Charakterystyka próby. Doświadczenia jednoczynnikowe – analiza wariancji i testy szczegółowe
Weryfikacja hipotez dotyczących jednorodności wielu wariancji rozkładów normalnych
Test nieparametryczny Kruskala- Wallisa

Estymatory punktowe średniej, wariancji i odchylenia standardowego

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$\hat{\sigma}^2 = \hat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$$

$$\hat{s} = \sqrt{\hat{s}^2}$$

Współczynnik zmienności

$$v = \frac{\hat{s}}{\bar{x}} \cdot 100\%$$

Zadanie1 1. W ramach gospodarowania zasobami środowiska przyrodniczego, wykonano pomiary zanieczyszczenia powietrza. Badano poziom stężenia fluoru w mg/m^3 na terenie Lubonia w lipcu 2010 roku, jakie występowało tam w związku z działalnością zakładów chemicznych Luvena. Pomiary wykonane zostały przez Wojewódzką Stację Sanitarno-Epidemiologiczną w Poznaniu. Za pomocą pakietu STATISTICA podaj i **zinterpretuj** statystyki opisowe dla poziomu stężenia fluoru w powietrzu w badanym okresie dla Lubonia. Wykonaj histogram przedstawiający empiryczny rozkład cechy.

3,9	5,8	5,5	4,6	6,1	4,8	5,8	4,1	4,5	4,9
5,2	5,3	6,2	6	5,7	4,5	4,1	2,4	4,7	4,4
3,9	4,9	4,2	3,6	4,1	3,5	8	4,4	7,7	8,2

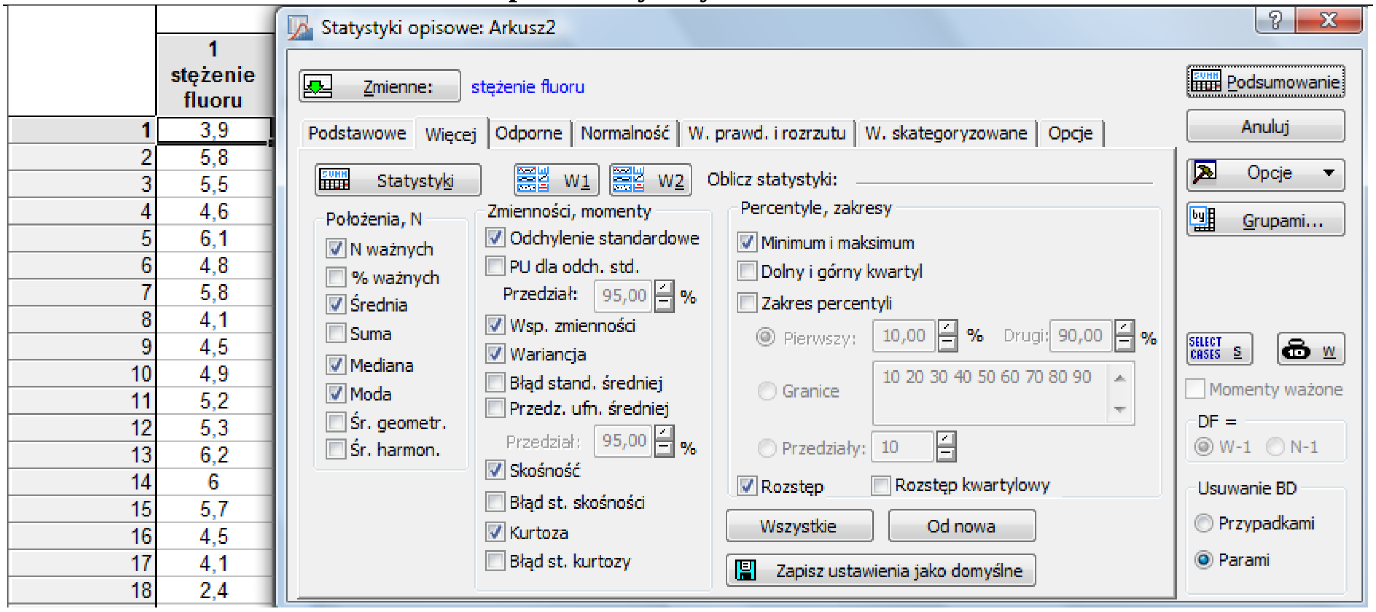
Obliczone statystyki umieść w tabeli.

Charakterystyki próby	Stężenie fluoru w powietrzu w mg/m^3
$\bar{x}$ (średnia)	
m_e (mediana)	
m_o (moda)	
$x_{\min}$ (minimum)	
$x_{\max}$ (maximum)	
R (rozstęp)	
s (odchylenie standardowe)	
v (współczynnik zmienności)	
($\bar{x}$ - s, $\bar{x}$ + s)	

W celu obliczenia większości tych miar wpisujemy dane do jednej kolumny (zmiennej) oraz nadajemy nazwę zmiennej (klikając nagłówek dwa razy) do arkusza pakietu STATISTICA (patrz rysunek 1).

Następnie postępujemy według schematu: **STATYSTYKA → STATYSTYKI PODSTAWOWE I TABELE → STATYSTYKI OPISOWE → OK → wprowadzamy zmienną → klikamy w zakładkę WIĘCEJ → Zaznaczamy interesujące nas wielkości i klikamy PODSUMOWANIE**

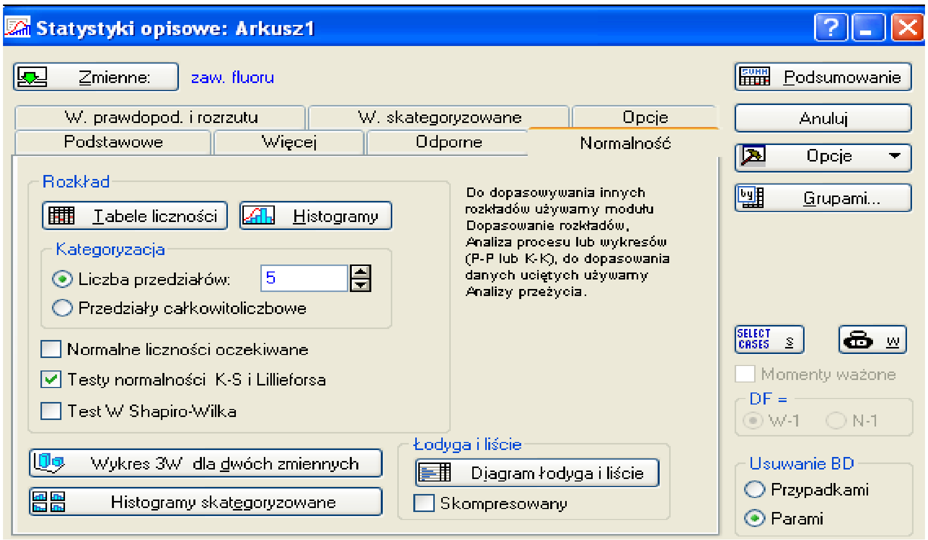
Charakterystyka próby. Doświadczenia jednoczynnikowe – analiza wariancji i testy szczegółowe Weryfikacja hipotez dotyczących jednorodności wielu wariancji rozkładów normalnych Test nieparametryczny Kruskala- Wallisa



