

Analiza korelacyjna miesięcznych danych meteorologicznych (średniej temperatury i sum opadów) z stacji meteorologicznych w Bielsko-Białej i Katowicach.

Projekt ma na celu uzyskanie podstawowych informacji dotyczących temperatury powietrza i sum opadów w Bielsku-Białej oraz Katowicach na przestrzeni lat oraz ich analizę korelacyjną.

2. Materiały

Dla zrealizowania tematu posłużono się danymi wartości miesięcznych sum opadów oraz średnich miesięcznych temperatur powietrza ze stacji meteorologicznych w Bielsko-Białej i Katowicach. Obejmują one okres 70 lat, 1951 – 2021.

3. Metody

3.1 Metodami analizy wykorzystanymi w tym projekcie są **analiza rozkładu zmiennej i analiza korelacyjna**.

3.2 Wczytano dane z stacji meteorologicznych w Bielsko-Białej i w Katowicach:

```
TempBielsko <- read.delim("Temp_biel-biała.txt", sep = "\t", header = FALSE)
```

```
OpadyBielsko <- read.delim("Sum_opad_bial-biała.txt", sep = "\t", header = FALSE)
```

```
TempKatowice <- read.delim("Temp_katowice.txt", sep = "\t", header = FALSE)
```

```
OpadyKatowice <- read.delim("Sum_opad_katowice.txt", sep = "\t", header = FALSE)
```

Dodano nazwy kolumn:

```
Colnames(TempBielsko) <- c("Rok", "I", "II", "III", "IV", "V", "VI", "VII", "VIII", "IX", "X", "XI", "XII")
```

Stworzono wykres zbiorczy ramka-wąsy prezentujący rozkłady sum opadów w Bielsko-Białej dla wszystkich miesięcy:

```
boxplot(OpadyBielsko[,2:13], border="black",  
col="indianred1", main="Rozkład sum opadów",  
xlab="miesiące", ylab="sumy opadów [mm]")
```

Stworzono wykresy liniowe z dodanymi modelami liniowymi badające zmienność sum opadów w Bielsko-Białej dla poszczególnych miesięcy na przestrzeni lat np.:

```
ggplot(OpadyBielsko, aes(x = Rok, y = I)) +
```

```
geom_line(size = 2) +  
geom_smooth(method = "lm", color = "red", size = 2) +  
labs(title = "Styczen", x = "Rok", y = "Średnia suma opadów") +  
theme(text = element_text(size = 30))
```

Dla każdego miesiąca zbadano związek pomiędzy temperaturą a opadami używając współczynnika korelacji Pearsona oraz współczynnika korelacji rang Spearmana, a następnie stworzono tabelę z wynikami.

Stworzono wykresy rozrzutu z dodaną prostą regresji badające związek pomiędzy temperaturą a opadami w Bielsko-Białej na przestrzeni lat:

```
ggplot(temp_opady_Bielsko, aes(I, I2)) +  
geom_point(size = 5, colour = "darkblue") +  
geom_smooth(method = "lm", color = "black", size = 2) +  
labs(title = "Styczen", x = NULL, y = NULL) +  
theme(text = element_text(size = 30))
```

Zbadano związek pomiędzy średnią miesięczną temperaturą w Bielsko-Białej, a w Katowicach (dla 6 wybranych miesięcy) używając współczynnika korelacji Pearsona oraz współczynnika korelacji rang Spearmana, a następnie stworzono tabelę z wynikami.

Zbadano związek pomiędzy miesięczną sumą opadów w Bielsko-Białej, a w Katowicach (dla 6 wybranych miesięcy) używając współczynnika korelacji Pearsona oraz korelacji rang Spearmana i stworzono tabelę z wynikami.

