

**AUTOR PRACY:** Adam Bajguz; e-mail: adam.bajguz@gmail.com

**SZKOŁA:** III Liceum Ogólnokształcące im. K. K. Baczyńskiego w Białymstoku, ul. Pałacowa 2/1, 15-042 Białystok, telefon: +48 85 741 60 33

**OPIEKUN NAUKOWY:** Małgorzata Schroeder; e-mail: szumi01@o2.pl

**CEL PRACY:** charakterystyka warunków naturalnych w Polsce w oparciu o dane meteorologiczne na podstawie rozdziału 1. *Małego Rocznika Statystycznego Polski 2015*

#### OBJAŚNIENIA ZASTOSOWANYCH SKRÓTÓW:

<b>K-S-L</b>	Test normalności Kołmogorowa - Smirnowa z poprawką Lillieforsa	<b>*T1</b>	Średnie roczne temperatury (°C) powietrza w latach 1991-2000
<b>SW-W</b>	Test W Shapiro-Wilka	<b>*T2</b>	Średnie roczne temperatury (°C) powietrza w latach 2001-2010
<b>Q<sub>1</sub></b>	dolny kwartył (kwantyl rzędu 1/4)	<b>T3</b>	Średnie roczne temperatury (°C) powietrza w roku 2014
<b>Q<sub>2</sub></b>	mediana (kwantyl rzędu 1/2)	<b>*O1</b>	Średnie roczne sumy opadów (mm) w latach 1991-2000
<b>Q<sub>3</sub></b>	górny kwartył (kwantyl rzędu 3/4)	<b>*O2</b>	Średnie roczne sumy opadów (mm) w latach 2001-2010
<b>α</b>	poziom istotności	<b>O3</b>	Średnie roczne sumy opadów (mm) w roku 2014
<b>p</b>	p-value	<b>U</b>	Usłonecznienie (godz.) w roku 2014
		<b>W</b>	Średnia prędkość wiatru (m/s) w roku 2014

\* – zmienna dodatkowa

**MATERIAŁ I METODY:** Dane meteorologiczne zawarte na stronie 25 *Małego Rocznika Statystycznego 2015* (TABL. 7.) posłużyły do przeprowadzenia analizy warunków naturalnych za pomocą:

- podstawowych miar statystycznych i dwóch testów: K-S-L oraz SW-W;
- statystyki wielowymiarowej: analizy składowych głównych (w celu pogłębienia analizy dodano podział na krainy geograficzne: pobrzeże, pojezierze, niziny, wyżyny, Sudety, kotliny podkarpackie, Karpaty).

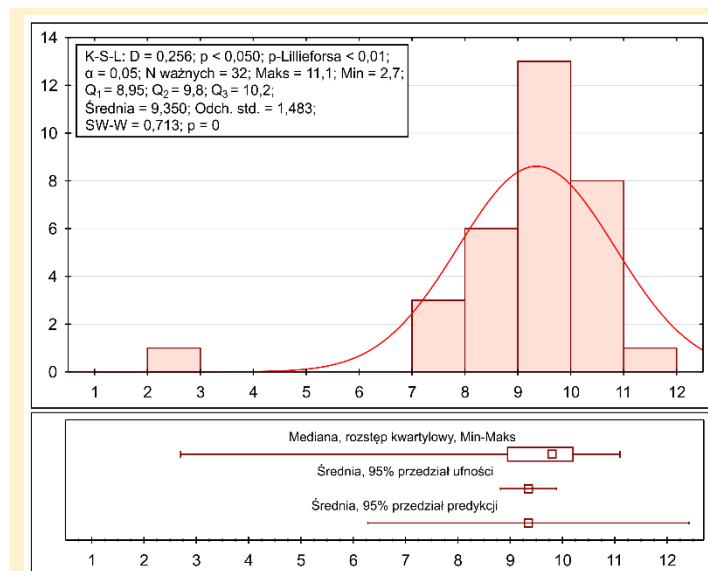
Do wykonania analiz statystycznych oraz wykresów wykorzystano pakiet StatSoft Statistica 12.5.

