Kolorowa mapa korelacji (BEZ EKSTREMALNYCH) Oznaczone wsp. korelacji są istotne z p < .05000									
	p < .05000									
	CO WS (1g)	NO WS (1g)	NO2 WS (1g)	Nox WS (1g)	O3 WS (1g)	PM10 WS (1g)	SO2 WS (1g)	ciśnienie WS (1g)	kier WS (1g)	predk WS (1g)
CO WS (1g)	1.000000	0.802848	0.822372	0.884922	-0.559392	0.807346	0.829706	0.125045	-0.215984	-0.111086
NO WS (1g)	0.802848	1.000000	0.761116	0.922740	-0.644117	0.790546	0.676591	0.185112	-0.204557	-0.137476
NO2 WS (1g)	0.822372	0.761116	1.000000	0.932300	-0.558865	0.846804	0.863338	0.134732	-0.233559	0.046638
Nox WS (1g)	0.884922	0.922740	0.932300	1.000000	-0.633171	0.898978	0.820812	0.146788	-0.231532	-0.066760
O3 WS (1g)	-0.559392	-0.644117	-0.558865	-0.633171	1.000000	-0.486251	-0.495478	-0.103311	0.052597	0.067464
PM10 WS (1g)	0.807346	0.790546	0.846804	0.898978	-0.486251	1.000000	0.837069	0.199399	-0.218422	-0.044560
SO2 WS (1g)	0.829706	0.676591	0.863338	0.820812	-0.485478	0.837069	1.000000	0.134460	-0.198910	0.081198
ciśnienie WS (1g)	0.125045	0.185112	0.134732	0.146788	-0.103311	0.199399	0.134460	1.000000	0.042410	-0.215494
kier WS (1g)	-0.215984	-0.204557	-0.233559	-0.231532	0.052597	-0.218422	-0.198910	-0.042410	1.000000	0.405246
predk WS (1g)	-0.111086	-0.137476	0.046638	-0.066760	0.067464	-0.044560	0.081198	-0.215494	0.405246	1.000000
predk2 WS (1g)	-0.306510	-0.309156	0.294212	-0.301432	0.029630	-0.255998	-0.137139	-0.000535	0.176518	-0.058054
prom. WS (1g)	-0.429520	-0.398982	-0.413066	-0.422196	0.781290	-0.360561	-0.441031	0.010194	-0.069797	-0.128082
temp. WS (1g)	-0.598930	-0.446786	-0.591502	-0.547226	0.611267	-0.552113	-0.714018	-0.079482	-0.043054	-0.322655
wilg. WS (1g)	0.227609	0.309223	0.210192	0.265713	-0.712626	0.172536	0.157576	-0.058084	0.135117	-0.073048
As(PM10) DG (24g)	0.371944	0.325061	0.409792	0.398939	-0.366166	0.384233	0.500063	-0.134562	-0.040641	0.203150
Ba(PM10) DG (24g)	0.619158	0.451348	0.629738	0.563614	-0.528181	0.578990	0.705764	0.001325	0.003344	0.286036
C6H6 DG (1g)	0.768654	0.663704	0.703767	0.745986	-0.566900	0.738678	0.713533	0.149231	-0.137594	-0.102101
CO DG (1g)	0.804250	0.667712	0.768296	0.776719	-0.543599	0.750869	0.748839	0.133912	-0.142515	-0.112010
Cd(PM10) DG (24g)	0.282443	0.229020	0.252511	0.300709	-0.212069	0.294993	0.354662	0.122855	0.018307	0.260330
NO DG (1g)	0.553255	0.616793	0.573887	0.646614	-0.352781	0.548539	0.424509	0.177445	-0.033247	-0.131367
NO2 DG (1g)	0.598541	0.527064	0.723275	0.686383	-0.308397	0.592333	0.580090	0.081350	0.013662	0.074179
Nox DG (1g)	0.619557	0.646772	0.674788	0.712371	-0.368490	0.593991	0.509155	0.162009	-0.022333	-0.062248
Ni(PM10) DG (24g)	0.185583	0.117135	0.151335	0.133919	-0.145297	0.186232	0.177525	0.122460	-0.017897	0.220440
O3 DG (1g)	-0.546500	-0.605299	-0.565071	-0.610243	0.867165	-0.461549	-0.463147	-0.085930	0.059511	-0.103528
PM10 DG (24g)	0.710539	0.584473	0.731623	0.720093	-0.360498	0.782769	0.705811	0.182936	-0.156980	-0.092852
Pb(PM10) DG (24g)	0.382739	0.293627	0.359219	0.383548	-0.257781	0.363334	0.472434	0.010255	0.026123	0.367659
SO2 DG (1g)	0.625546	0.442356	0.646872	0.598762	-0.410244	0.586105	0.742377	0.073160	0.049637	0.207675
ciśnienie DG (1g)	0.148889	0.176100	0.188341	0.190905	-0.121687	0.236967	0.147461	0.082524	-0.030455	-0.119358
kier DG (1g)	-0.224710	-0.209872	-0.260377	-0.247561	0.038722	-0.230792	-0.226713	-0.048899	0.785883	0.299123
opad atm DG (1g)	-0.197369	-0.172033	-0.257334	-0.215728	0.029440	-0.252058	-0.230985	-0.250202	0.261443	0.440747

Im kolor komórki jest bardziej czerwony tym zmienne są silniej skorelowane ujemnie, a im bardziej niebieskie tym dodatnie.
Współczynnik Pearsona dla NO2 i SO2 w Wodzisławiu: 0,863338.
Współczynnik Pearsona dla CO i PM10 w Wodzisławiu: 0,907346.
Wybrane współczynniki są istotne statystycznie.

Obliczenie współczynników Spearmana wszystkich zmiennych. Określa on poziom monotonicznej korelacji. Czyli liniowej, wykładniczej lub logarytmicznej itd.

Tabela 16. Współczynnik Spearmana dla wszystkich zmiennych

Zmienna	Korelacja porządku rang Spearmana (BEZ EKSTREMALNYCH) BD usuwane parami Oznaczone wsp. korelacji są istotne z p < .05000									
	p < .05000									
	CO WS (1g)	NO WS (1g)	NO2 WS (1g)	Nox WS (1g)	O3 WS (1g)	PM10 WS (1g)	SO2 WS (1g)	ciśnienie WS (1g)	kier WS (1g)	predk WS (1g)
CO WS (1g)	1.000000	0.710699	0.803079	0.817170	-0.586750	0.854479	0.823322	0.146597	-0.272975	-0.104919
NO WS (1g)	0.710699	1.000000	0.827328	0.909336	-0.694444	0.718389	0.648675	0.254026	-0.278521	-0.111187
NO2 WS (1g)	0.803079	0.827328	1.000000	0.979290	-0.605701	0.863860	0.834381	0.172309	-0.270363	0.090645
Nox WS (1g)	0.817170	0.909336	0.979290	1.000000	-0.676224	0.869088	0.811673	0.192654	-0.284590	0.016854
O3 WS (1g)	-0.586750	-0.694444	-0.605701	-0.676224	1.000000	-0.501048	-0.515939	-0.113699	0.070360	0.041451
PM10 WS (1g)	0.854479	0.718389	0.863860	0.869088	-0.501048	1.000000	0.842397	0.220230	-0.279002	0.024465
SO2 WS (1g)	0.823322	0.648675	0.834381	0.811673	-0.515939	0.842397	1.000000	0.144144	-0.252434	0.157183
ciśnienie WS (1g)	0.146597	0.254026	0.172309	0.192654	-0.113699	0.220230	0.144144	1.000000	-0.017724	-0.166265
kier WS (1g)	-0.272975	-0.278521	-0.270363	-0.284590	0.070360	-0.279002	-0.252434	-0.017724	1.000000	0.427908
predk WS (1g)	-0.104919	-0.111187	0.090645	0.016854	0.041451	0.024465	0.157183	-0.166265	0.427908	1.000000
predk2 WS (1g)	-0.297534	-0.347347	-0.256065	-0.289477	0.039009	-0.200669	-0.009965	0.042325	0.217435	-0.046443
prom. WS (1g)	-0.499952	-0.405708	-0.438172	-0.451526	0.777247	-0.384592	-0.497051	0.011473	-0.061387	-0.116532
temp. WS (1g)	-0.678174	-0.462049	-0.592033	-0.577945	0.618644	-0.595750	-0.747154	-0.078551	-0.049174	-0.319811
wilg. WS (1g)	0.234394	0.314268	0.219225	0.260972	-0.706281	0.157699	0.134168	-0.065423	0.120852	-0.029942
As(PM10) DG (24g)	0.336964	0.225481	0.378391	0.357517	-0.329594	0.396021	0.526125	-0.056949	-0.021171	0.282589
Ba(PM10) DG (24g)	0.747795	0.489016	0.634010	0.617161	-0.632558	0.656024	0.778704	0.032093	-0.102603	0.087418
C6H6 DG (1g)	0.827920	0.649478	0.726647	0.734160	-0.621258	0.746239	0.739088	0.150434	-0.205969	-0.149850
CO DG (1g)	0.821041	0.631275	0.757082	0.754993	-0.578565	0.762317	0.750604	0.157372	-0.202064	-0.128016
Cd(PM10) DG (24g)	0.373784	0.238693	0.306463	0.308961	-0.207204	0.364729	0.447574	0.071257	0.002580	0.308960
NO DG (1g)	0.495001	0.665851	0.593072	0.642764	-0.381833	0.505843	0.374746	0.171593	-0.049140	-0.158135
NO2 DG (1g)	0.552502	0.571483	0.693926	0.694482	-0.309894	0.568783	0.524837	0.083848	-0.009129	0.072412
Nox DG (1g)	0.555658	0.650774	0.676780	0.699781	-0.367026	0.560042	0.464824	0.151217	-0.031729	-0.051102
Ni(PM10) DG (24g)	0.285043	0.161549	0.244281	0.187313	-0.131272	0.254694	0.299436	0.118228	-0.001220	0.283357
O3 DG (1g)	-0.593715	-0.667281	-0.610828	-0.658314	0.876936	-0.489428	-0.494500	-0.097001	0.081122	0.084114
PM10 DG (24g)	0.663762	0.558394	0.712208	0.699565	-0.320918	0.804512	0.663124	0.190735	-0.266431	-0.130963
Pb(PM10) DG (24g)	0.513591	0.336200	0.404967	0.420973	-0.260077	0.469352	0.583560	-0.042036	0.005401	0.400523
SO2 DG (1g)	0.648460	0.408449	0.603610	0.581583	-0.405933	0.599700	0.736300	0.052539	0.026181	0.237302
ciśnienie DG (1g)	0.145267	0.257652	0.225651	0.242897	-0.125446	0.275045	0.168655	0.975031	-0.018475	-0.047545
kier DG (1g)	-0.213472	-0.236530	-0.264173	-0.263406	0.065585	-0.262396	-0.235136	-0.054465	0.797091	0.289949
opad atm DG (1g)	-0.207093	-0.214311	-0.278784	-0.262451	0.006540	-0.340363	-0.271824	-0.271768	0.297548	0.610622