Stworzono wykresy rozrzutu z dodaną prostą regresji badające związek pomiędzy temperaturą w Bielsko-Białej, a w Katowicach dla 6 wybranych miesięcy (to samo zrobiono dla związku pomiędzy sumą miesięcznych opadów w tych miastach):

```
ggplot(temp_Bielsko_Katowice, aes(II, II2)) +  
geom_point(size = 5, colour = "darkblue") +  
geom_smooth(method = "lm", color = "black", size = 2) +  
labs(title = "Luty", x = NULL, y = NULL) +  
theme(text = element_text(size = 30))
```

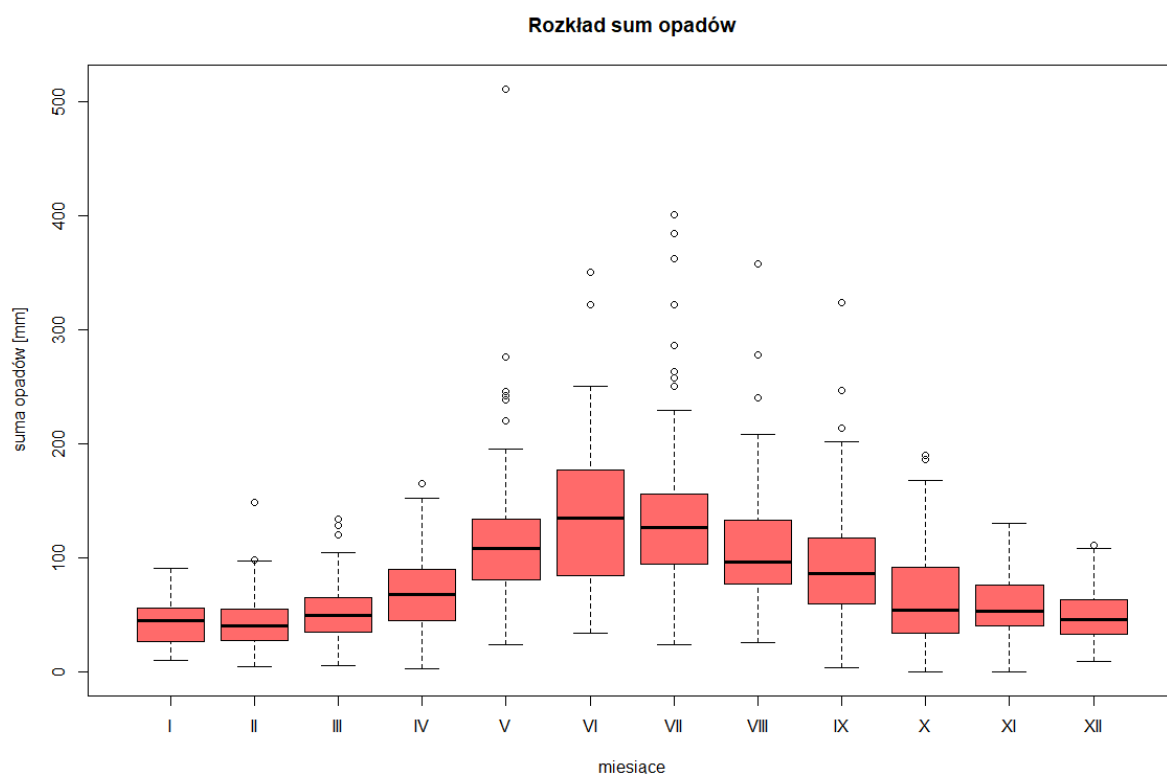
4. Wyniki

Analizując zmienność opadów w kolejnych miesiącach (rys.1) można zauważyć, że najczęściej opadów występuje w miesiącach letnich.

Sumy opadów rosną od lutego do lipca i maleją od lipca do stycznia.