#### WYNIKI:

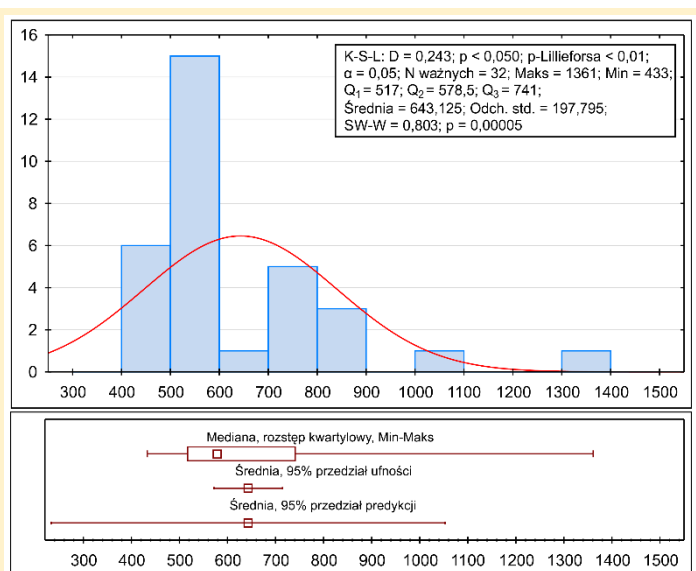
##### A) STATYSTYKA OPISOWA I MATEMATYCZNA

W celu wyciągnięcia podstawowych wniosków dotyczących warunków naturalnych panujących w Polsce wykorzystano dane dotyczące średnich rocznych temperatur (T3), średnich rocznych sum opadów (O3), usłonecznienia (U) i średniej prędkości wiatru (W) w 2014 roku. Z danych obliczono następujące miary statystyczne: liczbę N ważnych (liczbę obserwacji, które posłużyły do wyliczenia wartości statystycznych), minimum, maksimum, dolny i górny kwartył, medianę, średnią arytmetyczną oraz odchylenie standardowe. Przeprowadzono również dwa testy statystyczne (**K-S-L** oraz **SW-W**) weryfikujące czy rozkład obserwacji jest zgodny z rozkładem normalnym.

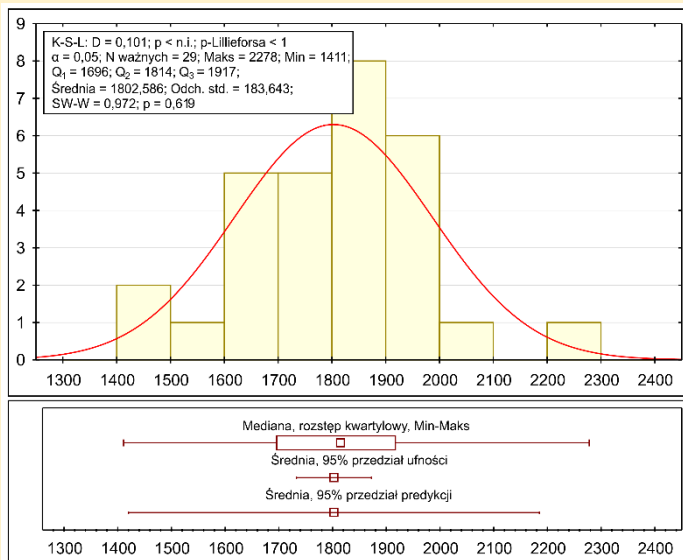
Dane meteorologiczne, wartości wyżej wymienionych miar statystycznych oraz wyniki testów zgodności rozkładów zostały przedstawione na poniższych histogramach (**Ryc. 1-4**):



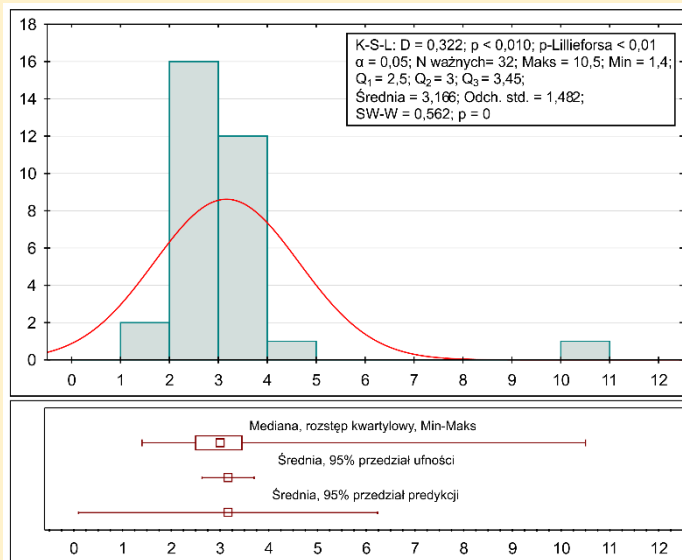
**Ryc. 1.** Histogram i wykres ramka-wąsy zmiennej T3