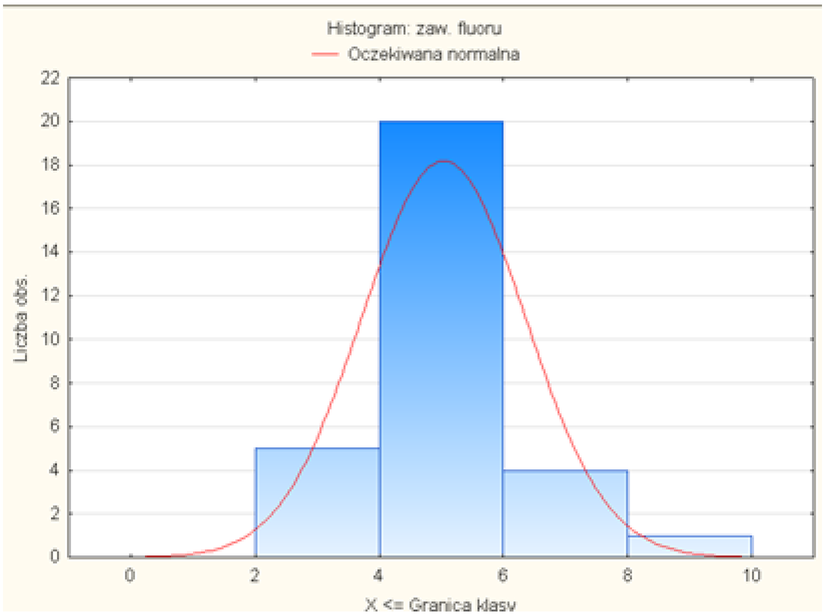
Do skoroszytu i raportu zapisują nam się wyniki w postaci tabelki:

Zmienna	Statystyki opisowe Nważnych	Średnia	Mediana	Moda	Liczność Mody	Minimum	Maksimum	Rozstęp	Wariancja	Odch.std	Wsp.zmn.
stężenie fluoru	30	5,0333	4,7500	4,10	3	2,40	8,20	5,80	1,7340	1,3168	26,16205

Aby sporządzić histogram postępujemy według schematu: **STATYSTYKA → STATYSTYKI PODSTAWOWE I TABELE → STATYSTYKI OPISOWE → wprowadzamy zmienne → klikamy w zakładkę NORMALNOŚĆ (jak wyżej) → zmieniamy ewentualnie liczbę przedziałów ($\approx \sqrt{n}$) i wybieramy HISTOGRAM.**
Otrzymujemy:



Charakterystyka próby. Doświadczenia jednoczynnikowe – analiza wariancji i testy szczegółowe
Weryfikacja hipotez dotyczących jednorodności wielu wariancji rozkładów normalnych
Test nieparametryczny Kruskala- Wallisa



Interpretacja:
Zbadano stężenie fluoru w powietrzu w Luboniu w lipcu 2010 roku, uzyskując 30 obserwacji. Wartość średnia stężenia fluoru wyniosła 5,0333 mg/m^3. Mediana wyniosła 4,75 mg/m^3, co oznacza, że w próbie połowa obserwacji ma wartość mniejszą bądź równą 4,75 mg/m^3 i tyle samo obserwacji ma wartość większą bądź równą 4,75 mg/m^3. Wartość modalna (moda), czyli wartość obserwowana najczęściej wynosi 4,1 mg/m^3, przy czym wartość ta występuje trzy razy (liczność mody wynosi 3).
Z relacji $m_o < m_e < \bar{x}$ wynika, że rozkład empiryczny cechy jest prawostronnie asymetryczny (przypomnijmy, że dla zmiennej o rozkładzie normalnym wartości tych 3 miar położenia są jednakowe).

Najmniejszą obserwowaną wartością stężenia fluoru wśród 30 obserwacji było 2,4 mg/m^3, a największą obserwowaną wartością tego pestycydu było 8,2 mg/m^3. Rozstęp, czyli zakres zmienności obserwacji wyniósł 5,8 mg/m^3. Wariancja i odchylenie standardowe stężenia fluoru wynoszą odpowiednio 1,734 (mg/m^3)^2 oraz 1,3168 mg/m^3. Stąd współczynnik zmienności wynosi około 26 % (>10 %), co oznacza, ze obserwowana zmienna nie jest stabilna.

Zadanie 2. W ramach monitoringu wdrażania strategii rozwoju gminy, przeprowadzono badanie w wybranych losowo gospodarstwach rolniczych o powierzchni powyżej 30 ha zużycia energii elektrycznej w kilowatogodzinach na 1 ha użytków rolnych uzyskano następujące wyniki: 335; 196; 220; 113; 232; 205; 160; 263; 302; 221; 121; 245; 232; 115. Wyznacz i zinterpretuj średnią arytmetyczną, rozstęp, wariancję, odchylenie standardowe, współczynnik zmienności, medianę, modę, współczynnik skośności i spłaszczenia.

	Statystyki opisowe									
Zmienna	Średnia	Mediana	Moda	Liczność Mody	Rozstęp	Wariancja	Odch.std	Wsp.zmn.	Skośność	Kurtoza
zużycie energii	211,4286	220,50	232,00	2	222,00	4489,19	67,00	31,69	0,0281	-0,419532

Zadanie 3. Otwórz plik temperatura umieszczony w katalogu DYDAKTYKA / Budka.
Scharakteryzuj temperaturę średnią dla miesiąca czerwca i dla lipca na podstawie zamieszczonych tam obserwacji z okresu 228 lat. Wykonaj histogramy rozkładu cechy.