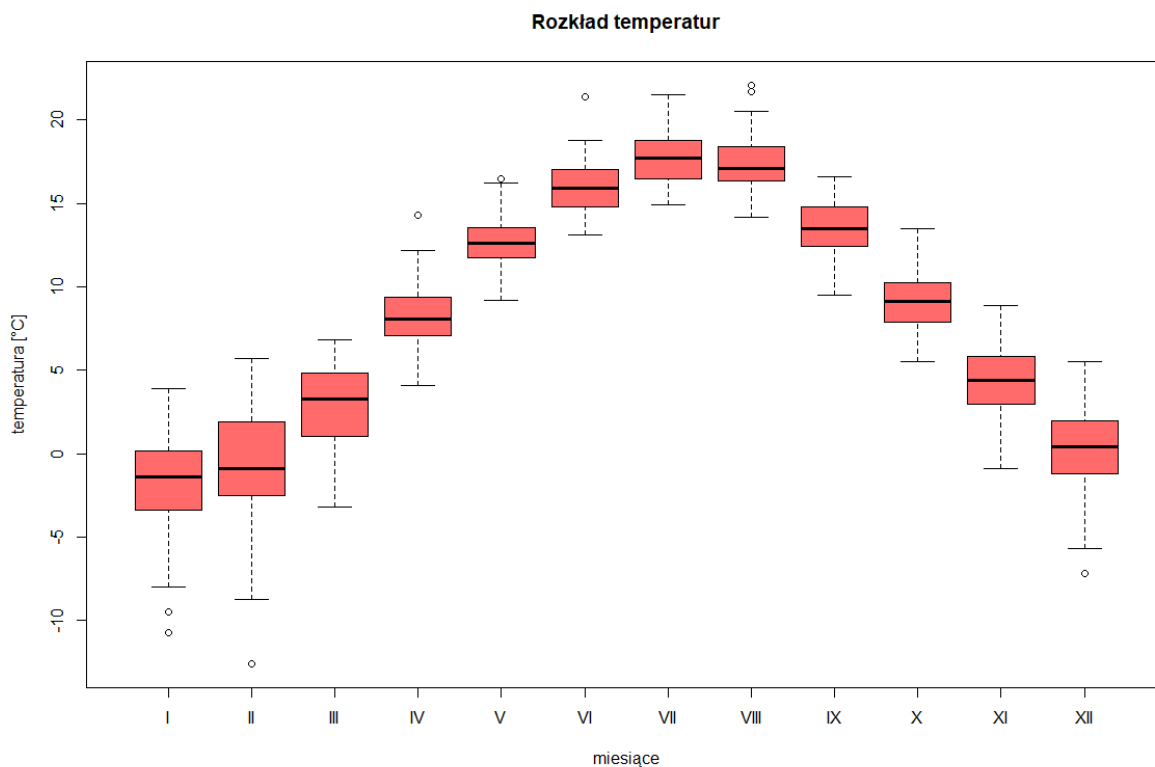
Najwyższą sumę opadów na przestrzeni ostatnich 70 lat zanotowano w maju i wynosiła ona nieco ponad 500mm.

Największy rozstęp międzykwartyłowy sum opadów notowany jest w czerwcu, a najmniejszy w lutym.



Rysunek 1 Wykres zbiorczy ramka-wąsy prezentujący rozkłady zmienności sum opadów dla wszystkich miesięcy.

Porównując wyniki z analogicznym wykresem wykonanym dla średnich temperatur (rys.2) okazuje się, że największe sumy opadów występują w miesiącach z najwyższą średnią temperaturą.