Czerwone korelacje są istotne statystycznie.
Współczynnik Spearmana dla NO2 i SO2 w Wodzisławiu: 0,834381.
Współczynnik Spearmana dla CO i PM10 w Wodzisławiu: 0,854479.
Wybrane współczynniki są istotne statystycznie.

Wartości współczynnika Pearsona jak i Spearmana są podobne dla wybranych zmiennych, co oznacza, że korelacja jest monotoniczna. Gdyby współczynnika Pearsona był znacznie większy od tego drugiego, to korelacja byłaby liniowa.

3.1.1. Test istotności współczynnika korelacji

3.1.1.1. Test istotności współczynnika Pearsona dla NO₂ i SO₂ w Wodzisławiu

H₀: Współczynnik wynosi 0.

H₁: Współczynnik różni się od 0.

W tym przypadku współczynnik wynosi 0,863338, więc odrzucamy hipotezę zerową na rzecz alternatywnej.

3.1.1.2. Test istotności współczynnika Spearmana dla NO₂ i SO₂ w Wodzisławiu

H₀: Współczynnik wynosi 0.

H₁: Współczynnik różni się od 0.

W tym przypadku współczynnik wynosi 0,834381, więc odrzucamy hipotezę zerową na rzecz alternatywnej.

3.2. Regresja liniowa, budowa i weryfikacja

Regresja liniowa pozwala na określenie zależności między zmienną zależną (objaśnianą-y) a zmienną niezależną (objaśniającą-X). Różne metody wyznaczania regresji (np. metoda najmniejszych kwadratów) znajdują taką prostą, która najlepiej opisuje zależność między zmiennymi, czyli mówiąc inaczej, będzie przebiegała jak najbliżej wszystkich punktów wykresu. Ogólny zapis prostej: $y=a_0+a_1*x_1$, gdzie a_0 , a_1 to parametry regresji, a x_1 to zmienne niezależne.

Za pomocą współczynnika determinacji (R^2) możemy ocenić jakość modelu regresji liniowej, innymi słowy, określić jak dobrze model regresji liniowej dopasowuje się do danych. Współczynnik przyjmuje wartości z przedziału $<0;1>$ i im wyższa jest wartość, tym lepiej. Określa, jaki procent wariancji zmiennej objaśnianej został wyjaśniony przez funkcję regresji.

3.2.1. Regresja liniowa dla NO₂ i SO₂ w Wodzisławiu Śląskim

NO₂ będzie zmienną niezależną, a SO₂ zależną. Między zmiennymi istnieje korelacja liniowa.

Tabela 17. Podsumowanie regresji liniowej zmiennej zależnej SO₂ Wodzisław Śląski

N=2684	Podsumowanie regresji zmiennej zależnej: SO ₂ _WS (1g) (BEZ EKSTREMALNYCH) R= ,86333847 R ² = ,74535332 Popraw. R ² = ,74525838 F(1,2682)=7850,2 p<0,0000 Błąd std. estymacji: 7,6927					
	b*	Bł. std. z b*	b	Bł. std. z b	t(2682)	p
W. wolny			-8,41094	0,328287	-25,6207	0,00
NO ₂ _WS (1g)	0,863338	0,009744	1,10851	0,012511	88,6016	0,00

Według powyższej tabeli możemy odczytać, że $y=-8,41094+1,10851*x$. Błąd standardowy z b mówi jak dokładnie jest oszacowany parametr b w regresji liniowej. Im niższy parametr, tym dokładniejsza jest estymacja. Kolejna kolumna oblicza wartość testu t Studenta z następującego wzoru: $t=b/s$, gdzie b to parametr b, a s to błąd standardowy z b. Test ten sprawdza, czy współczynnik regresji jest istotny statystycznie.

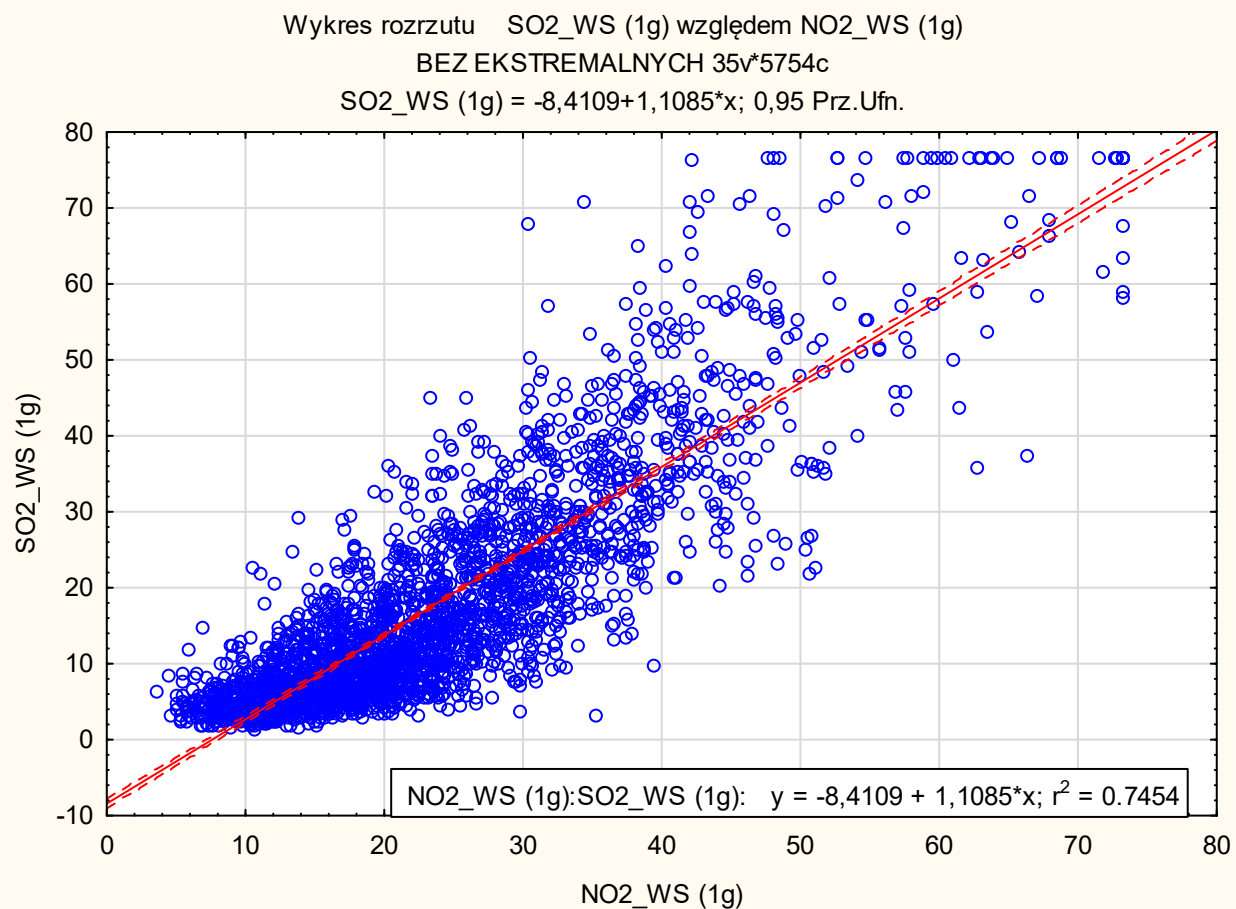
H₀: Współczynnik nie jest istotny statystycznie.

H₁: Współczynnik jest istotny statystycznie.

Poziom istotności p jest niższy niż poziom alfa 0,05, więc odrzucamy hipotezę zerową na rzecz alternatywnej. Co oznacza, że współczynnik b jest istotny statystycznie. Tyczy się to zarówno b₀ jak i b₁.

Powyższy model jest dobry, ponieważ współczynnik determinacji wynosi 0,7454.

Wykres rozrzutu. Przedział ufności został zostawiony domyślnie, czyli 95%.



