

Wydział Informatyki Przedmiot:	Data: 22.04.2024
<p>Grupa PS2</p> <p>Grupa 12</p> <p>Imię i nazwisko:</p> <p>Michał Sienicki</p> <p>Mateusz Milewski</p> <p>Damian Patalan</p>	<p>Prowadzący:</p> <p>Mgr. Joanna Gościk</p>

1. Analiza pierwszego zbioru danych:

1.1 Zmienne zależne: Przeanalizowano dwie zmienne zależne, w celu ustalenia różnicy między nimi.

1.2 Utworzenie zmiennej 'RÓŻNICA': Obliczono różnicę między zmiennymi pierwszą i drugą, której wartości zostały zapisane jako nowa zmienna 'RÓŻNICA'.

1.3 Statystyki opisowe:

N ważnych, średnia, mediana, odchylenie standardowe, przedział ufności dla odchylenia standardowego (95%), błąd standardowy średniej, przedział ufności dla średniej (95%), oraz ekstrema (minimum i maksimum) zostały obliczone dla trzech zmiennych, w tym zmiennej 'RÓŻNICA'.

Zmienna	Statystyki opisowe (Procent ATP sta w Statystyka 3.stw)										
	Nważnych	Średnia	Ufność -95,000%	Ufność 95,000%	Mediana	Minimum	Maksimum	Odch.std	P. ufności odch. std. -95,000%	P. ufności odch. std. +95,000%	Standard. Błąd
SZWAJCARIA	38	59,23684	57,47682	60,99687	59,50000	50,00000	69,00000	5,354635	4,365440	6,927545	0,868637
POLSKA	38	35,00000	33,99777	36,00223	35,50000	30,00000	40,00000	3,049147	2,485859	3,944826	0,494637
RÓŻNICA	38	24,23684	22,09924	26,37444	24,50000	11,00000	35,00000	6,503364	5,301957	8,413709	1,054985

1.4 Normalność rozkładu:

Przeprowadzono test normalności rozkładu Shapiro-Wilka dla zmiennej 'RÓŻNICA', a wyniki, w tym histogramy, zostały umieszczone w sprawozdaniu.