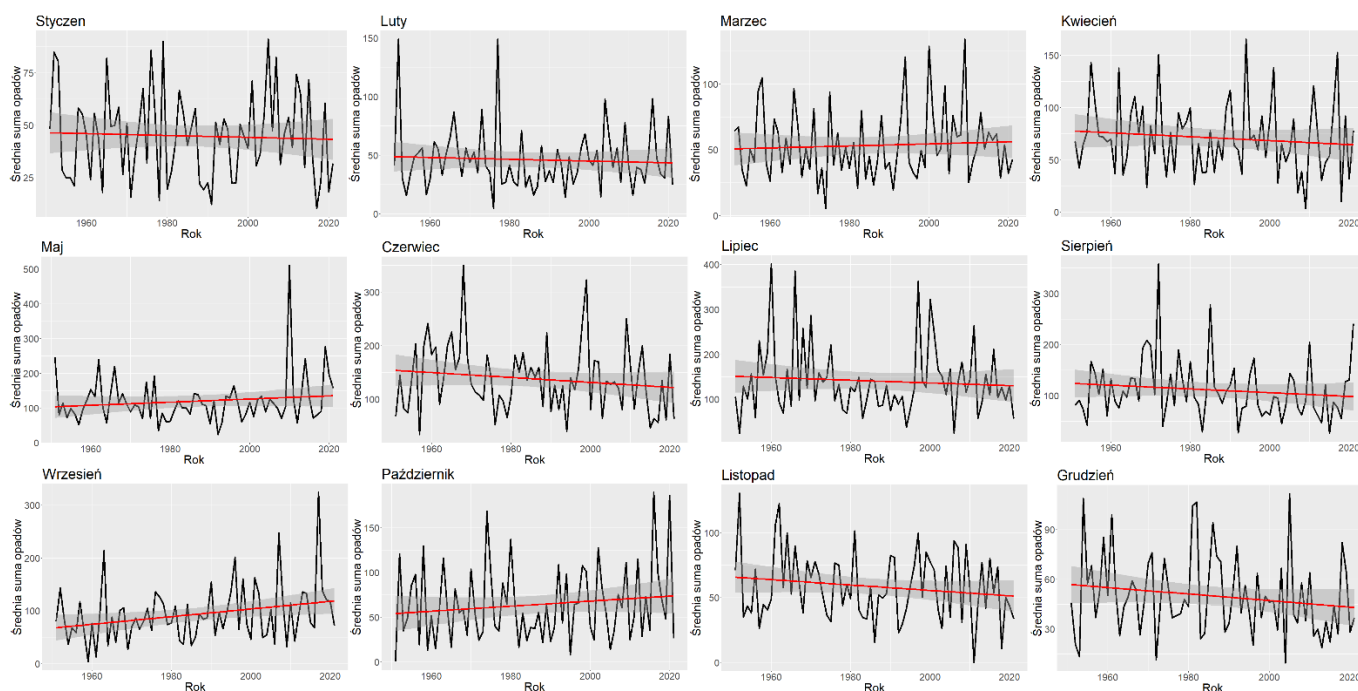


Rysunek 2 Wykres zbiorczy ramka-wąsy prezentujący rozkłady temperatur dla wszystkich miesięcy.

Bazując na wykresach liniowych z dodanymi modelami liniowymi badających zmienność sumy opadów dla poszczególnych miesięcy (rys. 3) można zaobserwować lekki trend spadkowy sumy opadów na przestrzeni kolejnych lat dla większości miesięcy.

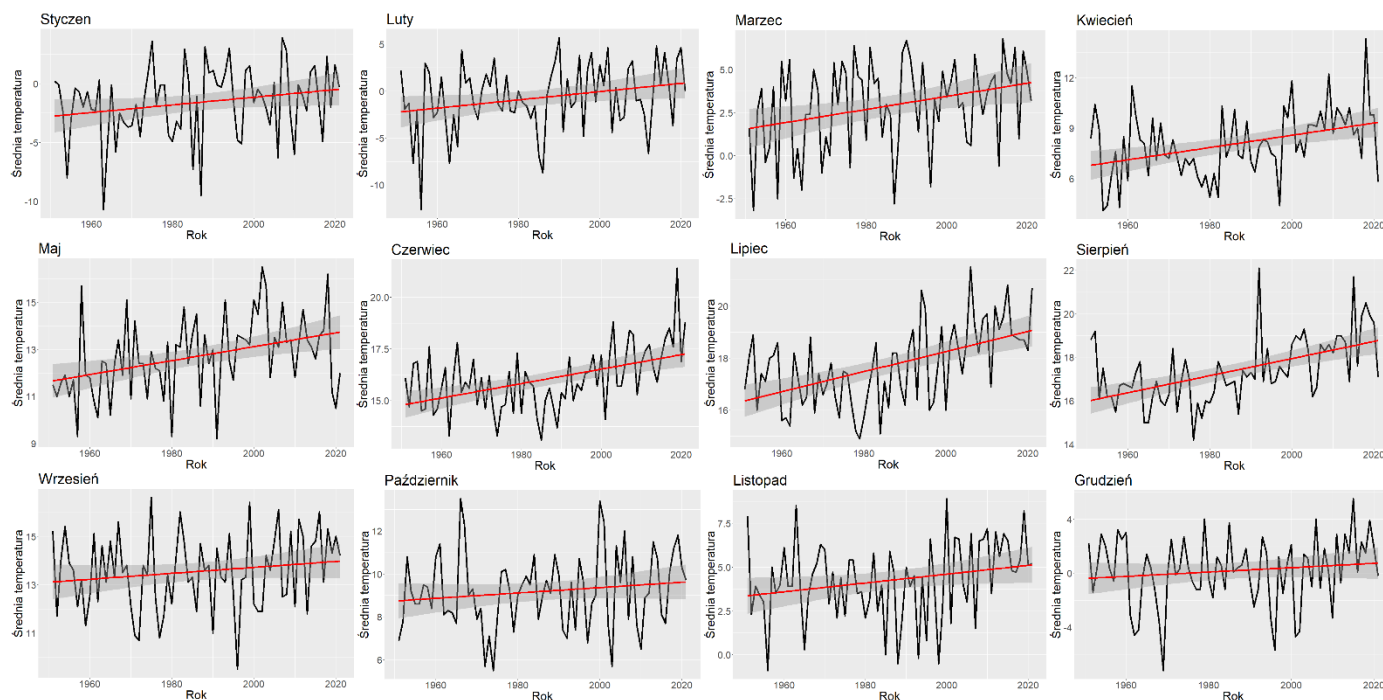
Jedynymi miesiącami z trendem wzrostowym sumy opadów są marzec, maj, wrzesień i październik. Z czego najwyraźniej można to zaobserwować we wrześniu i październiku.