Odp. $\bar{x}_{cz} = 17{,}05833$; $\bar{x}_{lip} = 18{,}7663$

Charakterystyka próby. Doświadczenia jednoczynnikowe – analiza wariancji i testy szczegółowe
Weryfikacja hipotez dotyczących jednorodności wielu wariancji rozkładów normalnych
Test nieparametryczny Kruskala- Wallisa

Zadanie 4. Elementem prowadzonej polityki sektorowej przemysłu, w rejonie składowiska odpadów Tarnów-Krzyż w ramach monitoringu zanieczyszczenia było badanie gleby. Pobrano 14 próbek do badań geochemicznych z trzech profili w celu zbadania stężenia pierwiastków w badanej glebie. Dane zebrano w poniższej tabeli. Korzystając z pakietu STATISTICA oblicz poznane miary położenia i rozrzutu dla otworu 1 (tło) oraz łącznie dla otworów 2 i 3 (oddziaływanie składowiska), także łącznie dla wszystkich próbek dla wszystkich oznaczanych pierwiastków.

Symbol próbki: Nr otworu	Głębokość pobrania próbki (m)	**Bar**	**Chrom**	**Cynk**	**Kobalt**	**Mangan**	**Miedź**	**Nikiel**	**Ołów**	**Rtęć**	**Stront**
1	**0,4**	22	10	22	4	159	7,2	6	6	0,034	7
1	**0,8**	25	13	23	2	42	10,9	8	6	0,033	9
1	**1,1**	42	13	35	10	248	15,3	22	8	0,038	11
1	**2**	59	18	65	9	541	21,7	29	12	0,066	98
1	**3**	31	16	63	13	943	20,7	39	12	0,07	100
1	**3,8**	33	14	57	9	954	19,2	29	11	0,059	98
2	**0,4**	32	10	28	6	493	9,4	7	8	0,029	7
2	**1**	27	13	29	4	208	10,8	11	7	0,034	9
2	**1,7**	32	11	31	5	241	13,2	17	7	0,026	41
2	**2,7**	26	10	30	5	343	12,9	17	7	0,029	62
2	**3,5**	29	11	37	7	472	15,1	22	8	0,037	67
3	**0,8**	63	19	65	14	944	19,6	31	14	0,078	14
3	**2,5**	36	15	61	10	984	20,1	29	11	0,062	99
3	**3,6**	40	13	62	10	902	17,3	31	11	0,061	94

Dane pochodzą z rozprawy doktorskiej p. K. Sobiak pod tytułem „Badanie wpływu składowisk odpadów na środowisko gruntowo-wodne na przykładzie wybranych obiektów zlokalizowanych w obrębie zlewni Dunajca."

Danych jest v populacji normalnych N(μ_i; σ_i), i=1,2,...,v. Z każdej populacji wylosowano próbkę n_i elementową. Formułuje się hipotezę:

$H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_v^2$

$H_1 : \sim H_0$.

Test Bartletta : $$X^2 = \frac{(n-v)\ln[\frac{1}{n-v}\sum_{i=1}^{v}(n_i-1)\hat{s}_i^2] - \sum_{i=1}^{v}(n_i-1)\ln[\hat{s}_i^2]}{1+\frac{1}{3(v-1)}[\sum_{i=1}^{v}(\frac{1}{n_i-1}) - \frac{1}{n-v}]}$$

Jeżeli $X^2 > X^2_{\alpha,v-1}$ hipotezę H_0 odrzucamy.

UKŁAD CAŁKOWICIE LOSOWY

Model obserwacji (model analizy wariancji) w doświadczeniu jednoczynnikowym o układzie całkowicie losowym

$$y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \text{ (wyraz wolny + efekt obiektowy + błąd doświadczalny)} ,$$

gdzie i =1,2,...,v, j=1,2,...,r_i oraz $\sum_{i=1}^{v}\alpha_i = 0$, przy czym spełnione są założenia, które zapisujemy krótko w postaci y_{ij}~N($\mu+\alpha_i;\sigma$) lub równoważnie w postaci ε_{ij}~N(0;σ).
Stawiamy hipotezę :

H_0: $\mu_1 = \mu_2 = ... = \mu_v$

H_1:~H_0

Przeprowadzamy analizę wariancji (układ całkowicie losowy), wyznaczamy statystykę F :

Źródła zmienności	Stopnie swobody	Sumy kwadratów SS	Średnie kwadraty MS	F
Obiekty	v-1	SS_O	MS_O	F_0
Błąd	n-v	SS_E	MS_E	
Całkowita	n-1	SS_C		

Wzory na końcu materiałów. Jeśli $F_0 > F_{\alpha;v-1;n-v}$ to hipotezę H_0 odrzucamy.

Po odrzuceniu hipotezy H_0 stawiamy hipotezy szczegółowe dla $i.i'$=1,…,v, $i \neq i'$:

$H_{0ii'}$: $\mu_i - \mu_{i'} = 0$

$H_{1ii'}$:~ $H_{0ii'}$

Weryfikację tych hipotez przeprowadzamy:

- testem wielokrotnym **Tukeya** jako testem najbardziej polecany do porównywania par średnich jego celem jest ustalenie, które grupy w próbce różnią się.
- testem jednokrotnym **Fishera**, stosowanym najczęściej tylko do wybranych porównań średnich.

(wzory na końcu materiału)

Zadanie 5 Ze względu na szczególną rolę uzdrowiskowo-wypoczynkową pięciu nadmorskich miast, porównywano powierzchnię zieleni ogólnodostępnej (w ha) według uchwalonych planów miejscowych w miastach A, B, C, D i E na przestrzeni kilu lat . Przeprowadzono odczyty w tych latach i otrzymano:

A	9,5	9,0	9,6	9,5
B	9,0	9,0	9,2	
C	8,9	8,7	8,8	
D	8,8	9,2	9,0	9,0
E	8,5	8,6	8,5	8,7

Przy prawdziwości założenia analizy wariancji (założenia o rozkładzie normalnym zmiennej obserwowanej) :

a) Na poziomie istotności $\alpha = 0{,}01$ zweryfikować hipotezę o jednorodności wariancji dla powierzchni zielonych w badanych miastach. Zastosować test Bartletta.

b) Przy prawdziwości założeń analizy wariancji na poziomie istotności $\alpha = 0{,}01$ zweryfikować hipotezę ogólną o braku istotnych różnic między rzeczywistą średnią powierzchnię zieleni ogólnodostępnej w badanych miast na przestrzeni interesujących lat.

c) Wykorzystać test **Fishera** do porównań średnich powierzchni zielonych parami. Postawić i zweryfikować odpowiednie hipotezy szczegółowe. Przyjąć poziom istotności α = 0,01.