**Ryc. 2.** Histogram i wykres ramka-wąsy zmiennej O3



Ryc. 3. Histogram i wykres ramka-wąsy zmiennej U



Ryc. 4. Histogram i wykres ramka-wąsy zmiennej W

Stwierdzono, że jedynie wartości usłonecznienia w 2014 roku (zmienna U) mają w przybliżeniu rozkład normalny – z prawdopodobieństwem większym niż 95% (wartość współczynnika  $p$  dla testu SW-W jest większa niż  $\alpha=0,05$ , natomiast wartość  $p$  dla testu K-S-L jest nieistotna (n.i.), czyli jest większa niż  $\alpha=0,05$ ).

Na podstawie analizy wartości minimum, maksimum oraz wartości dolnego i górnego kwartyła (z definicji: *około 25% rejestrowanych przypadków jest poniżej wartości dolnego kwartyła, a około 25% – powyżej wartości górnego kwartyła*) (Ryc. 1-4.) wraz z TABL. 7. Małego Rocznika Statystycznego stwierdzono, że w 2014 roku w stacjach meteorologicznych:

- w **Gorzowie Wielkopolskim, Kaliszu, Opolu, Poznaniu, Szczecinie, Wrocławiu** (absolutne maksimum) oraz w **Zielonej Górze** zanotowano najwyższe wartości średnich temperatur (T3), tymczasem najniższe wartości średnich temperatur stwierdzono w **Białymstoku, Mławie, Suwałkach**, na **Śnieżce** (absolutne minimum), w **Terespolu, Włodawie** oraz w **Zakopanem**;
- w **Częstochowie, Jeleniej Górze, Kielcach, Lublinie, Nowym Sączu**, na **Śnieżce** oraz w **Zakopanem** (absolutne maksimum) odnotowano najwyższe wartości średnich opadów (O3), natomiast najniższe wartości średnich opadów w **Białymstoku, Chojnicach, Helu, Łebie** (absolutne minimum), **Mławie, Terespolu** oraz **Toruniu**;
- w **Chojnicach, Łebie, Opolu, Poznaniu, Terespolu, Toruniu** oraz w **Warszawie** (absolutne maksimum) rejestrowano najwyższe wartości usłonecznienia (U), podczas gdy najniższe wartości usłonecznienia w **Gorzowie Wielkopolskim, Łodzi, Nowym Sączu, Suwałkach**, na **Śnieżce** (absolutne minimum), w **Zakopanem** oraz w **Zielonej Górze**;
- w **Chojnicach, Kaliszu, Łebie, Łodzi, Rzeszowie, Szczecinie**, na **Śnieżce** (absolutne maksimum) oraz we **Włodawie**, odnotowano najwyższe wartości średniej prędkości wiatru (W), z kolei najniższe wartości średniej prędkości wiatru w **Białymstoku, Częstochowie, Katowicach, Nowym Sączu, Opolu, Terespolu, Toruniu** oraz w **Zakopanem** (absolutne minimum).

## B) ANALIZA SKŁADOWYCH GŁÓWNYCH

Zjawiska fizyczne – meteorologiczne cechuje ogromna złożoność. W celu redukcji liczby zmiennych, wykrycia struktury i ogólnych prawidłowości w związkach między zmiennymi oraz zweryfikowania wykrytych prawidłowości i powiązań zastosowano analizę składowych głównych. Jest ona zespołem procedur statystycznych, pozwalających na klasyfikację badanych obiektów w nowych ortogonalnych tzn. wzajemnie nieskorelowanych przestrzeniach, zdefiniowanych przez nowe zmienne – czynniki. Buduje ona teoretyczny model, który opisuje korelacje między badanymi zmiennymi, tzn. średnią roczną temperaturą powietrza, średnią roczną sumą opadów, usłonecznieniem i średnią prędkością wiatru w 2014 roku.