Rys. 10. Wykres rozrzutu regresji liniowej

Na wykresie rozrzutu ukazana jest wyliczona prosta. Przerywane, czerwone linie to przedziały ufności dla wartości oczekiwanej modelowanej zmiennej zależnej. Przedział zależy między innymi od współczynnika alfa, wartości zmiennych, co można zauważyć na wykresie. Bliżej lewego dolnego rogu wartości są skupione bliżej linii regresji i przedział jest węższy, a bliżej prawego górnego rogu rozszerza się, gdyż wiele punktów odbiega od prostej.

Tabela 18. Wartości przewidywane i reszty z regresji liniowej

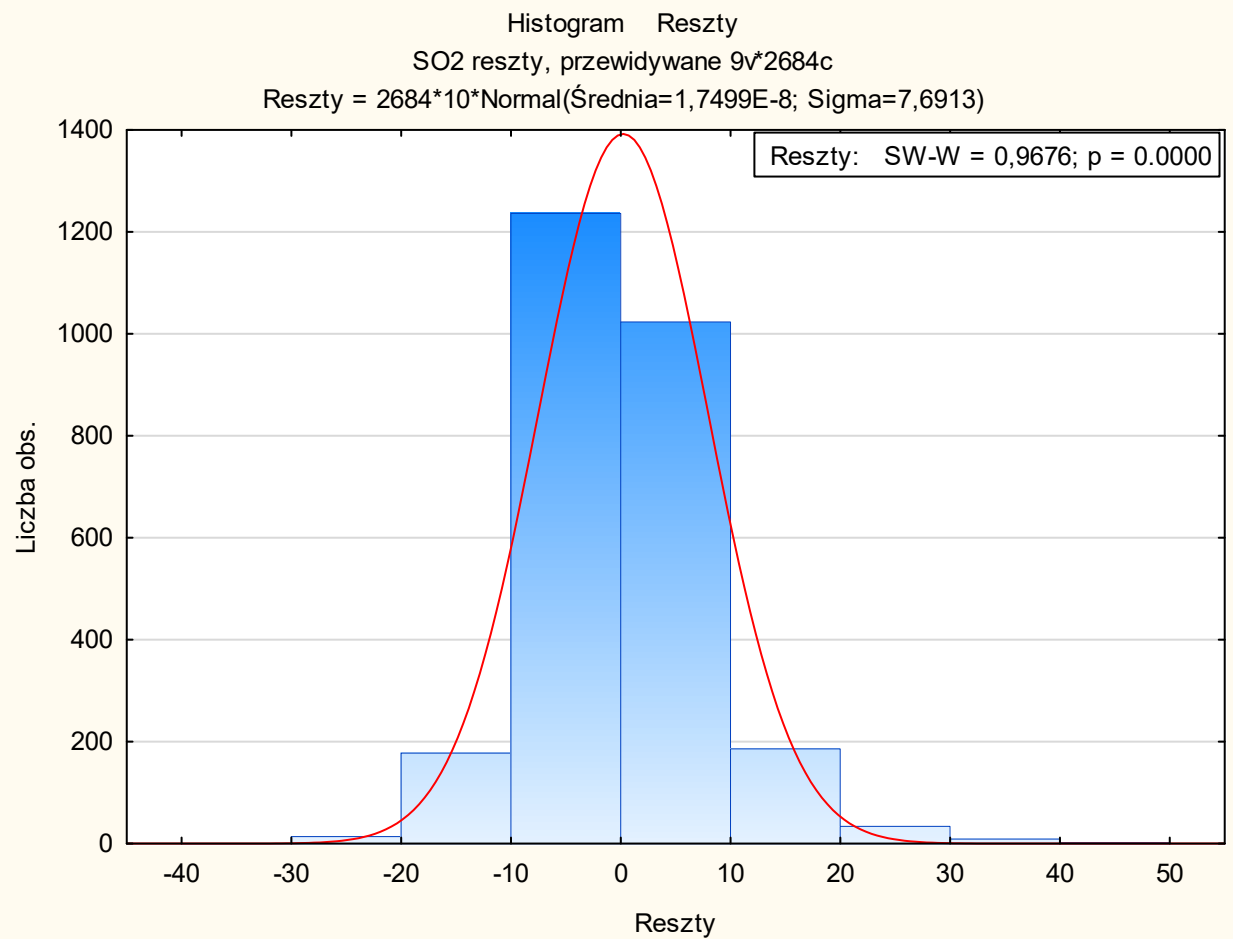
	Wartości przewidywane i reszty SO ₂ WS (1g)							
	Obserw. Wartość	Przewidyw. Wartość	Reszta	Standard. Przewid.	Standard. Reszta	Bl. std. W.przew.	Mahaln. Odlegl.	Usunięte Reszta
2007-01-01	20,727272	15,060656	5,666616	-0,187691	0,736620	0,151081	0,035228	5,668803
2007-01-02	23,454546	26,483175	-3,028629	0,680371	-0,393700	0,179607	0,462904	-3,030281
2007-01-03	13,909091	21,133389	-7,224298	0,273810	-0,939108	0,153955	0,074972	-7,227192
2007-01-04	17,227272	14,771478	2,455794	-0,209667	0,319236	0,151717	0,043960	2,456750
2007-01-05	19,363636	18,771769	0,591867	0,094338	0,076939	0,149147	0,008900	0,592090
2007-01-06	24,090910	25,471054	-1,380144	0,603454	-0,179409	0,173437	0,364157	-1,380846
2007-01-07	10,863636	9,566278	1,297358	-0,605240	0,168647	0,173575	0,366315	1,298018
2007-01-08	18,363636	24,073360	-5,709724	0,497235	-0,742224	0,165837	0,247243	-5,712379
2007-01-09	20,227272	26,434978	-6,207706	0,676708	-0,806958	0,179301	0,457934	-6,211081
2007-01-10	26,869566	25,085484	1,784082	0,574152	0,231918	0,171229	0,329651	1,784967
2007-01-11	12,636364	7,349249	5,287115	-0,773724	0,687288	0,187757	0,598649	5,290267
2007-01-12	9,700000	3,888296	5,811704	-1,036741	0,755480	0,213906	1,074833	5,816201
2007-01-13	9,136364	8,554156	0,582208	-0,682157	0,075683	0,179756	0,465337	0,582526
2007-01-14	13,681818	7,638427	6,043391	-0,751748	0,785598	0,185777	0,565125	6,046918
2007-01-15	27,772728	27,880869	-0,108141	0,786589	-0,014058	0,188932	0,618723	-0,108206
2007-01-16	28,190475	30,914940	-2,724464	1,017165	-0,354161	0,211823	1,034625	-2,726532
2007-02-06	22,272728	21,904528	0,368200	0,332413	0,047863	0,156479	0,110499	0,368353
2007-02-07	26,090910	28,796598	-2,705688	0,856181	-0,351720	0,195492	0,733045	-2,707437
2007-02-08	37,952381	29,026625	8,925756	0,873662	1,160285	0,197190	0,763285	8,931625
2007-02-09	34,090908	36,218826	-2,127918	1,420238	-0,276614	0,257950	2,017075	-2,130314
2007-02-10	42,863636	33,327049	9,536587	1,200475	1,239689	0,232024	1,441141	9,545271
2007-02-11	24,521740	17,807844	6,713896	0,021083	0,872759	0,148520	0,000445	6,716400
2007-02-12	34,636364	26,145802	8,490562	0,654732	1,103713	0,177492	0,428674	8,495085
2007-02-13	18,578947	24,527784	-5,948837	0,531769	-0,773307	0,168183	0,282779	-5,951682
2007-02-14	12,045455	18,289806	-6,244351	0,057710	-0,811722	0,148734	0,003330	-6,246686
2007-02-15	18,045454	27,688084	-9,642630	0,771938	-1,253474	0,187595	0,595889	-9,648367
2007-02-16	5,681818	15,446225	-9,764407	-0,158389	-1,269304	0,150339	0,025087	-9,768138
2007-02-17	17,181818	11,349541	5,832277	-0,469719	0,758155	0,164058	0,220636	5,834931
2007-02-18	19,818182	14,337710	5,480472	-0,242632	0,712423	0,152797	0,058870	5,482635
2007-02-19	15,090909	18,289806	-3,198897	0,057710	-0,415834	0,148734	0,003330	-3,200094
2007-02-20	33,000000	31,736570	1,263430	1,079606	0,164237	0,218533	1,165549	1,264450
2007-02-21	42,909092	29,567738	13,341354	0,914784	1,734282	0,201261	0,836830	13,350492
2007-02-22	23,636364	20,121265	3,515099	0,196893	0,456938	0,151339	0,038767	3,516459
2007-02-23	22,285715	16,379480	5,906235	-0,087466	0,767769	0,149054	0,007650	5,908453
2007-02-24	32,217392	24,169754	8,047638	0,504561	1,046136	0,166324	0,254582	8,051402

Powyższa tabela przedstawia obserwowaną wartość zmiennej zależnej SO₂ oraz przewidywaną wartość obliczoną na podstawie modelu regresji liniowej. Reszta to różnica między wartościami faktycznymi a przewidywanymi. Są one używane do oceny jakości modelu regresji liniowej. Jeśli reszty są małe i rozproszone wokół zera, oznacza to, że model jest dobrze dopasowany do danych. Natomiast duże i skoncentrowane reszty oznaczają, że model nie jest dobrze dopasowany do danych i może być potrzebne dalsze dopasowanie modelu.

Zapisanie powyższych danych do arkusza i wykonanie histogramu dla reszt oraz przeprowadzenie testu Shapiro-Wilka na normalność rozkładu.