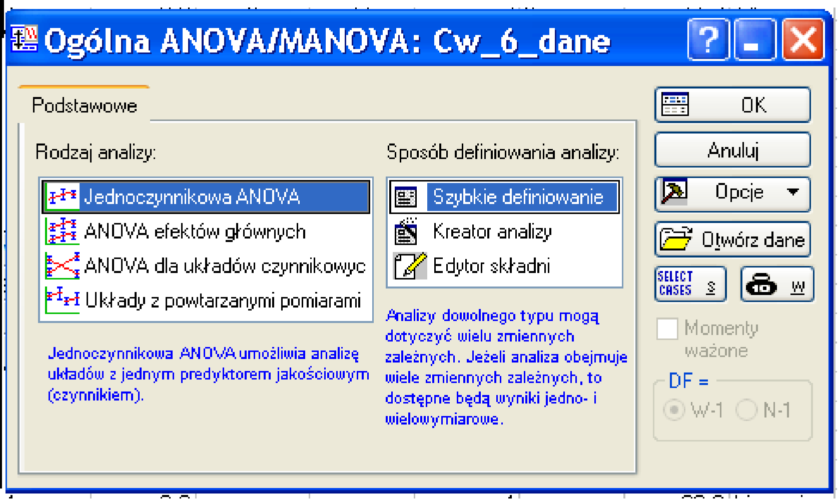
Rozwiązanie:

Wypełniamy arkusz danych

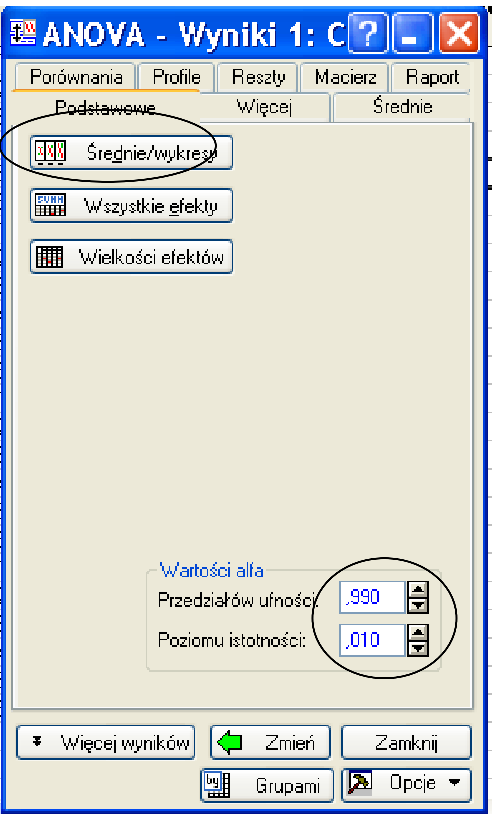
Za pomocą programu STATISTICA wybieramy : **STATYSTYKA → ANOVA → JEDNOCZYNNIKOWA ANOVA → SZYBKIE DEFINIOWANIE → OK**

Charakterystyka próby. Doświadczenia jednoczynnikowe – analiza wariancji i testy szczegółowe
Weryfikacja hipotez dotyczących jednorodności wielu wariancji rozkładów normalnych
Test nieparametryczny Kruskala- Wallisa

	3 obiekty (miasta)	4 pow. ziel. w ha
1	A	9,5
2	A	9
3	A	9,6
4	A	9,5
5	B	9
6	B	9
7	B	9,2
8	C	8,9
9	C	8,7
10	C	8,8
11	D	8,8
12	D	9,2
13	D	9
14	D	9
15	E	8,5
16	E	8,6
17	E	8,5
18	E	8,7



→ ustalamy zmienne w obrębie pola **Listy zmiennych zależnych** zaznaczamy zmienną obserwacji, czyli *powierzchnia zieleni*, a w polu **Predyktor jakościowy (czynnik)** wybieramy zmienną *obiekty* i akceptujemy **OK→OK**. Pojawia się okno **ANOVA-wyniki 1**. Ustalamy w nim poziom istotności 0,01, ufności 0,99.



a) Sprawdzanie założenia o jednorodności wariancji :
Stawiamy hipotezy: zerową i alternatywną

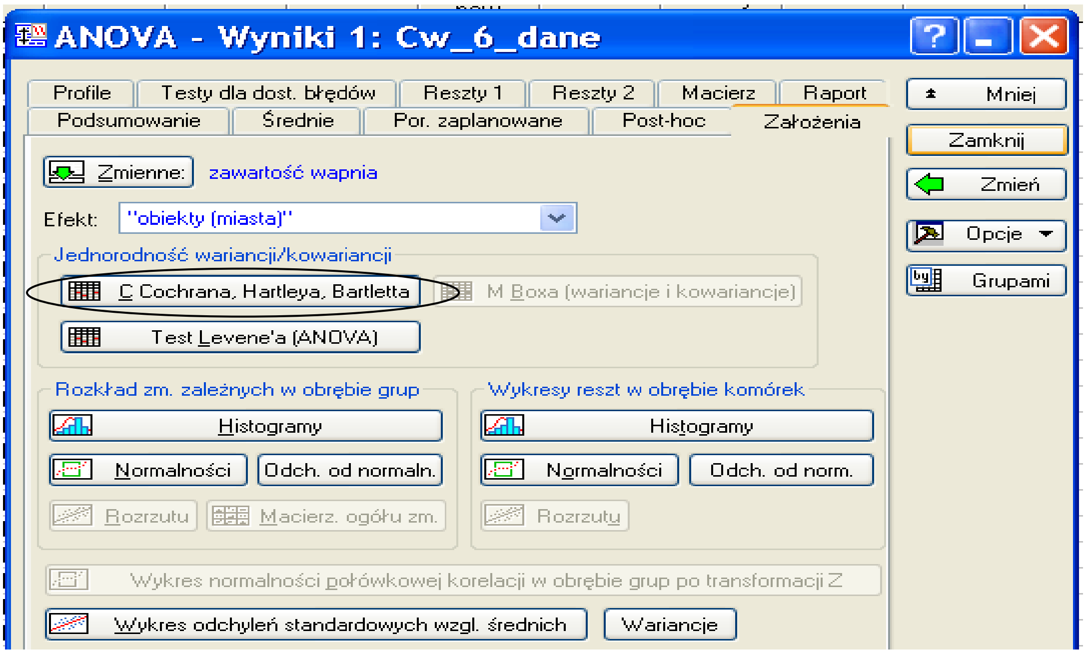
$H_0: \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2$ rzeczywista wariancja powierzchni zielonej w badanych miast **nie** różni się między sobą istotnie

$H_1: \sim H_0$ co najmniej dwie wariancje dla powierzchni zielonych w badanych miast **różnią się** między sobą istotnie

Za pomocą programu STATISTICA

Charakterystyka próby. Doświadczenia jednoczynnikowe – analiza wariancji i testy szczegółowe
Weryfikacja hipotez dotyczących jednorodności wielu wariancji rozkładów normalnych
Test nieparametryczny Kruskala- Wallisa

W oknie **ANOVA – Wyniki 1** → wybieramy zakładkę **WIĘCEJ WYNIKÓW** → następnie zakładkę **ZAŁOŻENIA** i w niej wybieramy test Bartletta, klikając na przycisk między innymi z tym testem.