Zmienne pierwotne zawarte na stronie 25 *Małego Rocznika Statystycznego 2015* (TABL. 7.) mają różne jednostki, w związku z tym do analizy wykorzystana została macierz korelacji, która przyjmuje postać przedstawioną obok:

$$R = \begin{pmatrix} 1 & -0,390 & 0,465 & 0,715 \\ -0,390 & 1 & -0,555 & -0,027 \\ 0,465 & -0,555 & 1 & -0,214 \\ -0,715 & -0,027 & -0,214 & 1 \end{pmatrix} \begin{matrix} T3 \\ O3 \\ U \\ W \end{matrix}$$

Istotne korelacje zostały zaznaczone czcionką pogrubioną. Stwierdzono niską korelację między średnią roczną sumą opadów (O3) a średnią prędkością wiatru (W) w roku 2014. Pozostałe korelacje wykazują wysoką zależność. Można przypuszczać, że w macierzy występuje wyraźna struktura, która będzie miała wpływ na postać składowych głównych.

W celu przetestowania słuszności zastosowania analizy składowych głównych wykorzystano:

- **współczynnik Kaisera-Mayera-Olkina (KMO)** – oceniający adekwatność macierzy korelacji, poprzez porównanie korelacji cząstkowej z dwuzmiennowymi współczynnikami korelacji. Przyjmuje on wartość z przedziału  $<0,1>$ . Im mniejsza wartość KMO tym redukcja wymiarowości (ilości zmiennych) będzie mniejsza. Przyjmuje się, że wartość współczynnika KMO powinna wynosić co najmniej 0,5;
- **test sferyczności Bartletta** – zakłada, że macierz jest macierzą jednostkową, czyli taką w której wszystkie współczynniki korelacji (poza współczynnikami identycznych zmiennych) są równe zero, tzn. są nieistotne statystycznie. Zatem odrzucenie testowanej hipotezy zerowej jest podstawą do zastosowania analizy składowych głównych.

Dla analizowanych danych meteorologicznych określono:

współczynnik KMO:	0,508
test sferyczności Bartletta:	statystyka chi-kwadrat: 41,669
	liczba stopni swobody (df): 6
	p: 0,000

Stwierdzono, że współczynnik KMO jest większy od 0,5. Natomiast poziom istotności testu Bartletta (p) jest mniejszy od  $\alpha$  ( $\alpha=0,05$ ) – oznacza to, że należy odrzucić hipotezę zerową. Istnieją zatem wystarczające podstawy do przeprowadzenia analizy składowych głównych.

Na podstawie danych zawartych w **Tab. 1.** stwierdzono, że pierwsza wartość własna ( $\lambda_1$ ) wyjaśnia około 54,8% całkowitej wariancji (zmienności) danych meteorologicznych, druga ( $\lambda_2$ ) – około 84,8%, trzecia ( $\lambda_3$ ) – około 95,6%, a czwarta ( $\lambda_4$ ) – 100%. I-ta wartość własna jest równa wariancji i-tej składowej głównej  $Z_i$ . Pozwoliło to zdefiniować część wariancji danych meteorologicznych wyodrębnionej przez i-tą składową (% ogółu wariancji). Wartości własne wskazują wpływ zmiennych pierwotnych na poszczególne składowe główne.