Największy spadek sum opadów można zauważyć w grudniu.



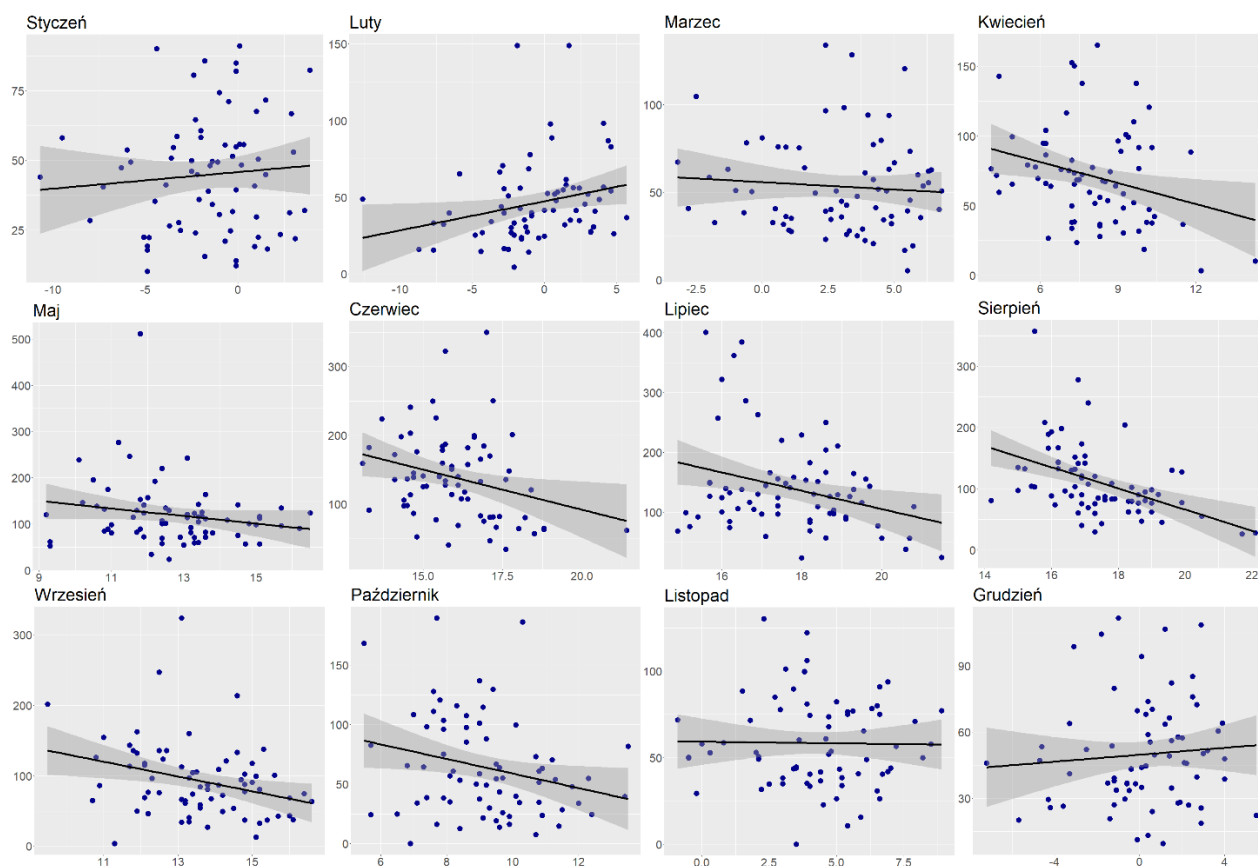
Rysunek 3 Zmienność sum opadów dla poszczególnych miesięcy na przestrzeni lat.