Otrzymujemy:

	Testy jednorodności wariancji (Zadanie 6-1.sta) Efekt: "obiekty (miasta)"				
	Hartleya F-maks	Cochrana C	Bartlett Chi-kw.	a df	p
zawartość wapnia	8,00000C	0,553459	3,88626:	4	0,421617

Decyzja i wniosek:

Ponieważ p = 0,421617 > α = 0,01, to na poziomie istotności 0.01 nie mamy podstaw do odrzucenia hipotezy zerowej, orzekającej o jednorodności wariancji.

b) Stawiamy hipotezy: zerową i alternatywną

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ rzeczywiste średnie powierzchnie zielone w badanych miast **nie** różnią się między sobą istotnie

$H_1 : \sim H_0$ co najmniej dwie średnie powierzchnie zielone w badanych miastach **różnią się** między sobą istotnie

Za pomocą programu STATISTICA : Ponownie wracamy do początkowego okna **ANOVA – Wyniki 1** (klikając przycisk **MNIEJ**) można w nim otworzyć kartę **WIĘCEJ**

Charakterystyka próby. Doświadczenia jednoczynnikowe – analiza wariancji i testy szczegółowe
Weryfikacja hipotez dotyczących jednorodności wielu wariancji rozkładów normalnych
Test nieparametryczny Kruskala- Wallisa

→ klikając w niej przycisk **WYNIKI JEDNOWYMIAROWE**. Otrzymamy tabelę analizy wariancji

Tabela analizy wariancji ANOVA

Efekt	Wyniki jednowymiarowe dla każdej ZZ (Zadanie 6-1.sta) Parametryzacja z sigma-ograniczeniami Dekompozycja efektywnych hipotez				
	Stopnie swobody	pow. ziel. w ha SS	pow. ziel. w ha MS	pow. ziel. w ha F	pow. ziel. w ha p
Wyraz wolny	1	1419,371	1419,371	49314,44	0,000000
obiekty (miasta)	4	1,482	0,370	12,87	0,000186
Błąd	13	0,374	0,029		
Ogół	17	1,856			

<u>Decyzja</u>:
Ponieważ prawdopodobieństwo p = 0,000186 < α = 0,01, to na poziomie istotności α = 0,01 odrzucamy hipotezę H_0 na korzyść hipotezy H_1.

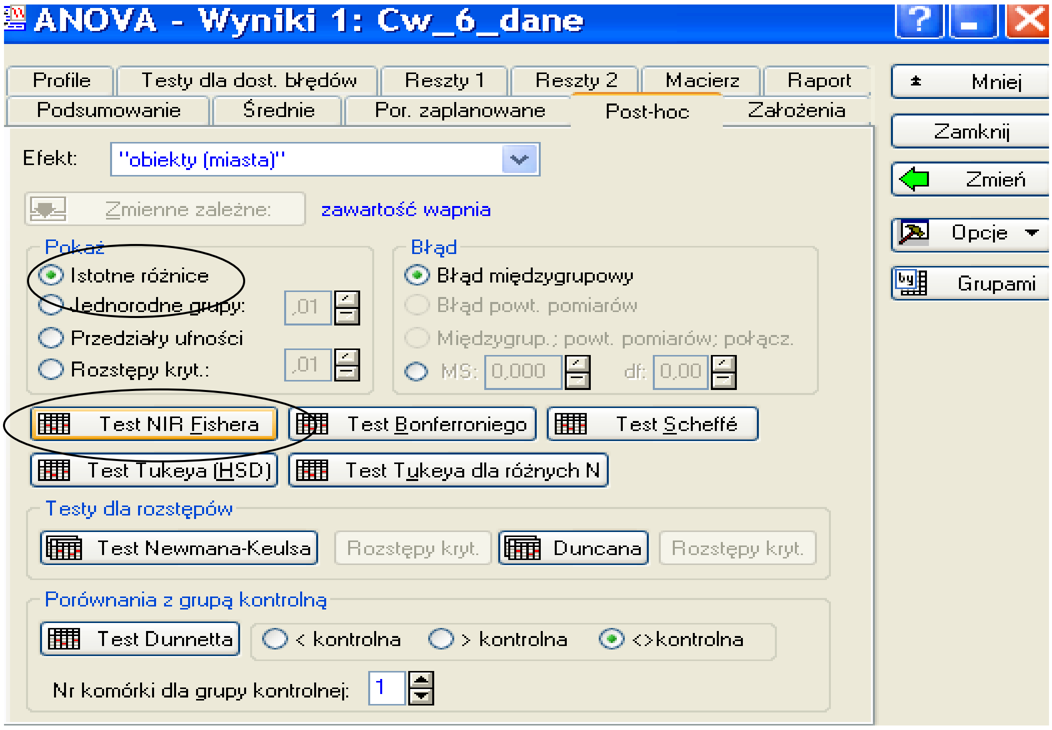
<u>Wniosek:</u>
Na poziomie istotności 0.01 stwierdzamy istotne zróżnicowanie rzeczywistych średnich powierzchni zielonych w badanych miastach.

c) Stawiamy hipotezy szczegółowe:

$$H_{0ii'}: \mu_i - \mu_{i'} = 0 \qquad i, i' = 1, 2, 3, 4, 5$$
$$H_{1ii'}: \sim H_{0ii'} \qquad i \neq i'$$

W celu wykonania dalszej analizy szczegółowej po analizie wariancji wracamy znów do okna **ANOVA – Wyniki 1 → WIĘCEJ WYNIKÓW,** a następnie klikamy opcję **POST-HOC**. W polu **Pokaż** zaznaczamy **Istotne różnice** i wybieramy **TEST NIR FISHERA**.

Charakterystyka próby. Doświadczenia jednoczynnikowe – analiza wariancji i testy szczegółowe
Weryfikacja hipotez dotyczących jednorodności wielu wariancji rozkładów normalnych
Test nieparametryczny Kruskala- Wallisa



Otrzymujemy tabelę, w której prawdopodobieństwa *p* mniejsze od $\alpha = 0{,}01$ oznaczają odrzucenie hipotez zerowych szczegółowych.