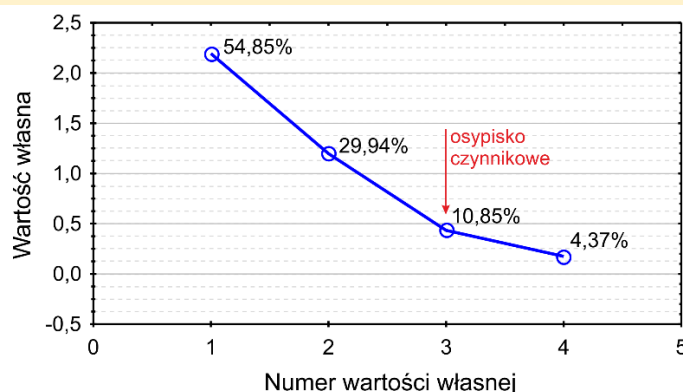
Tab. 1. Wartości własne na podstawie macierzy korelacji zmiennych aktywnych					Tab. 2. Wektory własne macierzy korelacji				
Nr wartości (i)	1	2	3	4	Zmienna	T3	O3	W	U
Wartość własna $\lambda_i$	2,194	1,197	0,434	0,175	Czynnik 1	0,602	-0,424	-0,443	0,511
% ogółu wariancji	54,845	29,937	10,847	4,371	Czynnik 2	0,268	0,607	-0,648	-0,373
Skumulowana wartość własna	2,194	3,391	3,825	4	Czynnik 3	0,271	-0,571	0,023	-0,774
Skumulowany % wariancji	54,845	84,782	95,629	100	Czynnik 4	0,701	0,353	0,619	0,003

**Tab. 2.** przedstawia wektory własne odpowiadające znalezionym wartościom własnym. Są to współczynniki składowych  $Z_1$ ,  $Z_2$ ,  $Z_3$  oraz  $Z_4$ , które odzwierciedlają wpływ zmiennych pierwotnych dotyczących danych meteorologicznych na daną składową główną:

Wartość własna:	Składowa główna:
$\lambda_1 = 2,194$	$Z_1 = 0,602T3 - 0,424O3 - 0,443W + 0,511U$
$\lambda_2 = 1,197$	$Z_2 = 0,268T3 + 0,607O3 - 0,648W - 0,373U$
$\lambda_3 = 0,434$	$Z_3 = 0,271T3 - 0,571O3 + 0,023W - 0,774U$
$\lambda_4 = 0,175$	$Z_4 = 0,701T3 + 0,353O3 - 0,619W + 0,003U$

Przy ustaleniu liczby składowych zostały zastosowane następujące kryteria wyboru:

- 1. skumulowany procent wyjaśnionej wariancji analizowanych zmiennych** – liczbę składowych wyznacza się do momentu aż suma ich wariancji przekroczy pewną wartość, np. 75% lub 80%. Według tego kryterium, dla danych z **Tab. 1.** należy pozostawić dwie pierwsze składowe;
- 2. kryterium Kaisera** – wykorzystuje się tylko te składowe, którym odpowiadają wartości własne większe od jednego, tzn. dla danych z **Tab. 1.** należy pozostawić pierwsze dwie składowe;
- 3. kryterium Cattella** – bazuje na tzw. wykresie osypiska (**Ryc. 6.**). Jest to wykres liniowy przedstawiający kolejne wartości własne. Według Cattella na wykresie należy odnaleźć miejsce, od którego na prawo występuje łagodny spadek załamania – tzw. osypisko czynnikowe (punkt w którym przyrost informacji doznaje załamania). Zatem nie należy uwzględniać więcej czynników niż te znajdujące się na lewo od tego punktu. Sam punkt jest przez niektórych statystyków uwzględniany, a przez innych pomijany.



**Ryc. 6.** Wykres osypiska Cattella

Na podstawie trzeciego kryterium należy ograniczyć liczbę składowych do dwóch lub trzech. Natomiast pierwsze dwa kryteria przemawiają za tym, aby ograniczyć liczbę składowych tylko do dwóch. Reasumując, najwłaściwszy wybór to dwie składowe, które wyjaśniają około 85% całkowitej zmienności, całkowicie spełniając warunki zawarte w dwóch pierwszych kryteriach, częściowo w trzecim.