Porównując zmienność danych opadowych z temperaturowymi dla kolejnych miesięcy (rys. 4) można zauważyć, że trend wzrostowy temperatury we wrześniu i październiku, gdzie jest największy trend wzrostowy sum opadów, jest mniejszy niż w pozostałych miesiącach.



Rysunek 4 Zmienność średniej temperatury dla poszczególnych miesięcy na przestrzeni lat.

Związek pomiędzy temperaturą, a opadami w Bielsko-Białej (rys.5) jest najbardziej zauważalny w lutym, sierpniu i wrześniu.



Rysunek 5 Wykresy rozrzutu z dodaną prostą regresji pokazujące związek pomiędzy temperaturą, a opadami w Bielsko-Białej.

Analizując otrzymane wartości współczynników korelacji (tab. 1) można zauważyć, że związek pomiędzy temperaturą, a opadami w Bielsko-Białej jest raczej niski.

Znak współczynnika wskazuje na kierunek zależności. Jeśli wraz ze wzrostem wartości jednej zmiennej odpowiednio wzrastają wartości drugiej to znak jest dodatni, natomiast jeśli wraz ze spadkiem wartości jednej zmiennej odpowiednio maleją wartości drugiej to znak jest ujemny.

Natomiast wartość bezwzględna współczynnika korelacji wskazuje na siłę istniejącej korelacji. Im wartość ta jest większa tym korelacja jest silniejsza.

Najwyższe wartości współczynnika korelacji, zarówno Pearsona jak i Spearmana zanotowany został w sierpniu.

W przypadku współczynnika korelacji Pearsona obliczenie istotności statystycznej pozwoliło na zakwalifikowanie stycznia, marca, maja, listopada i lutego jako miesięcy o nie istotnej korelacji. Pozostałe miesiące według tego współczynnika wykazują korelację liniową, mają jednak w większości małą siłę.

Natomiast po interpretacji wartości współczynnika korelacji rang Spearmana okazało się, że styczeń, marzec, lipiec, listopad i grudzień mają najslabszą korelację (oznaczający praktycznie brak związku), zaś sierpień najwyższy (oznaczający, że zależność ta jest istotna).