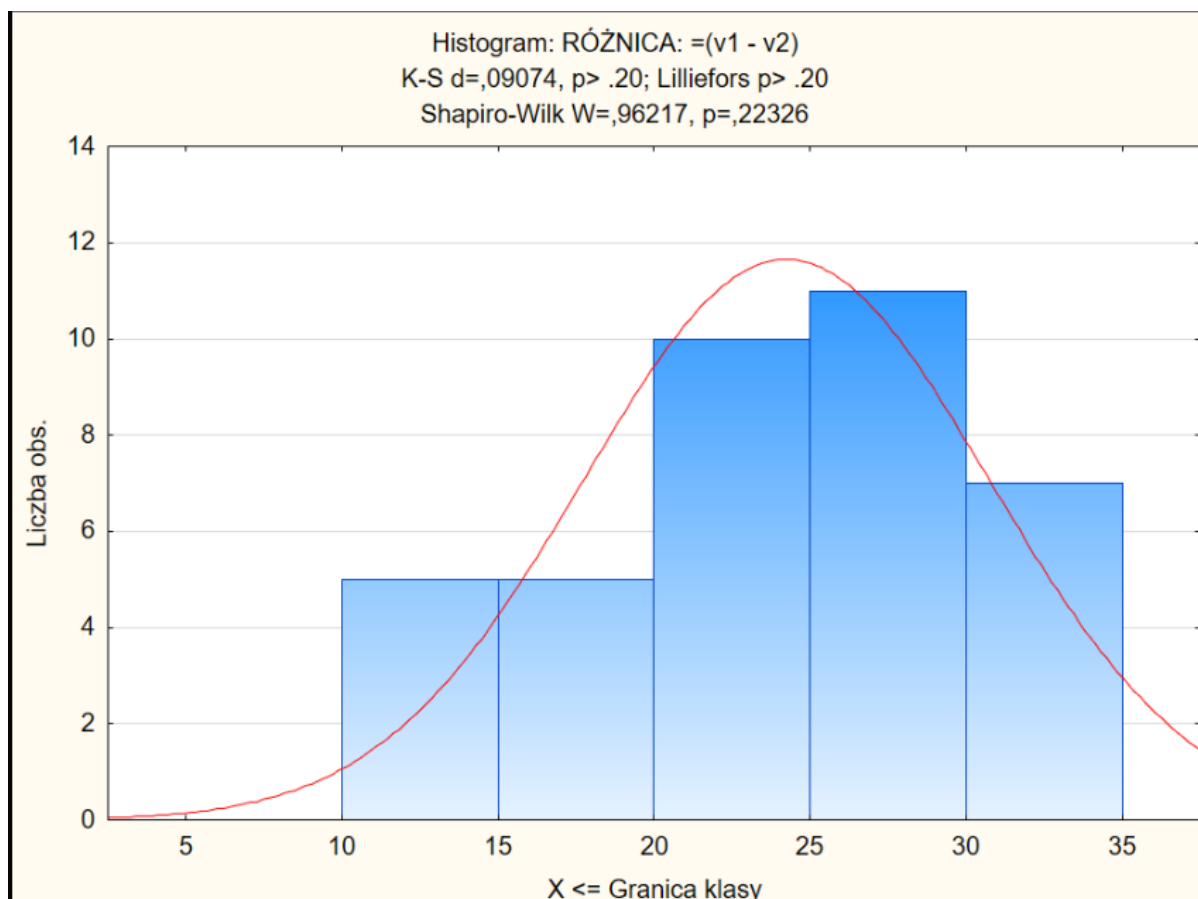


Tabela licznosci: RÓŻNICA: =(v1 - v2) (Procent_ATP.sta w Statystyka 3.stw)
K-S d=,09074, p> .20; Lilliefors p> .20
Shapiro-Wilk W=,96217, p=,22326

Klasa	Liczba	Skumulow. Liczba	Procent Ważnych	Skumul. % Ważnych	% ogółu Przypadki	Skumulow. % Ogółu
5,000000<x<=10,00000	0	0	0,00000	0,0000	0,00000	0,0000
10,00000<x<=15,00000	5	5	13,15789	13,1579	13,15789	13,1579
15,00000<x<=20,00000	5	10	13,15789	26,3158	13,15789	26,3158
20,00000<x<=25,00000	10	20	26,31579	52,6316	26,31579	52,6316
25,00000<x<=30,00000	11	31	28,94737	81,5789	28,94737	81,5789
30,00000<x<=35,00000	7	38	18,42105	100,0000	18,42105	100,0000
Braki	0	38	0,00000		0,00000	100,0000

1.5 Testy statystyczne:

W zależności od wyniku testu normalności rozkładu, wybrano odpowiedni test statystyczny:

Test t dla prób zależnych, jeśli dane miały rozkład normalny.

Test średnich względem stałej wartości odniesienia (Procent_ATP.sta w Statystyka 3.stw)

Zmienna	Średnia	Odch.st.	Ważnych	Bł. std.	Ufność -95,000%	Ufność +95,000%	Odniesienie Stała	t	df	p
RÓŻNICA	24,23684	6,503364	38	1,054985	22,09924	26,37444	0,00	22,97364	37	0,000000

Test kolejności par Wilcoxona, jeśli dane nie miały rozkładu normalnego.

Test średnich względem stałej wartości odniesienia (Procent_ATP.sta w Statystyka 3.stw)

Zmienna	Średnia	Odch.st.	Ważnych	Bł. std.	Ufność -95,000%	Ufność +95,000%	Odniesienie Stała	t	df	p
RÓŻNICA	24,23684	6,503364	38	1,054985	22,09924	26,37444	0,00	22,97364	37	0,000000

Test kolejności par Wilcoxona (Procent_ATP.s Zaznaczone wyniki są istotne z $p < ,05000$				
Para zmiennych	N	T	Z	p
Ważnych				
SZWAJCARIA & RÓŻNICA	38	0,00	5,373093	0,000000

Test kolejności par Wilcoxona (Procent_ATP.sta w Zaznaczone wyniki są istotne z $p < ,05000$				
Para zmiennych	N	T	Z	p
Ważnych				
POLSKA & RÓŻNICA	38	19,50000	5,090299	0,000000

1.6 Zapis skoroszytu: Wszystkie wyniki i działania zostały zapisane.

Statystyki opisowe:

Dla zmiennych SZWAJCARIA, POLSKA oraz RÓŻNICA obliczono średnią, medianę, odchylenie standardowe oraz przedział ufności. Wyniki wykazały, że średnie wartości dla SZWAJCARIA i POLSKA różnią się znacząco, co wskazuje na potencjalną różnicę w analizowanych danych.