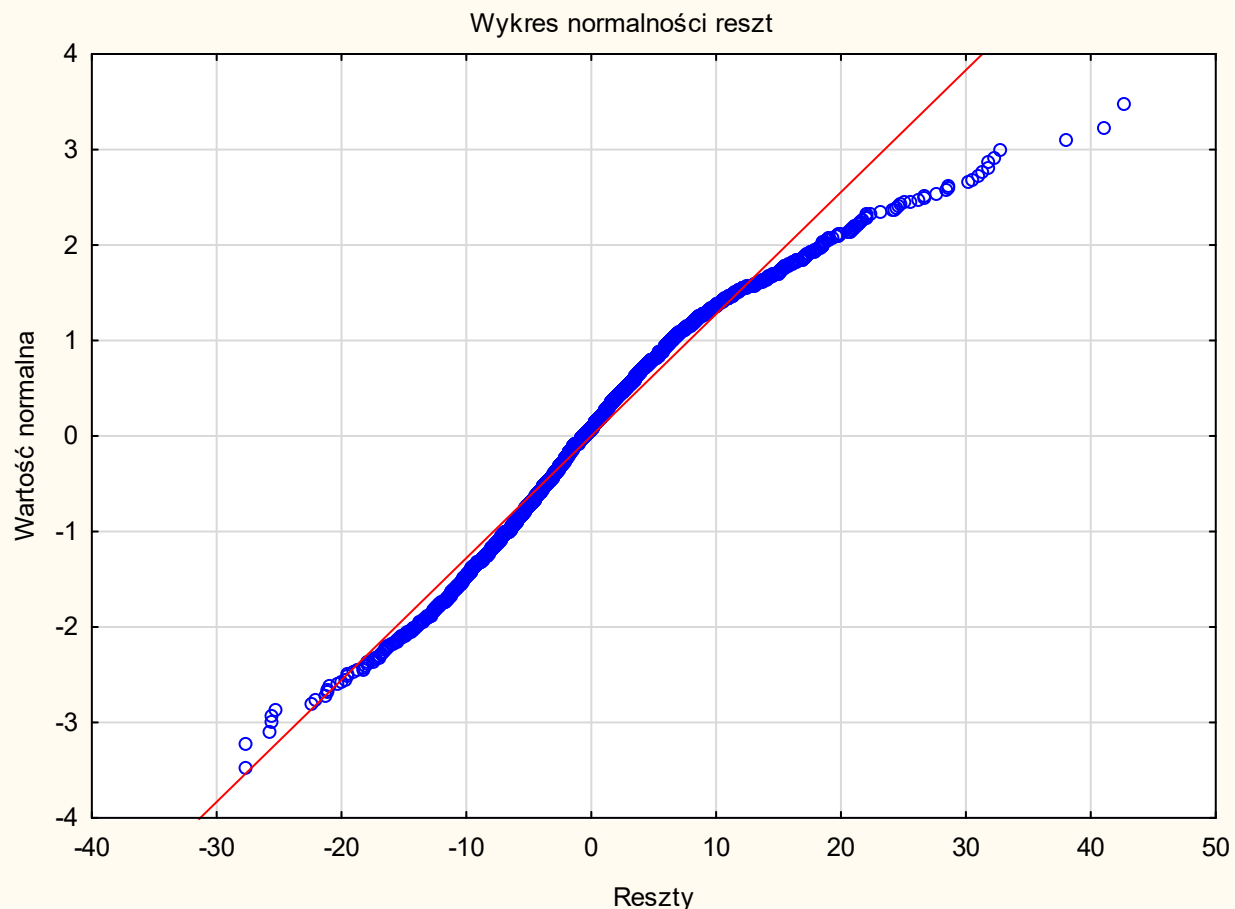
H0: Rozkład badanej zmiennej jest normalny.

H1: Rozkład badanej zmiennej nie jest normalny.



Rys. 11. Histogram reszt regresji liniowej

W teście Shapiro-Wilka wartość p wynosi 0, czyli jest mniejsza od 0.05. Zatem możemy odrzucić hipotezę zerową na rzecz alternatywnej. Co oznacza, że zmienna posiada rozkład odbiegający od krzywej Gaussa.



Rys. 12. Wykres normalności reszt

Porównując z histogramem możemy zauważyć, że im większa wartość dodatnia reszty, tym bardziej odbiegają one od prostej.

Model nie spełnia założenia normalności reszt. Może to prowadzić do nieprawidłowych wyników testów statystycznych oraz błędnych interpretacji wyników. Problem ten można przykładowo rozwiązać stosując inny model, który nie wymaga założenia normalności reszt lub transformacji danych.

Przewidywanie

Tabela 19. Statystyki podstawowe dla NO2 Wodzisław Śląski

Zmienna	Statystyki opisowe (BEZ EKSTREMALNYCH)			
	Nważnych	Minimum	Maksimum	Odch.std
NO2_WS (1g)	2755	3,666667	73,26087	11,79384

Na ich podstawie, bierzemy wartość wyższą od maksimum, przykładowo 85,34 do predykcji zmiennej zależnej SO2.
Przedział ufności 95%.

Tabela 20. Przewidywanie wartości zmiennej SO2 Wodzisław Śląski

Zmienna	Obliczanie wartości (BEZ EKSTREMALNYCH) zmiennej: SO2_WS (1g)		
	Wagi b	Wartość	Wagi b *Wartość
NO2_WS (1g)	1,108515	85,34000	94,60064
W. wolny			-8,41094
Przewidyw.			86,18970
-95,0%GU			84,64255
+95,0%GU			87,73685

Przewidywana zmienna z 95% dokładnością będzie znajdowała się w przedziale [84,64;87,74].

3.2.2. Regresja wieloraka

Regresja wieloraka pozwala na określenie zależności między zmienną zależną (objaśnianą-y) a zmiennymi niezależnymi (objaśniającą-X).

NO2, CO PM10 dla Wodzisławia Śląskiego będą zmiennymi niezależnymi, a SO2 zależną. Między zmiennymi istnieje korelacja liniowa.

Tabela 21. Podsumowanie regresji wielorakiej zmiennej zależnej SO2 Wodzisław Śląski

N=2198	Podsumowanie regresji zmiennej zależnej: SO2_WS (1g) (BEZ EKSTREMALNYCH) R= ,88745306 R^2= ,78757293 Popraw. R2= ,78728246 F(3,2194)=2711,4 p<0,0000 Błąd std. estymacji: 6,7261					
	b*	Bł. std. z b*	b	Bł. std. z b	t(2194)	p
W. wolny			-6,40217	0,340189	-18,8195	0,000000
NO2_WS (1g)	0,511760	0,018768	0,63913	0,023439	27,2675	0,000000
CO_WS (1g)	0,256433	0,024560	10,03068	0,960706	10,4409	0,000000
PM10_WS (1g)	0,164595	0,026843	0,05276	0,008605	6,1317	0,000000

Według powyższej tabeli możemy odczytać, że $y = -6,40 + 0,64 \cdot x_1 + 10,03 \cdot x_2 + 0,05 \cdot x_3$.

H0: Współczynnik nie jest istotny statystycznie.

H1: Współczynnik jest istotny statystycznie.

Poziom istotności p jest niższy niż poziom alfa 0,05, więc odrzucamy hipotezę zerową na rzecz alternatywnej. Co oznacza, że współczynnik b jest istotny statystycznie.

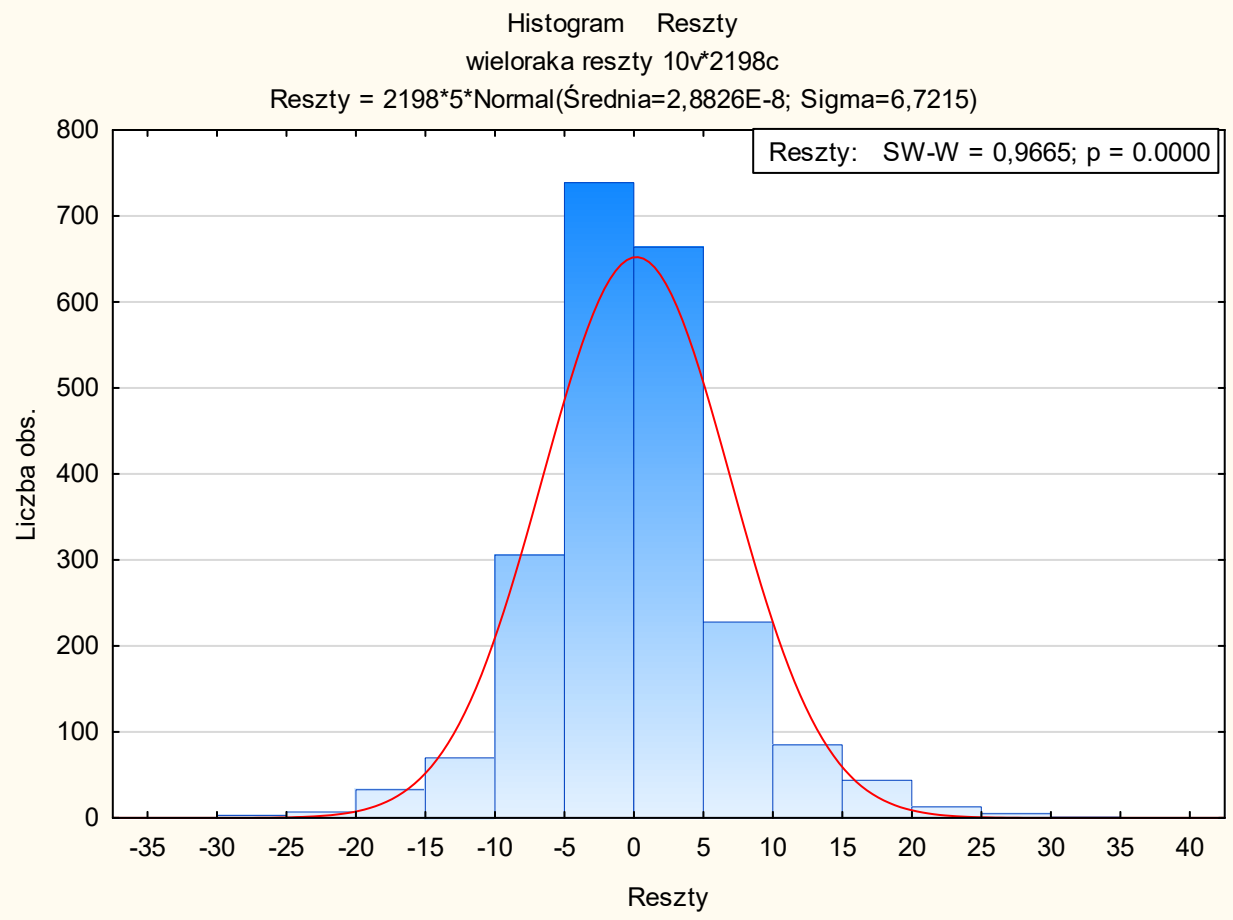
Tyczy się to wszystkich współczynników b.