**Tab. 3.** Współrzędne czynnikowe przypadków wraz z odpowiadającymi im krainami geograficznymi

Przypadek	Czynnik 1	Czynnik 2	Stacja meteorologiczna	Przypadek	Czynnik 1	Czynnik 2	Stacja meteorologiczna
1	0,336	-0,553	Hel A	16	0,495	-0,265	Kalisz C
2	0,695	-1,880	Łeba A	17	-0,113	-0,082	Łódź C
3	0,535	-0,180	Koszalin A	18	-0,760	0,325	Włodawa C
4	-0,854	-0,261	Suwałki B	19	-0,317	0,517	Lublin D
6	0,561	-0,943	Chojnice B	20	1,206	-0,037	Wrocław C
7	0,493	-0,079	Szczecin A	21	-0,174	0,837	Jelenia Góra E
8	0,044	-0,347	Białystok C	22	-0,273	0,723	Kielce D
9	1,231	-0,404	Toruń B	23	0,408	0,698	Częstochowa D
10	0,362	-0,612	Mława C	24	-6,315	-2,588	Śnieżka E
11	0,137	0,660	Gorzów Wielkopolski B	25	1,316	0,153	Opole C
12	1,017	-0,448	Poznań B	26	0,005	0,042	Kłodzko E
13	1,688	-1,138	Warszawa C	27	0,494	0,568	Katowice D
14	0,712	-0,597	Terespol C	28	0,068	-0,051	Rzeszów F
15	0,246	0,432	Zielona Góra B	31	-0,323	1,991	Nowy Sącz G
				32	-2,919	3,519	Zakopane G

**Oznaczenia krain geograficznych:** A – pobrzeże, B – pojezierze, C – nizina, D – wyżyna, E – Sudety, F – kotlina podkarpacka, G – Karpaty

Przyjmując otrzymane dwie składowe  $Z_1$  i  $Z_2$  (odpowiednio czynnik 1 i 2) jako osie nowego układu, obliczono współrzędne wszystkich przypadków w tak utworzonym układzie (**Tab. 3.**). Liczba przypadków wynosi 29, ponieważ przypadki o numerze: 5, 29, 30 (**TABL. 7. Malego Rocznika Statystycznego** ze strony 25) nie zostały ujęte w analizie z powodu braku pomiaru usłonecznienia w 2014 roku przez odpowiednie stacje meteorologiczne.

Z kolei w celu dokładniejszej interpretacji składowych stworzona została **Tab. 4.** Zawiera ona ładunki czynnikowe, które należy interpretować jako korelacje między składowymi a zmiennymi. Odzwierciedlają one wpływ pierwotnych zmiennych meteorologicznych na składowe główne (analogicznie jak wektory własne).



**Tab. 4.** Ładunki (współrzędne) czynnikowe zmiennych aktywnych i \*dodatkowych

Zmienna	T3	O3	W	U	*T1	*T2	*O1	*O2
Czynnik 1	0,892	-0,628	-0,656	0,757	0,925	0,924	-0,829	-0,787
Czynnik 2	0,294	0,665	-0,709	-0,408	0,258	0,249	0,144	0,269

**Tab. 5.** Wkłady zmiennych

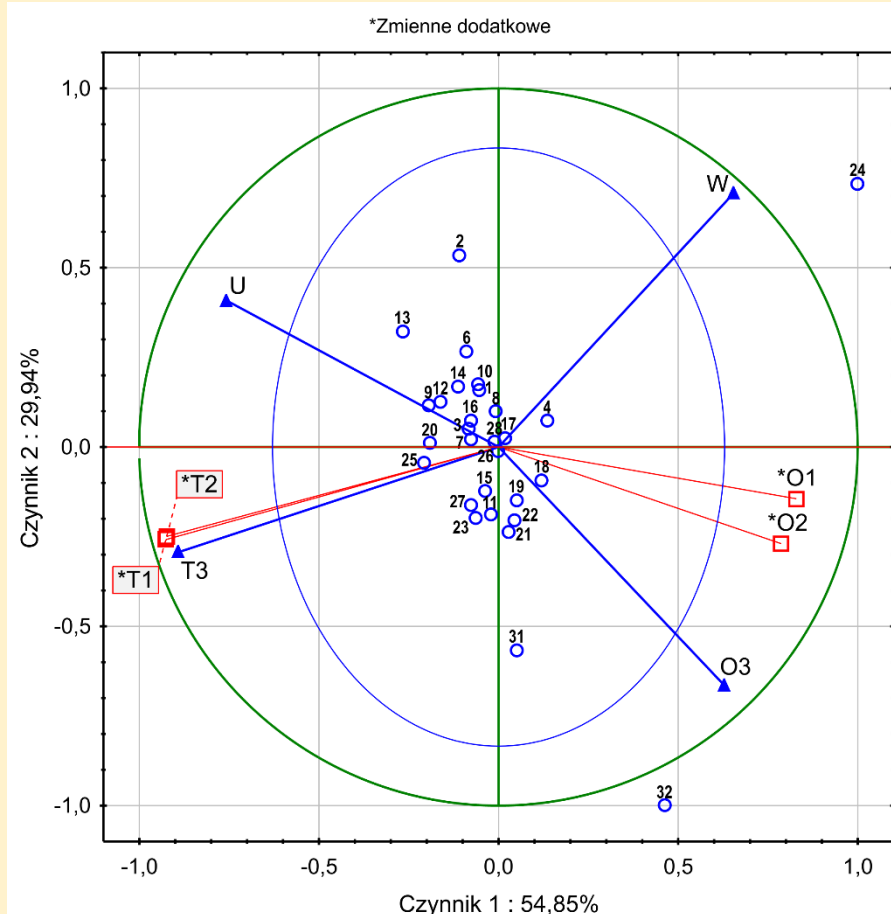
Zmienna	T3	O3	W	U
Czynnik 1	<b>0,363</b>	0,180	0,196	<b>0,261</b>
Czynnik 2	0,072	<b>0,369</b>	<b>0,420</b>	0,139