	Test NIR; zmienna zawartość wapnia (Zadanie 6-1.sta) Prawdopodobieństwa dla testów post-hoc Błąd: MS międzygrupowe = ,02878, df = 13,000					
Nr podkl.	obiekty (miasta)	{1} 9,4000	{2} 9,0667	{3} 8,8000	{4} 9,0000	{5} 8,5750
1	A		0,023183	0,000471	0,005380	0,000011
2	B	0,023183		0,076377	0,615534	0,002231
3	C	0,000471	0,076377		0,146692	0,106105
4	D	0,005380	0,615534	0,146692		0,003606
5	E	0,000011	0,002231	0,106105	0,003606	

<u>Decyzje:</u>

Na poziomie istotności $\alpha = 0{,}01$, na podstawie testu Fishera, odrzucamy hipotezy szczegółowe dla par miast: A-C (bo p=0,000471 < $\alpha = 0{,}01$), A-D (bo p=0,005380 < $\alpha = 0{,}01$), A-E (bo p=0,000011 < $\alpha = 0{,}01$), B-E (bo p=0,002231 < $\alpha = 0{,}01$), D-E (bo p=0,003606 < $\alpha = 0{,}01$).

<u>Wnioski:</u>

Na poziomie istotności 0,01 stwierdzamy, że prawdziwe średnie powierzchnie zielone wypisanych powyżej par miast różnią się między sobą istotnie.

Charakterystyka próby. Doświadczenia jednoczynnikowe – analiza wariancji i testy szczegółowe
Weryfikacja hipotez dotyczących jednorodności wielu wariancji rozkładów normalnych
Test nieparametryczny Kruskala- Wallisa

TEST KRUSKALA-WALLISA jest jedną z najpopularniejszych alternatyw dla jednoczynnikowej analizy wariancji. Przeprowadzany jest w przypadku, gdy zostały złamane założenia ANOVA bądź gdy charakter naszych zmiennych nie pozwala na wykorzystanie analizy wariancji.
Weryfikujemy hipotezy:
H_0: Rozkłady dla k populacji są takie same
H_1: Nie wszystkie rozkłady są takie same
Wszystkim obserwacjom zmiennej losowej X uszeregowanym niemalejąco nadajemy rangi, czyli kolejne numery od 1 do n (jednakowym wartościom obserwacji przypisujemy liczbę będącą średnią arytmetyczną rang, które kolejno miały następować. Dla każdej próbki obliczamy sumy rang $T_{i,\ i=1,2,\ldots,v.}$
Wyznaczamy wartość statystyki:

$$X^2 = \frac{12}{n(n+1)} \sum_{i=1}^{v} \frac{T_i^2}{n_i} - 3(n+1).$$

Jeżeli H_0 jest prawdziwa to powyższa statystyka ma asymptotyczny rozkład chi-kwadrat o (v-1) stopniach swobody.
Przyjmujemy poziom istotności α i odczytujemy wartość krytyczną $X^2_{\alpha,v-1}$. Jeżeli $X^2 > X^2_{\alpha,v-1}$ hipotezę H_0 odrzucamy.

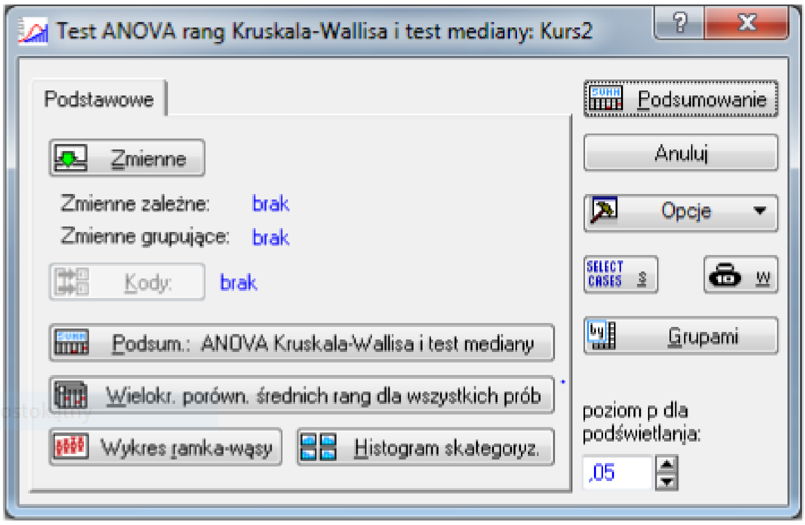
Zadanie 6. Inspektor do spraw polityki przestrzennej i planowania miejscowego, zlecił przeprowadzenie analiz w celu sprawdzenie, czy preferencje studentów różnych kierunków (Gospodarki Przestrzennej, Ochrony Środowiska, Ekonomii, Medycyny) do zamieszkania w małej miejscowości różnią się istotnie na poziomie istotności 0,01. Wyniki ankiet przeprowadzonych pośród 10 studentów każdego kierunku podano w umownej skali sumarycznej.

Grupa ankietowanych	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
Gospodarki Przestrzennej	38	39	48	31	43	37	38	47	33	44
Ochrony Środowiska	56	48	47	54	50	55	48	46	53	49
Ekonomii	44	49	40	40	49	45	48	41	39	50
Medycyny	44	45	48	53	52	43	44	49	52	50

Z racji, że porównywanych grup było więcej niż 2 (4 kierunki studiów) a zmienna zależna (punktacja) była mierzona na <u>skali porządkowej</u> przeprowadzono test Kruskala-Wallisa, aby porównać ze sobą grupy studentów pod względem preferencji zamieszkania w małej miejscowości.
W celu przeprowadzenia analizy, należy wprowadzić dane do arkusza:

Charakterystyka próby. Doświadczenia jednoczynnikowe – analiza wariancji i testy szczegółowe Weryfikacja hipotez dotyczących jednorodności wielu wariancji rozkładów normalnych Test nieparametryczny Kruskala- Wallisa

	1 Kierunek studiów	2 Punktacja
1	Gospodarki Przestrzenna	38
2	Gospodarki Przestrzenna	39
3	Gospodarki Przestrzenna	48
4	Gospodarki Przestrzenna	31
5	Gospodarki Przestrzenna	43
6	Gospodarki Przestrzenna	37
7	Gospodarki Przestrzenna	38
8	Gospodarki Przestrzenna	47
9	Gospodarki Przestrzenna	33
10	Gospodarki Przestrzenna	44
11	Ochrona Środowiska	56
12	Ochrona Środowiska	48



Z menu **Statystyka** opcję **Statystyki nieparametryczne**. Następnie w otwierającym się oknie wybieramy opcję **Porównanie wielu prób niezależnych (grup).** Po kliknięciu na przycisku **OK** otworzy się okno **Test ANOVA rang Kruskala-Wallisa i test mediany** (rysunek powyżej).
Klikając przycisk **Podsumowanie. ANOVA Kruskala-Wallisa** i test mediany otrzymujemy dwa arkusze wyników. Nas interesuje ten dotyczący nieparametrycznej analizy wariancji.