Powyższy model jest dobry, ponieważ współczynnik determinacji wynosi 0,7876 i jest on wyższy niż r^2 dla pierwszej regresji liniowej, jednakże niewiele.

Zapisanie powyższych danych do arkusza i wykonanie histogramu dla reszt oraz przeprowadzenie testu Shapiro-Wilka na normalność rozkładu.

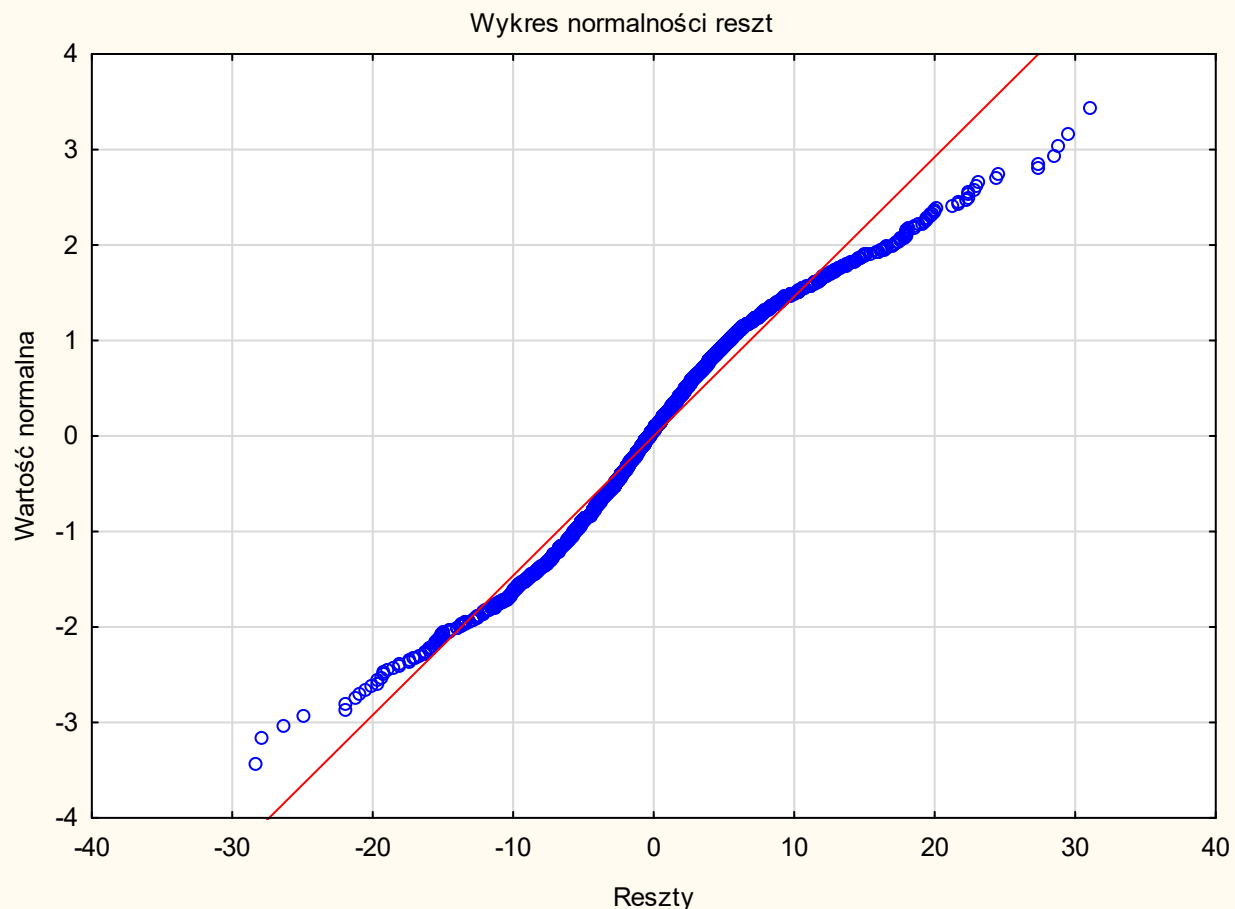
H0: Rozkład badanej zmiennej jest normalny.

H1: Rozkład badanej zmiennej nie jest normalny.



Rys. 13. Histogram reszt regresji wielorakiej

W teście Shapiro-Wilka wartość p wynosi 0, czyli jest mniejsza od 0.05. Zatem możemy odrzucić hipotezę zerową na rzecz alternatywnej. Co oznacza, że zmienna posiada rozkład odbiegający od krzywej Gaussa.



Rys. 14. Wykres normalności reszt regresji wielorakiej

Na wykresie można zauważyć, że zarówno dla dużych wartości ujemnych reszty jak i dodatnich punkty odbiegają od prostej.

Model nie spełnia założenia normalności reszt. Może to prowadzić do nieprawidłowych wyników testów statystycznych oraz błędnych interpretacji wyników. Problem ten można przykładowo rozwiązać stosując inny model, który nie wymaga założenia normalności reszt lub transformacji danych.

Przewidywanie

Tabela 22. Statystyki podstawowe dla NO₂, CO i PM₁₀.

Zmienna	Statystyki opisowe (BEZ EKSTREMALNYCH)			
	Nważnych	Minimum	Maksimum	Odch.std
NO ₂ _WS (1g)	2755	3,666667	73,2609	11,79384
CO_WS (1g)	2821	0,117826	1,9383	0,37473
PM ₁₀ _WS (1g)	2343	2,125000	225,1250	45,57596

Na ich podstawie, bierzemy wartości wyższe od maksimum, przykładowo 85,34, 3,12 oraz 228,96 do predykcji zmiennej zależnej SO₂. Przedział ufności 95%.

Tabela 23. Przewidywanie wartości zmiennej SO₂ Wodzisław Śląski - regresja wieloraka

Zmienna	Obliczanie wartości (BEZ EKSTREMALNYCH) zmiennej: SO ₂ _WS (1g)		
	Wagi b	Wartość	Wagi b *Wartość
NO ₂ _WS (1g)	0,63913	85,3400	54,54294
CO_WS (1g)	10,03068	3,1200	31,29573
PM ₁₀ _WS (1g)	0,05276	228,9600	12,08016
W. wolny			-6,40217
Przewidyw.			91,51666
-95,0%GU			88,75527
+95,0%GU			94,27805

Przewidywana zmienna z 95% dokładnością będzie znajdowała się w przedziale [88,75;94,27]. Sama przewidziana wartość jak i przedział są większe niż w przypadku regresji liniowej.

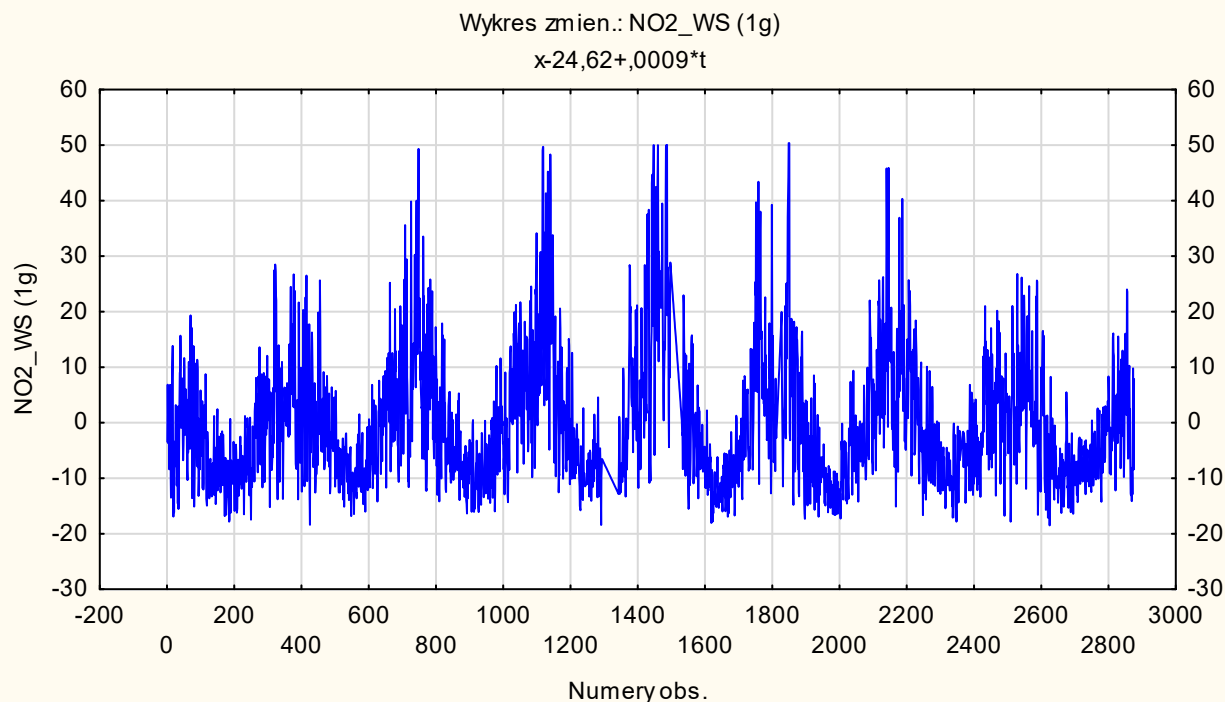
4. Analiza szeregów czasowych

Szereg czasowy to ciąg obserwacji uporządkowanych w czasie, których dokonuje się w określonych (zwykle stałych) jednostkach czasu. Główne cele analizy szeregów czasowych to: badanie wzorców w danych oraz przewidywanie przyszłych wartości.