W **Tab. 5.** zawarto dane dotyczące wkładów zmiennych, czyli wartości za pomocą których można stwierdzić jaka część zmienności danej składowej głównej może być tłumaczona zmiennością poszczególnych zmiennych pierwotnych.

Pierwsza składowa reprezentuje głównie zmienne pierwotne dotyczące średnich temperatur powietrza (T3) oraz usłonecznienia (U) w 2014 roku (**Tab. 5.**). Pozostałe zmienne pierwotne są odzwierciedlone w mniejszym stopniu. Uwzględniając dane dotyczące wektorów własnych (**Tab. 2.**), można stwierdzić, że pierwsza składowa wiąże wysokie odnotowane temperatury z wysokimi czasami usłonecznienia, oraz odpowiednio – niskie temperatury z niskimi czasami usłonecznienia (**kolor niebieski** w **Tab. 2. i 5.**)

Druga składowa przedstawia głównie średnie sumy opadów (O3) oraz średnią prędkość wiatru (W) w 2014 roku. Pozostałe zmienne pierwotne są odzwierciedlone w niewielkim stopniu (**Tab. 5.**). Przeciwnie wartości wektorów własnych dla powyższych zmiennych (**Tab. 2.**) wskazują na ujemną korelację, co świadczy o powiązaniu wysokich poziomów opadów z niskimi prędkościami wiatru, oraz odpowiednio – niskich poziomów opadów z wysokimi prędkościami wiatru (**kolor zielony** w **Tab. 2. i 5.**).

Dane zawarte w **Tab. 3.** oraz **Tab. 4.** zostały przedstawione graficznie na biplotie (**Ryc. 7.**) – wykresie łączącym wykres współrzędnych czynnikowych przypadków oraz wykres współrzędnych czynnikowych zmiennych.

**Ryc. 7.** Biplot 1x2

Punkty 24 i 32 (**Ryc. 7.**) odstają od elipsy rozkładu normalnego (**kolor niebieski**). Na podstawie rzutów prostopadłych punktów 24 i 32 na wektory T3, O3, U i W stwierdzono:

- podwyższenie opadów deszczu (O3), prędkości wiatru (W) oraz obniżenie temperatury (T3) i usłonecznienia (U) w stacji meteorologicznej na Śnieżce (24) w stosunku do wartości średniej w 2014 roku;
- podwyższenie opadów deszczu (O3) oraz obniżenie usłonecznienia (U), temperatury (T3) i prędkości wiatru (W) w stacji meteorologicznej w Zakopanem (32) w stosunku do wartości średniej w 2014 roku.

Na podstawie **Ryc. 7.** stwierdzono, że w 2014 roku poziom prędkości wiatru (W), opadów deszczu (O3), temperatur (T3) i usłonecznienia (U) odnotowany w Łodzi (przypadek 17), Kłodzku (26) i Rzeszowie (28) (środek biplotu) jest najbardziej zbliżony do wartości średnich poszczególnych parametrów meteorologicznych w Polsce.