Zależna: Punktacja	Wartość p dla porównań wielokrotnych (dwustronych); Punktacja (Arkusz2) Zmienna niezależna (grupująca): Kierunek studiów Test Kruskala-Wallisa: H (3, N= 40) =18,11213 p =,0004 **Gospodarki Przestrzenna** **R:9,2000**	 Ochrona Środowiska R:30,150	 Ekonomia R:17,850	 Medycyna R:24,800
Gospodarki Przestrzenna		0,000369	0,588142	0,017078
Ochrona Środowiska	0,000369		0,111839	1,000000
Ekonomia	0,588142	0,111839		1,000000
Medycyna	0,017078	1,000000	1,000000	

Decyzje:
Na poziomie istotności $\alpha = 0{,}05$, na podstawie testu Kruskala-Wallisa odrzucamy hipotezę zerową o braku istotnych różnic pomiędzy preferencjami studentów rozważanych kierunków H(3,N=40)=18,11213 (bo p=0,0004 < $\alpha = 0{,}05$). Podane wartości p w tabeli, takie, że p< $\alpha = 0{,}05$ wskazują na pary kierunków, studiów po których preferencje studentów różnią się istotnie.

Wnioski:
Na poziomie istotności 0,05 stwierdzamy, że preferencje studentów z wybranych kierunków różnią się istotnie. Na podstawie testu porównań wielokrotnych stwierdzono różnice w preferencjach dla kierunków: Gospodarka Przestrzenna-Ochrona Środowiska (bo p=0,000369 < $\alpha = 0{,}05$), Gospodarka Przestrzenna-Medycyna (bo p=0,017078 < $\alpha = 0{,}05$).

Charakterystyka próby. Doświadczenia jednoczynnikowe – analiza wariancji i testy szczegółowe
Weryfikacja hipotez dotyczących jednorodności wielu wariancji rozkładów normalnych
Test nieparametryczny Kruskala- Wallisa

Zadanie 7 Porównywano zawartość magnezu (w mg/l) w wodach mineralnych czterech rodzajów (obiektów) dostępnych na polskim rynku. Otrzymane obserwacje zapisano w tabeli (dane pobierz z pliku cw2.sta).

Muszynianka	Muszynianka plus	Piwniczanka	Galicjanka
*	*	*	*
*	*	*	*

Przy prawdziwości założeń analizy wariancji
a) Na poziomie istotności $\alpha = 0{,}01$ zweryfikować hipotezę ogólną o braku istotnych różnic między rzeczywistymi średnimi zawartościami magnezu w badanych wodach mineralnych.
b) Na poziomie istotności $\alpha = 0{,}01$ zweryfikować hipotezę o jednorodności wariancji zawartości magnezu w badanych wodach. Zastosować test Bartletta.
c) Wykorzystać test **Tukeya** do porównań średnich zawartości magnezu w wodach mineralnych parami. W tym celu postawić i zweryfikować odpowiednie hipotezy szczegółowe. Przyjąć α = 0,01.

Odp. **a)** p = 0,0000 < $\alpha = 0{,}01$; **b)** p = 0,038777 > $\alpha = 0{,}01$;

Zadanie 8 Podczas zajęć laboratoryjnych 4-osobowa grupa studentek (S1, S2, S3, S4) uczyła się ustalać kwasowość gleby mierzonej współczynnikiem pH. Każda studentka pobrała losowo 23 próbki gleby i poddała je badaniom chemicznym. Obserwacje pobierz z pliku cw2.sta.
Przy prawdziwości założeń analizy wariancji :
a) Na poziomie istotności α = 0,02 zweryfikować hipotezę ogólną o braku różnic między średnimi kwasowościami gleby oznaczonymi przez poszczególne studentki.
b) Wykorzystać test **Tukeya** do porównań wielokrotnych. W tym celu utworzyć grupy jednorodne średnich na poziomie istotności α = 0,02.
c) Wykonać wykres średnich

Odp. **a)** p = 0,0000; **b)** grupy: (S2, S3), (S1, S4); **c)** graficzna prezentacja średnich potwierdza ich zróżnicowanie, bo wyznaczone 98 % przedziały ufności są rozłączne dla co najmniej jednej pary obiektów.

Zadanie 9 W ramach monitoringu zanieczyszczenia powietrza benzenem w województwie małopolskim zebrano dane dotyczące stężenia benzenu dla 6 miejscowości tego województwa. W każdej miejscowości było 12 stanowisk pomiarowych, dla których obliczono średnie dobowe (w µg/m^3). Dane te znajdują się w pliku cw2.sta.

a) Na poziomie istotności $\alpha = 0{,}05$ zweryfikować hipotezę ogólną o braku istotnych różnic między rzeczywistym poziomem średniodobowego stężenia benzenu w poszczególnych miejscowościach województwa małopolskiego.
b) Wykorzystać test Fishera do porównań rzeczywistych średnich stężeń benzenu w badanych miejscowościach W tym celu postawić i zweryfikować odpowiednie hipotezy szczegółowe. Przyjąć poziom istotności α = 0,05.

Odp. **a)** p = 0,001337;
b) istotne różnice M1 z M3, M1 z M4, M1 z M5, M1 z M6, M2 z M6, M2 z M3.