Szereg czasowy składa się z dwóch podstawowych elementów - **trendu i sezonowości**. Pierwsze pojęcie odnosi się do długotrwałych i systematycznych zmian wielkości danego zjawiska, czyli przykładowo wzrostu lub spadku wartości zmiennej w dłuższym okresie czasu. Natomiast sezonowość odnosi się do rytmicznych i powtarzalnych zmian wielkości danego zjawiska, które charakteryzują się różną długością całego cyklu, przykładowo rok, kwartał, miesiące lub tygodnie.

4.1. Trend szeregu czasowego

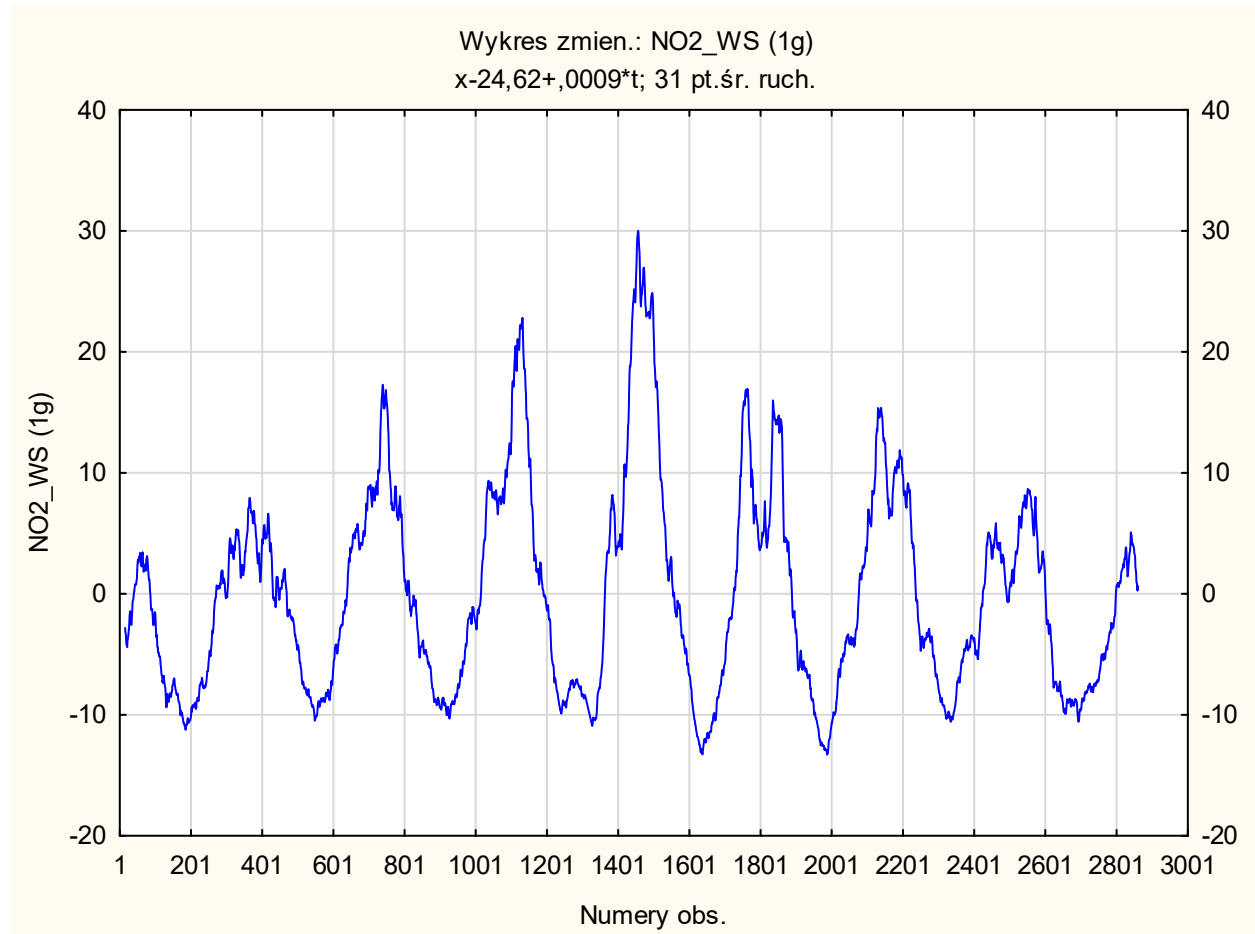
Odjęcie trendu polega na usunięciu długoterminowej tendencji z danych. Odbywa się to poprzez odjęcie przewidywanego trendu od rzeczywistych danych. Dzięki temu możemy zobaczyć np. sezonowość.



Rys. 15. Wykres zmiennej NO₂ Wodzisław Śląski

4.1.1. Wygładzenie

Wygładzanie pozbywa się szumu (nieprzewidywalne czynniki losowe, np. anomalie pogodowe, wybuch). W przykładzie użyłam metody średniej ruchomej, która zastępuje każdy element w sekwencji średnią n sąsiednich wartości, gdzie n jest szerokością okna wygładzania. W moim przypadku $n=31$, czyli 'tracimy' 15 pierwszych i 15 ostatnich wartości na rzecz ich średniej.

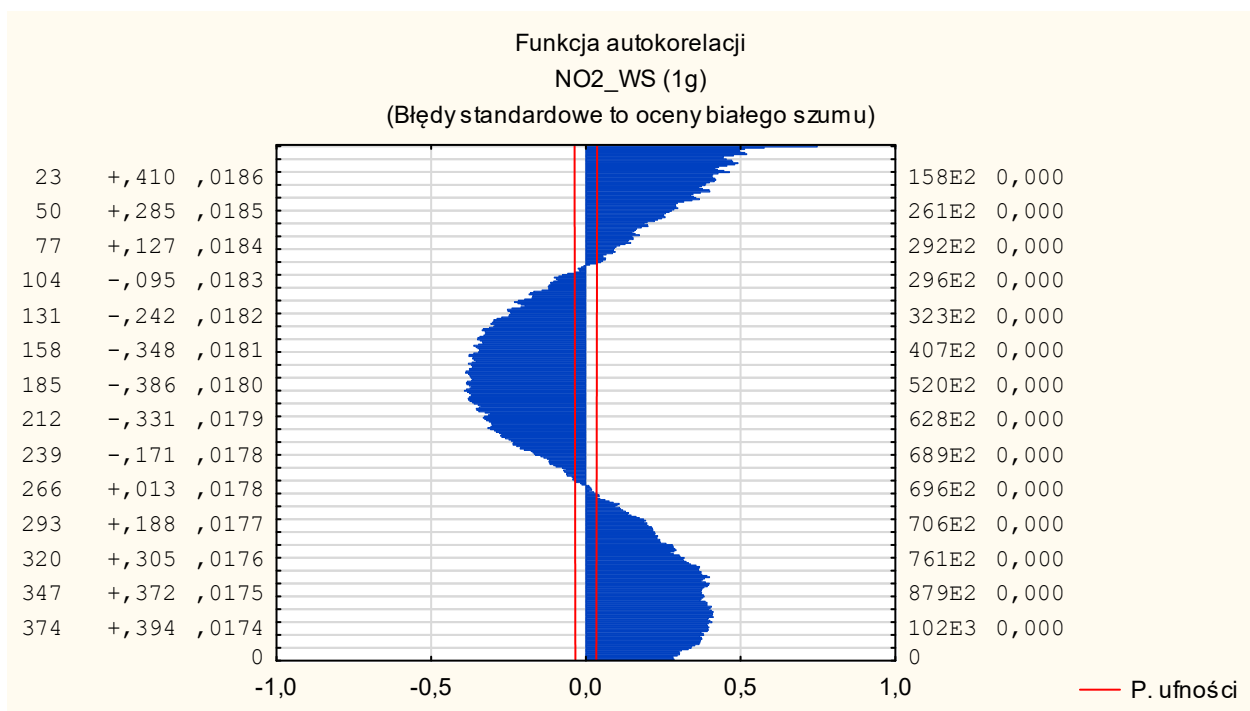


Rys. 16. Wykres z wygładzeniem zmiennej NO2 Wodzisław Śląski

4.2. Autokorelacja szeregu

4.2.1. Dla danych bez wygładzenia

Autokorelacja określa w jakim stopniu poszczególne wartości w szeregu są ze sobą skorelowane. Jako opóźnienie bierzemy 400. Chcemy określić ile trwa cykl. Początkowe wartości są nieistotne statystycznie.