Wszystkie wektory przedstawione na **Ryc. 7.** są silnie zbliżone do tzw. koła jednostkowego (kolor zielony). Odzwierciedla to duże odchylenie standardowe danej zmiennej meteorologicznej i określa jej ogromne znaczenie z punktu widzenia tzw. mocy dyskryminacyjnej, tzn. zdolności zmiennej do różnicowania przypadków. Reasumując, w 2014 roku stwierdzono dużą zmienność wartości w obrębie analizowanych parametrów meteorologicznych, takich jak: temperatura, opady, usłonecznienie i prędkość wiatru.

Odległości w linii prostej pomiędzy przypadkami: 11, 15, 18, 19, 21, 22, 23, 25, 26, 27 wskazują na podobieństwo ich profili reakcji (**Ryc. 7.**). Należy zauważyć, że przypadki te reprezentują różne krainy geograficzne (**Tab. 3.**): pojezierze (11, 15), niziny (18, 25), wyżyny (19, 22, 23, 27) oraz Sudety (21, 26). Ponadto uwzględniając rzuty prostopadłe ww. punktów na wektory T3, O3, U i W, stwierdzono że charakteryzują się one podwyższonymi wartościami średnich temperatur (T3) i opadów (O3) oraz obniżonymi wartościami usłonecznienia (U) i średniej prędkości wiatru (W) w stosunku do wartości średnich w 2014 roku. Natomiast pomiędzy przypadkami: 1, 3, 4, 6, 7, 8, 9, 10, 12, 14, 16, 17, 28 występuje również podobieństwo ich profili reakcji. Przypadki te reprezentują: pojezierze (4, 6, 9, 12), pobraże (1, 3, 7), niziny (8, 10, 14, 16, 17) i kotlinę podkarpacką (28) oraz cechują się obniżonymi wartościami średnich temperatur (T3) i opadów (O3) oraz podwyższonymi wartościami usłonecznienia (U) i średniej prędkości wiatru (W) w stosunku do wartości średnich w 2014 roku.

Stwierdzono również, że kąt pomiędzy wektorami \*O1 i \*O2 oraz wektorami \*T1 i \*T2 jest mały (**Ryc. 7.**). Oznacza to wysoką dodatnią korelację pomiędzy średnimi temperaturami w latach 1991-2000 a temperaturami w latach 2001-2010 oraz pomiędzy średnimi opadami w latach 1991-2000 a opadami w latach 2001-2010. Średnie wartości temperatur i opadów odnotowanych w większości stacji meteorologicznych pomiędzy latami 1991-2000 a 2001-2010 uległy wzrostowi. Może to oznaczać, że warunki naturalne w Polsce uległy zmianie.

## WNIOSKI:

- ✓ Na podstawie statystyki opisowej i analizy głównych składowych stwierdzono, że w 2014 roku warunki naturalne w Polsce charakteryzowały się dużą zmiennością.
- ✓ Stwierdzono silne zróżnicowanie pod względem charakterystyki profili parametrów meteorologicznych, takich jak: temperatura, opady, usłonecznienie oraz prędkość wiatru w obrębie krain geograficznych – pobraża, pojezierza, nizin, wyżyn, Sudetów i Karpat.
- ✓ Stwierdzono, że warunki naturalne w Polsce między 1991-2000 a 2001-2010 uległy zmianie – odnotowano wzrost poziomu temperatur i opadów deszczu.

## BIBLIOGRAFIA:

### KSIĄŻKI:

- Mały Rocznik Statystyczny Polski 2015. Główny Urząd Statystyczny, Warszawa, Rok LVIII.  
<http://stat.gov.pl/obszary-tematyczne/roczniki-statystyczne/roczniki-statystyczne/maly-rocznik-statystyczny-polski-2015,1,16.html>
- Stanisław A., 2007. Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny, Tom III. StatSoft Polska, Kraków.

### ARTYKUŁ:

- Sagan A., 2014. Jeden obraz ukazuje więcej niż 10 liczb, czyli jak budować mapy zadowolenia klienta z wykorzystaniem programu STATISTICA. StatSoft Polska.  
<http://www.statsoft.pl/Portals/0/Downloads/04obraz.pdf>