Normalność rozkładu zmiennej RÓŻNICA:

Test Shapiro-Wilka nie doprowadził do odrzucenia hipotezy o normalności rozkładu ($p=0.22326$), sugerując, że dane mogą pochodzić z populacji o rozkładzie normalnym.

Test t dla zmiennej RÓŻNICA:

Przeprowadzony test t wykazał, że istnieje statystycznie istotna różnica względem wartości odniesienia ($p < 0.00001$). Wskazuje to na istotną różnicę między badanymi grupami.

Test kolejności par Wilcoxona:

Ze względu na wynik testu Shapiro-Wilka dla zmiennej RÓŻNICA, przeprowadzono również test Wilcoxona, który potwierdził wcześniejsze wyniki testu t, wykazując istotność statystyczną różnicy ($p < 0.00001$).

Wnioski dla pierwszego zbioru danych:

Średnia wartość zmiennej RÓŻNICA jest znacząco różna od zera, co sugeruje, że analizowane zmienne zależne różnią się między sobą w sposób statystycznie istotny.

Histogramy zmiennej RÓŻNICA wskazują na rozkład zbliżony do normalnego, co wspiera zastosowanie testu t.

2. Analiza drugiego zbioru danych:

2.1 Zmienne niezależne: Zanalizowano dwie zmienne niezależne.

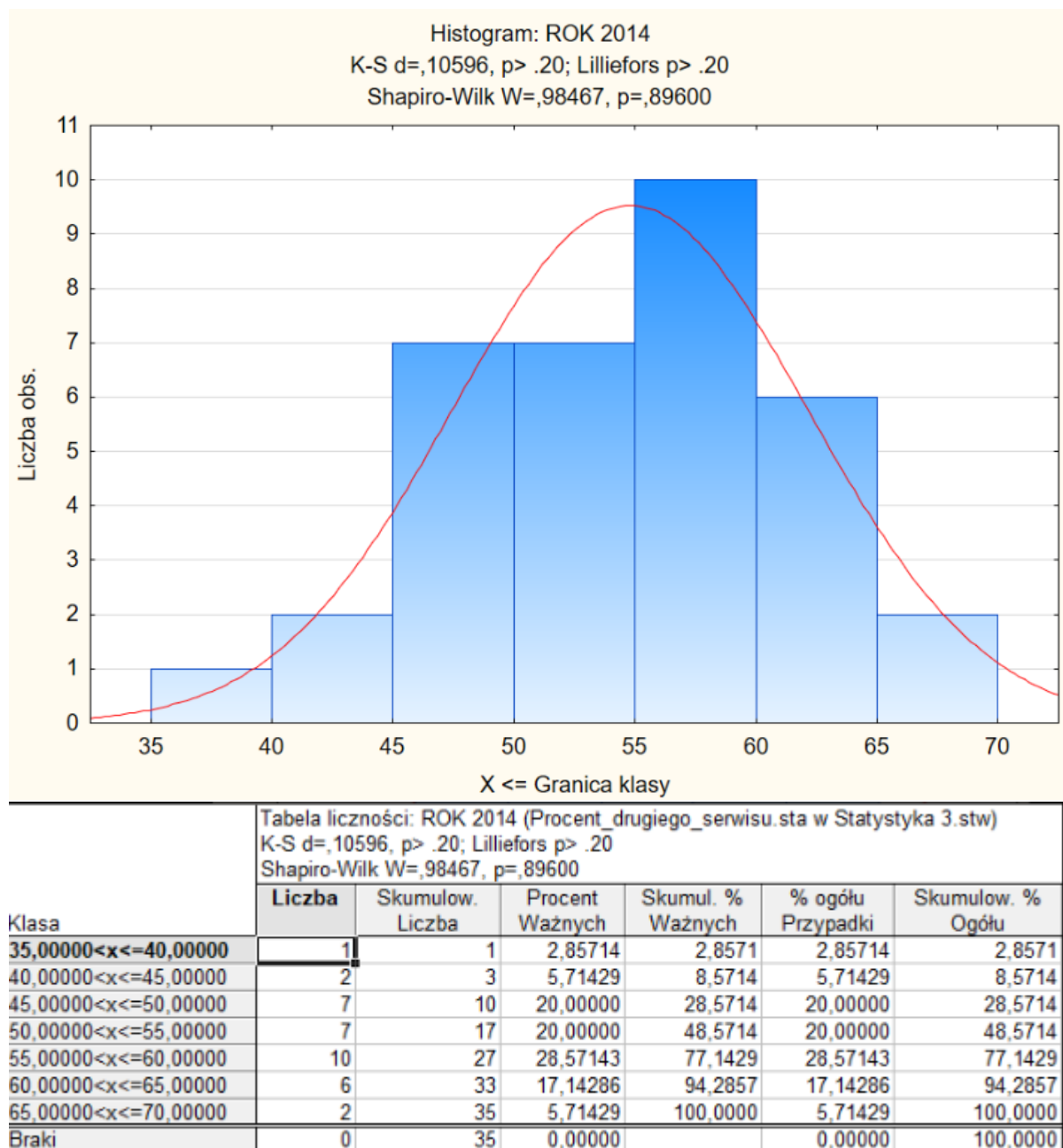
2.2 Statystyki opisowe:

Podobnie jak w pierwszym zbiorze danych, obliczono statystyki opisowe dla obu zmiennych niezależnych.

Statystyki opisowe (Procent drugiego serwisu sta w Statystyka 3.stw)											
Zmienna	Nważnych	Srednia	Ufność -95,000%	Ufność 95,000%	Mediana	Minimum	Maksimum	Odch.std	P. ufności odch. std. -95,000%	P. ufności odch. std. +95,000%	Standard. Błąd
ROK 2013	35	58,68571	56,20613	61,16530	58,00000	44,00000	74,00000	7,218324	5,838701	9,457463	1,220119
ROK 2014	35	54,80000	52,28287	57,31713	56,00000	40,00000	70,00000	7,327627	5,927113	9,600672	1,238595

2.3 Normalność rozkładów:

Dla każdej zmiennej niezależnej przeprowadzono test Shapiro-Wilka, a wyniki przedstawiono za pomocą histogramów oraz tabel.



2.6 Zapis skróty: Wszystkie wyniki i działania zostały zapisane.

Analiza drugiego zbioru danych:

Statystyki opisowe dla lat 2013 i 2014:

Obliczono średnie, mediany, odchylenia standardowe i przedziały ufności dla danych z obu lat. Wyniki wykazały niewielkie różnice między latami.

Test t dla prób niezależnych (ROK 2013 vs. ROK 2014):

Przeprowadzono test t, który nie wykazał statystycznie istotnej różnicy w średnich między latami ($p=0.930680$).

Test Levene'a na równość wariancji:

Test Levene'a wykazał, że wariancje między grupami są jednorodne ($p=0.891406$), co jest założeniem do przeprowadzenia testu t dla prób niezależnych.

Wnioski dla drugiego zbioru danych:

Brak istotnych różnic w średnich między latami 2013 i 2014 może wskazywać na stabilność badanego zjawiska w czasie.

Jednorodność wariancji między grupami pozwala na stosowanie testów parametrycznych, takich jak test t.

Podsumowanie:

Analiza statystyczna dwóch zbiorów danych przeprowadzona za pomocą testu t, testu Shapiro-Wilka oraz testu Wilcoxa dostarczyła informacji o istotnych różnicach w obrębie pierwszego zbioru danych, co sugeruje, że analizowane zmienne zależne są od siebie statystycznie różne. W przypadku drugiego zbioru danych, brak wykazanych istotnych różnic w średnich między latami sugeruje stabilność analizowanego zjawiska w czasie.

Dodatkowe uwagi:

Dołączono histogramy i tabele licznosci, które wizualizują rozkład danych i uzupełniają interpretację statystyczną.

Politechnika Białostocka Wydział Informatyki,

22.04.24r