Rys. 17. Funkcja autokorelacji dla danych bez wygładzenia

Sortowanie kolumny autokorelacje malejąco.

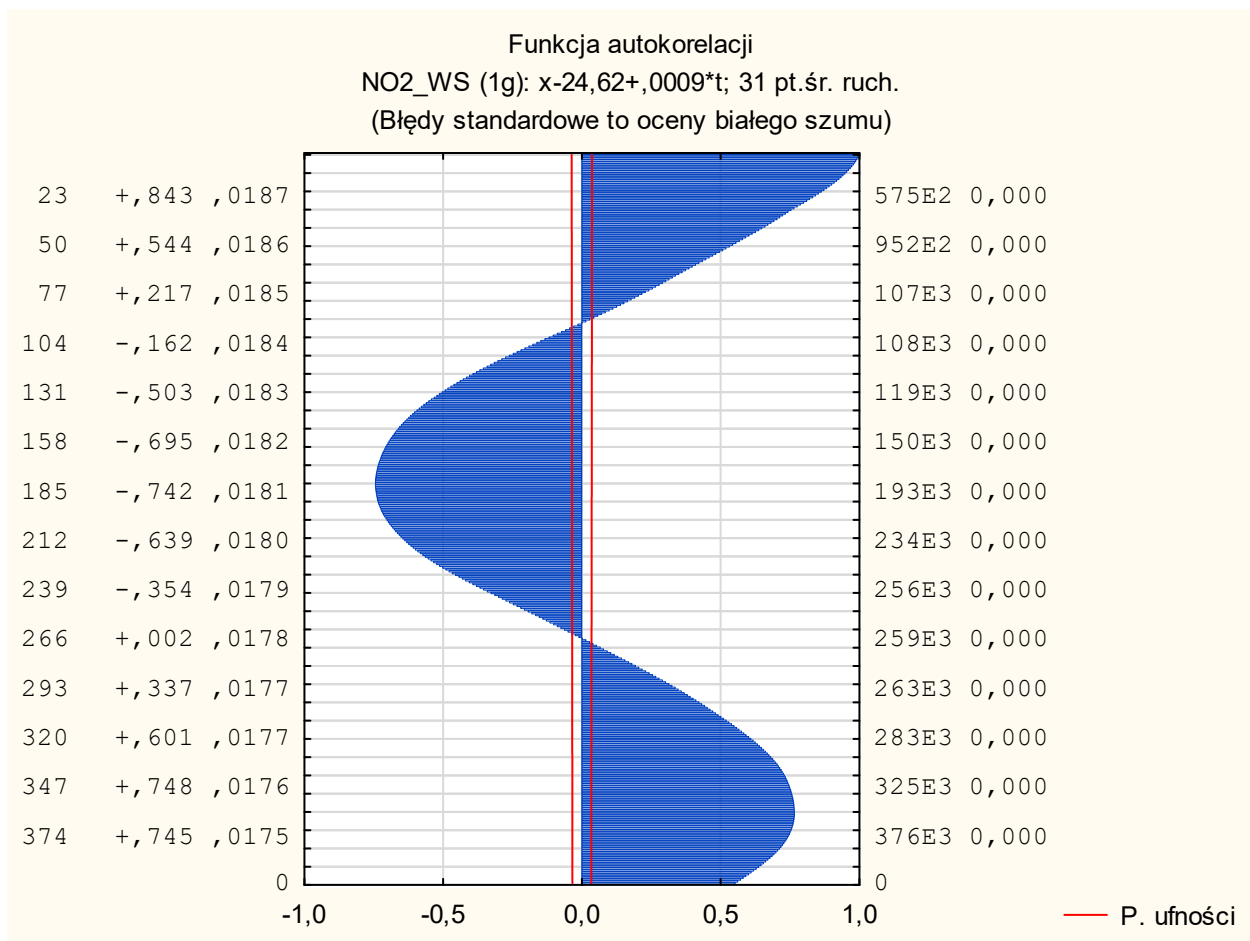
Tabela 24. Funkcja autokorelacji bez wygładzenia

Funkcja autokorelacji (BEZ EKSTREMALNYCH) NO2_WS (1g): x-24,62+0,009*t (Błędy standardowe to oceny białego szumu)				
Opóźn.	Auto-kor.	Błąd std	Q Boxa i Ljunga	p
1	0,746934	0,018637	1606,2	0,00
2	0,575379	0,018634	2559,7	0,00
7	0,518197	0,018618	6257,1	0,00
6	0,512143	0,018621	5482,4	0,00
3	0,511062	0,018631	3312,2	0,00
5	0,496023	0,018624	4726,0	0,00
4	0,494414	0,018627	4016,6	0,00
14	0,490391	0,018595	10613,7	0,00
15	0,479272	0,018592	11278,2	0,00
8	0,475445	0,018614	6909,5	0,00
13	0,472893	0,018598	9918,2	0,00
12	0,470147	0,018601	9271,7	0,00
21	0,462713	0,018572	14616,4	0,00
16	0,454160	0,018588	11875,2	0,00
11	0,452358	0,018605	8632,9	0,00
20	0,444589	0,018575	13995,7	0,00
9	0,442962	0,018611	7476,0	0,00
10	0,442575	0,018608	8041,7	0,00
17	0,427344	0,018585	12403,9	0,00
22	0,425288	0,018569	15141,0	0,00
19	0,423436	0,018579	13422,8	0,00
27	0,417817	0,018553	17587,7	0,00
18	0,415293	0,018582	12903,4	0,00
26	0,414861	0,018556	17080,5	0,00
28	0,413154	0,018549	18083,8	0,00
367	0,410240	0,017410	98611,0	0,00

Możemy zauważyć że większy przeskok jest z opóźnienia 28 na 367 i tyle też trwa nasz cykl. Jest to wartość większa niż faktyczny rok.

4.2.2. Dla danych z wygładzeniem

Jako opóźnienie przyjmujemy również 400. Początkowe wartości są nieistotne statystycznie.



Rys. 18. Funkcja autokorelacji dla danych z wykładzeniem

Sortowanie kolumny autokorelacje malejąco.

Tabela 25. Funkcja autokorelacji z wykładzeniem

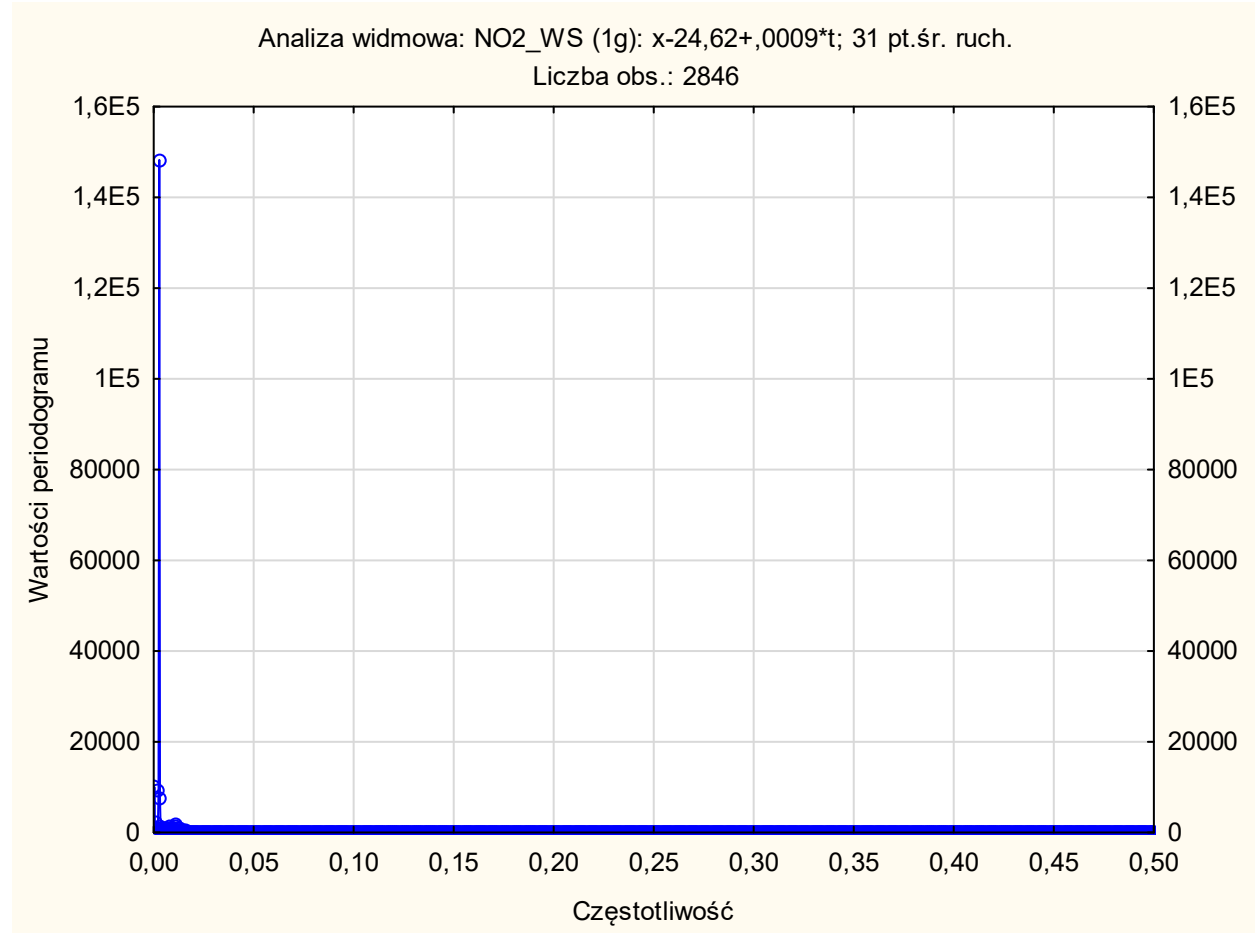
Funkcja autokorelacji (BEZ EKSTREMALNYCH)				
NO2_WS (1g): $x-24,62+0,0009 \cdot t$; 31 pt.śr. ruch.				
(Błędy standardowe to oceny białego szumu)				
Opóźn.	Auto-kor.	Błąd std	Q Boxa i Ljunga	p
1	0,998728	0,018735	2841,8	0,00
2	0,995957	0,018732	5668,8	0,00
3	0,992388	0,018728	8476,5	0,00
4	0,988328	0,018725	11262,3	0,00
5	0,983905	0,018722	14024,3	0,00
6	0,979113	0,018719	16760,3	0,00
7	0,973817	0,018715	19467,8	0,00
8	0,967945	0,018712	22143,6	0,00
9	0,961652	0,018709	24785,8	0,00
10	0,955125	0,018705	27393,0	0,00
11	0,948456	0,018702	29964,9	0,00
12	0,941595	0,018699	32500,7	0,00
13	0,934385	0,018695	34998,6	0,00
14	0,926738	0,018692	37456,7	0,00
15	0,918562	0,018689	39872,4	0,00
16	0,909939	0,018686	42243,9	0,00
17	0,901017	0,018682	44569,9	0,00
18	0,891928	0,018679	46850,0	0,00
19	0,882709	0,018676	49084,0	0,00
20	0,873298	0,018672	51271,4	0,00
21	0,863559	0,018669	53411,0	0,00
22	0,853393	0,018666	55501,3	0,00
23	0,842921	0,018662	57541,4	0,00
24	0,832263	0,018659	59530,8	0,00
25	0,821522	0,018656	61470,0	0,00
26	0,810678	0,018653	63358,9	0,00
27	0,799645	0,018649	65197,5	0,00
28	0,788386	0,018646	66985,3	0,00
29	0,776928	0,018643	68722,1	0,00
362	0,766700	0,017506	353399,7	0,00
361	0,766650	0,017510	351481,6	0,00