Zadanie 10 Dokonywano pomiarów zanieczyszczenia powietrza (emisji) na terenie Poznania. Oznaczano m.in. zawartość dwutlenku siarki w μg/m^3/miesiąc. Dla poszczególnych dzielnic otrzymano obserwacje:

Stare Miasto	3,9	4,0	4,3	4,0	
Nowe Miasto	3,2	3,4	3,0	3,4	
Wilda	4,5	4,6	5,2		
Jeżyce	2,5	2,6	2,5	2,8	
Grunwald	3,8	3,7	4,0	3,9	4,3

Charakterystyka próby. Doświadczenia jednoczynnikowe – analiza wariancji i testy szczegółowe
Weryfikacja hipotez dotyczących jednorodności wielu wariancji rozkładów normalnych
Test nieparametryczny Kruskala- Wallisa

a) Na poziomie istotności $\alpha = 0{,}01$ zweryfikować hipotezę ogólną o braku istotnych różnic między rzeczywistymi średnimi zawartościami dwutlenku siarki w dzielnicach Poznania.
b) Wykorzystać test Fishera do porównań średnich zawartości dwutlenku siarki. W tym celu postawić i zweryfikować odpowiednie hipotezy szczegółowe. Przyjąć poziom istotności $\alpha = 0{,}01$.
Odp. **a)** $p = 0{,}000000$; **b)** nie odrzucamy hipotezy dla porównania Stare Miasto - Grunwald

Zadanie 9 Badając wpływ składowania odpadów na glebę pobrano 14 próbek gleby do badań geochemicznych z trzech profili. Określono w nich między innymi stężenie rtęci. Otrzymując obserwacje:

Otwór 1	0,034	0,033	0,038	0,066	0,07	0,059
Otwór 2	0,029	0,034	0,026	0,029	0,037	
Otwór 3	0,078	0,062	0,061			

Przyjmując $\alpha = 0{,}05$ i założenie o normalności rozkładu
a) Sprawdzić czy spełnione jest założenie o jednorodności wariancji.
b) Odpowiedzieć czy badane profile (otwory) różniły się pod względem zawartości badanego pierwiastka.
c) Jeżeli jest to uzasadnione wykorzystać test Tukeya do porównań wielokrotnych w tym celu utworzyć grupy jednorodne.
Odp. **a)** $p = 0{,}066857$; **b)** $p = 0{,}006437$; **c))** grupy: (otwór 2, otwór 1), (otwór 1, otwór 3).

Zadanie 11 Obserwowano dobowy przybór odpadów (w m^3) na terenie kilku gmin w ciągu 5 dni. Otrzymano dane:

Gmina 1	Gmina 2	Gmina 3	Gmina 4
6,5	8,0	8,0	5,3
6,7	8,4	9,4	6,4
7,2	9,2	7,8	7,2
6,4	8,3	8,3	6,7
6,9	7,9	8,5	6,9

a) Na poziomie istotności $\alpha = 0{,}01$ zweryfikować hipotezę ogólną o braku istotnych różnic między rzeczywistymi średnimi przyborami odpadów w badanych gminach.
b) Wykorzystać test Tukeya do porównań średnich dobowych przyborów odpadów. W tym celu postawić i zweryfikować odpowiednie hipotezy szczegółowe. Przyjąć poziom istotności α= 0,01.
Odp. **a)** $p = 0{,}000042$.

Zadanie 12 Poniższe dane dotyczą frekwencji (w procentach) w wyborach prezydenckich z wybranych komisji wyborczych w trzech miastach wojewódzkich:

Wrocław: 38,5; 40,8; 41,7; 41,2; 37.9; 38,3; 42; 39,8; 43,1; 42,6
Warszawa: 38,9; 43,1; 40,4; 41,8; 42,39, 43,7; 40; 39,7; 43; 43,1
Poznań: 43,9; 44,2; 45,2; 44,6; 42,5; 43,4; 44,8; 42,8; 43,1; 44,8; 45

Na poziomie istotności α= 0,05 zweryfikować hipotezę, że rozkłady frekwencji we wszystkich miastach są jednakowe przy alternatywie, że się różnią. Zastosować test **Kruskala-Wallisa.**

Charakterystyka próby. Doświadczenia jednoczynnikowe – analiza wariancji i testy szczegółowe
Weryfikacja hipotez dotyczących jednorodności wielu wariancji rozkładów normalnych
Test nieparametryczny Kruskala- Wallisa

Wzory 1 : TESTY PARAMETRYCZNE

Analiza wariancji - UKŁAD CAŁKOWICIE LOSOWY		
Sumy kwadratów SS i średnie kwadraty MS	$SS_C = \sum_{i=1}^{v}\sum_{j=1}^{r_i} x_{ij}^2 - \frac{(\sum_{i=1}^{v}\sum_{j=1}^{r_i} x_{ij})^2}{n}$, $SS_O = \sum_{i=1}^{v}\frac{(\sum_{j=1}^{r_i} x_{ij})^2}{r_i} - \frac{(\sum_{i=1}^{v}\sum_{j=1}^{r_i} x_{ij})^2}{n}$, $SS_E = SS_C - SS_O$	$MS_O = \frac{SS_O}{v-1}$ $MS_E = \frac{SS_E}{n-v}$
Hipoteza ogólna - test dla porównania wielu obiektów - test Fishera – Snedecora		
$H_0: \mu_1 = \mu_2 = \ldots = \mu_v$ $H_1: \sim H_0$	$F_0 = \frac{MS_O}{MS_E}$	Obszar krytyczny $F_0 > F_{\alpha;v-1;n-v}$
Hipotezy szczegółowe - test dla różnicy średnich :		
test Tukeya		
$H_{0ii'}: \mu_i - \mu_{i'} = 0$ $H_{1ii'}: \sim H_{0ii'}$,	gdy $\lvert\bar{x}_i - \bar{x}_{i'}\rvert > NIR^T$ H_0 odrzucamy	Najmniejsza istotna różnica $NIR^T = \sqrt{\frac{MS_E}{r}} \cdot q_{\alpha,v,v_E}$
test Fishera		
$H_{0ii'}: \mu_i - \mu_{i'} = 0$ $H_{1ii'}: \sim H_{0ii'}$,	gdy $\lvert\bar{x}_i - \bar{x}_{i'}\rvert > NIR^F$ H_0 odrzucamy	Najmniejsza istotna różnica $NIR^F = \sqrt{MS_E(\frac{1}{r_i} + \frac{1}{r_{i'}})} \cdot t_{\alpha,v_E}$

Wzory 1 : TESTY NIEPARAMETRYCZNE

Hipotezy szczegółowe - test dla różnicy średnich :

Obliczamy średnie rang $\bar{x}_i$ dla każdego obiektu (populacji), i=1,2, …, v.

test Chi - kwadrat		
$H_{0ii'}: \mu_i - \mu_{i'} = 0$ $H_{1ii'}: \sim H_0$	gdy $\lvert\bar{x}_i - \bar{x}_{i'}\rvert > NIR^{chi}$ H_0 odrzucamy	Najmniejsza istotna różnica $NIR^{chi} = \sqrt{X^2_{\alpha,v-1} \cdot \frac{n(n+1)}{12} \cdot (\frac{1}{n_i} + \frac{1}{n_{i'}})}$