Przeskok występuje z opóźnienia 29 na 362 i to jest właśnie nasz cykl. Największa wartość autokorelacji 0,41 jest dla 362, czyli długości cyklu. Z kolei wartość najmniejsza to mniej więcej połowa roku, w tym wypadku to jest autokorelacja wynosząca -0,4 dla 191.

4.3. Analiza Fouriera szeregu czasowego

Analiza Fouriera składa się z dwóch etapów: transformaty Fouriera i analizy widmowej. Ten pierwszy, przekształca sygnał czasu na częstotliwości. Ta transformacja jest reprezentowana przez funkcje amplitudy i fazy różnych składowych częstotliwości. Analiza widma polega na interpretacji tych wartości i identyfikacji składowych częstotliwościowych odpowiedzialnych za charakterystykę sygnału.

4.3.1. Periodogram



Rys. 19. Periodogram

Pik na diagramie odpowiadają okresom, które najbliższej korelują z danymi, co możemy odczytać z podsumowania.

W podsumowaniu analizy widmowej sortujemy kolumnę periodogram malejąco.

Tabela 26. Podsumowanie analizy widmowej

Analiza widmowa: NO2_WS (1g): x=24,62+0,0009*t; 31 pt.śr. ruch. (Arkuszy9_(Odzyskany))						
Liczba obs.: 2846						
	Częstotliwość	Okres	Cosinus współcz.	Sinus współcz.	Periodogram	Gęstość Hamminga wagi
8	0,002811	355,750	9,93505	2,34121	148257,3	70329,10
1	0,000351	2846,000	-2,63869	0,25277	9998,8	4931,94
7	0,002460	406,571	-2,08003	-1,47182	9239,2	40391,01
9	0,003162	316,222	-1,63036	-1,64659	7640,6	39768,17
3	0,001054	948,667	-1,20463	-0,00680	2065,0	1443,18
32	0,011244	88,938	-1,02904	-0,19653	1561,8	1132,58
31	0,010892	91,806	0,91705	-0,49823	1549,9	1202,75
24	0,008433	118,583	-0,93303	0,26952	1342,1	877,21
10	0,003514	284,600	0,39855	0,80073	1138,4	7753,75
20	0,007027	142,300	0,85168	0,20166	1090,1	721,11
16	0,005622	177,875	-0,17847	0,83580	1039,4	886,20
23	0,008082	123,739	0,60477	-0,55886	964,9	919,88
37	0,013001	76,919	-0,70175	-0,40666	936,1	570,35
6	0,002108	474,333	-0,26714	0,75899	921,3	8155,51
15	0,005271	189,733	-0,49294	-0,62205	896,4	885,78
5	0,001757	569,200	-0,47938	-0,61306	861,8	1106,64
14	0,004919	203,286	-0,56174	0,51592	827,8	735,14
12	0,004216	237,167	0,66524	0,27909	740,6	569,00
35	0,012298	81,314	-0,28617	-0,66046	737,2	563,45
29	0,010190	98,138	-0,26348	-0,66233	723,0	537,84
18	0,006325	158,111	0,64592	0,18776	643,9	581,76
17	0,005973	167,412	-0,62923	-0,22723	636,9	731,71
22	0,007730	129,364	-0,63505	-0,13273	598,9	712,22
21	0,007379	135,524	-0,24261	-0,55369	520,0	683,37
36	0,012649	79,056	0,59109	0,05076	500,8	638,68
30	0,010541	94,867	-0,54688	0,04368	428,3	802,72
4	0,001405	711,500	0,44520	0,28610	398,5	921,92
13	0,004568	218,923	0,13612	-0,48160	356,4	581,40
11	0,003865	258,727	-0,11799	-0,47520	341,2	890,88
34	0,011947	83,706	0,33671	0,32407	310,8	425,44

Teraz z tabeli możemy odczytać, że okres trwa 355 dni.

4.4. Współczynnik sezonowości

Sezonowość to okresowa zmienność danych statystycznych, która jest powtarzalna w danym okresie, w tym wypadku rok. Co oznacza, że wskaźnik sezonowości będzie się powtarzał co 365 dni. Jako model sezonowy wybrałam addytywny.

Tabela 27. Dekompozycja sezonowa

Dekompozycja sezonowa: Addytyw. sezon (365) (BEZ EKSTREMALNYCH)							
NO2_WS (1g)							
Obs.	NO2_WS (1g)	Średnie ruchome	Różnice	Wskaźnik sezonow.	Skoryg. Szereg	Wyrówn. Trend-c.	Składnik losowy
1	21,17391			14,1409	7,03299	11,19800	-4,1650
2	31,47826			12,6036	18,87468	10,74867	8,1260
3	26,65217			20,3138	6,33835	9,85001	-3,5117
4	20,91304			14,6640	6,24902	10,86798	-4,6190
5	24,52174			12,1671	12,35462	12,13894	0,2157
6	30,56522			7,7611	22,80416	14,78957	8,0146
7	16,21739			8,4755	7,74189	15,16774	-7,4258
8	29,30435			11,0528	18,25157	16,94446	1,3071
9	31,43478			12,6169	18,81786	16,13668	2,6812
10	30,21739			8,3956	21,82178	12,90635	8,9154
11	14,21739			13,3294	0,88799	5,85549	-4,9675
12	11,09524			18,0667	-6,97144	0,94151	-7,9130
13	15,30435			13,7875	1,51688	1,99986	-0,4830
14	14,47826			11,7218	2,75642	7,40495	-4,6485
15	32,73913			11,7490	20,99013	15,06475	5,9254
16	35,47619			15,1435	20,33269	17,96418	2,3685
17	38,43478			13,5175	24,91724	16,19998	8,7173
18	20,73913			14,6308	6,10835	10,20689	-4,0985
19	7,69565			10,5198	-2,82414	4,70218	-7,5263
20	21,26087			12,2428	9,01803	2,08583	6,9322
21	8,45455			12,8325	-4,37798	0,11587	-4,4939
22	20,34783			20,3336	0,01425	-0,35889	0,3731
23	19,91304			20,9767	-1,06366	-0,66603	-0,3976
24	20,56522			21,9727	-1,40745	1,38498	-2,7924
25	19,39130			15,0302	4,36113	3,76176	0,5994
26	27,78261			17,7047	10,07795	6,53617	3,5418
27	16,73913			12,2440	4,49514	7,13835	-2,6432
28	17,82609			5,5395	12,28663	6,98886	5,2978
29	13,30435			11,6349	1,66944	5,86080	-4,1914
30	15,95455			12,3218	3,63275	4,43359	-0,8008
31	21,36364			9,9587	11,40498	4,54922	6,8558
32	10,04348			19,4749	-9,43143	3,64808	-13,0795
33	27,52174			10,8658	16,65596	6,33962	10,3163
34	9,04348			7,6711	1,37239	7,65879	-6,2864
35	11,94390			6,5944	11,59484	11,59484	6,2544

Wartości wskaźnika addytywnego pokazują, jak bardzo dane statystyczne odbiegają od średniej sezonowej. Wartości ujemne oznaczają, że dane statystyczne są mniejsze od średniej sezonowej, a wartości dodatnie oznaczają, że dane statystyczne są większe od średniej sezonowej.