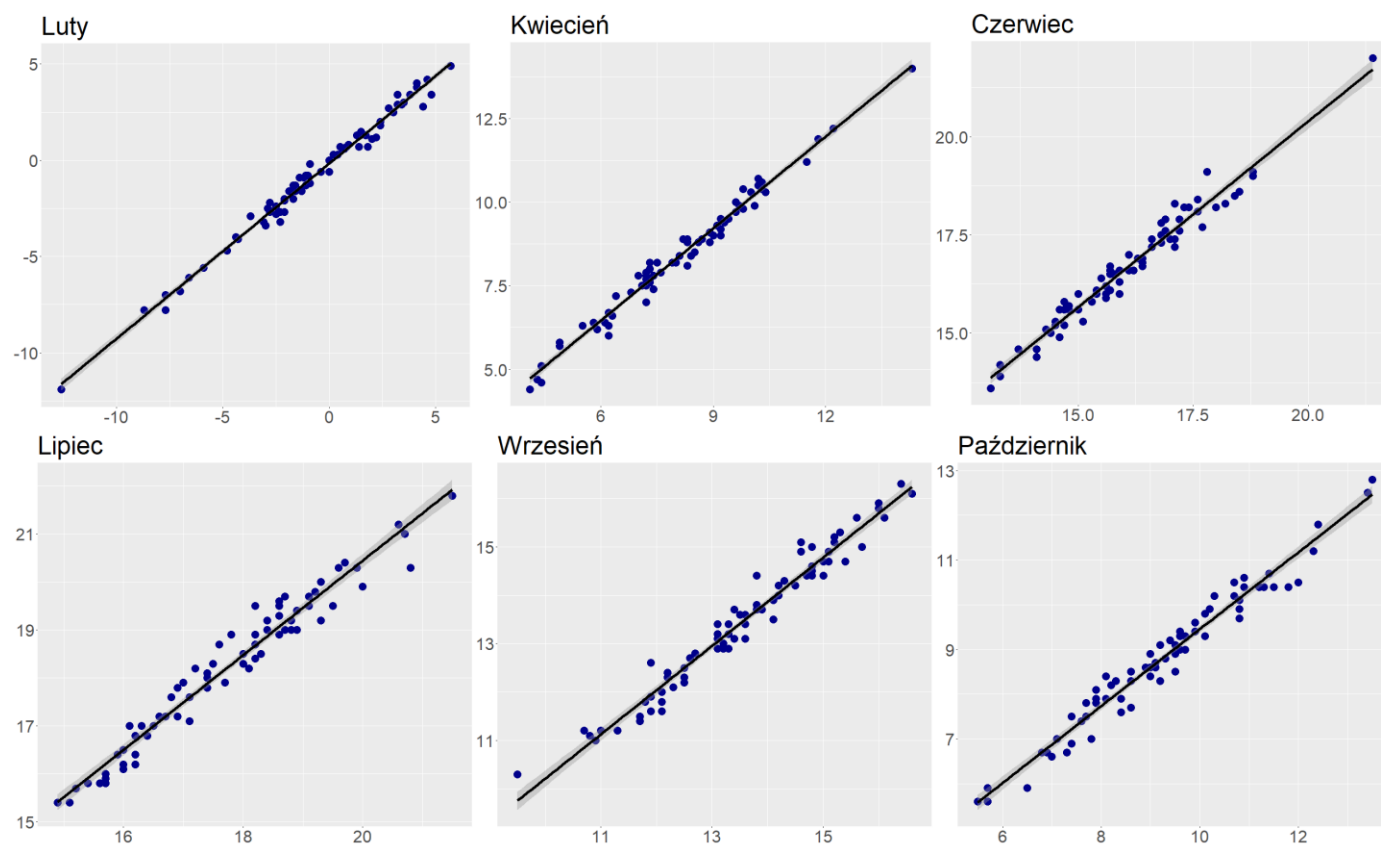
Największą zauważalną różnicę między wartościami współczynnika korelacji Pearsona i Spearmana można zauważyć w lipcu, gdzie współczynnik Spearmana jest zdecydowanie niższy. Może to być spowodowane bardzo dużym zróżnicowaniem sumy opadów przy mniejszych temperaturach (ok. 15-17°C). Notuje się wtedy zarówno bardzo duże jak i małe wartości.

Tabela 1 Wartości współczynników korelacji dla średnich temperatur i sum opadów w Bielsko-Białej.

Miesiąc	Współczynnik korelacji Pearsona	Istotność statyczna korelacji Pearsona	Współczynnik korelacji rang Spearmana	Interpretacja współczynnika korelacji rang Spearmana
Styczeń	0.0862823	nie istotna	0.083405	słaba
Luty	0.2531775	istotna	0.325504	niska
Marzec	-0.0805368	nie istotna	-0.059339	słaba
Kwiecień	-0.2833363	istotna	-0.263111	niska
Maj	-0.1916394	nie istotna	-0.203479	niska
Czerwiec	-0.2796757	istotna	-0.318135	niska
Lipiec	-0.3003604	istotna	-0.191173	słaba
Sierpień	-0.4499529	istotna	-0.530468	umiarkowana
Wrzesień	-0.3013383	istotna	-0.320751	niska
Październik	-0.2547897	istotna	-0.262702	niska
Listopad	-0.0159049	nie istotna	-0.030502	słaba
Grudzień	0.0817216	nie istotna	0.124160	słaba

Porównując średnie temperatury w Bielsko Białej z średnimi temperaturami w Katowicach dla wybranych sześciu miesięcy (rys. 6) można zauważyć silną korelację.

Temperatury w tych sąsiadujących miastach w danych miesiącach były bardzo podobne przez ostatnie 70 lat.



Rysunek 6 Wykresy rozrzutu z dodaną prostą regresji pokazujące związek pomiędzy temperaturą w Bielsko-Białej, a w Katowicach.

Wartości współczynników korelacji dla tych temperatur (tab. 2) potwierdzają obserwacje dokonane przy analizie wykresów rozrzutu (rys. 6).

Dla wszystkich sześciu miesięcy istotność statystyczna korelacji Pearsona jest „istotna”.

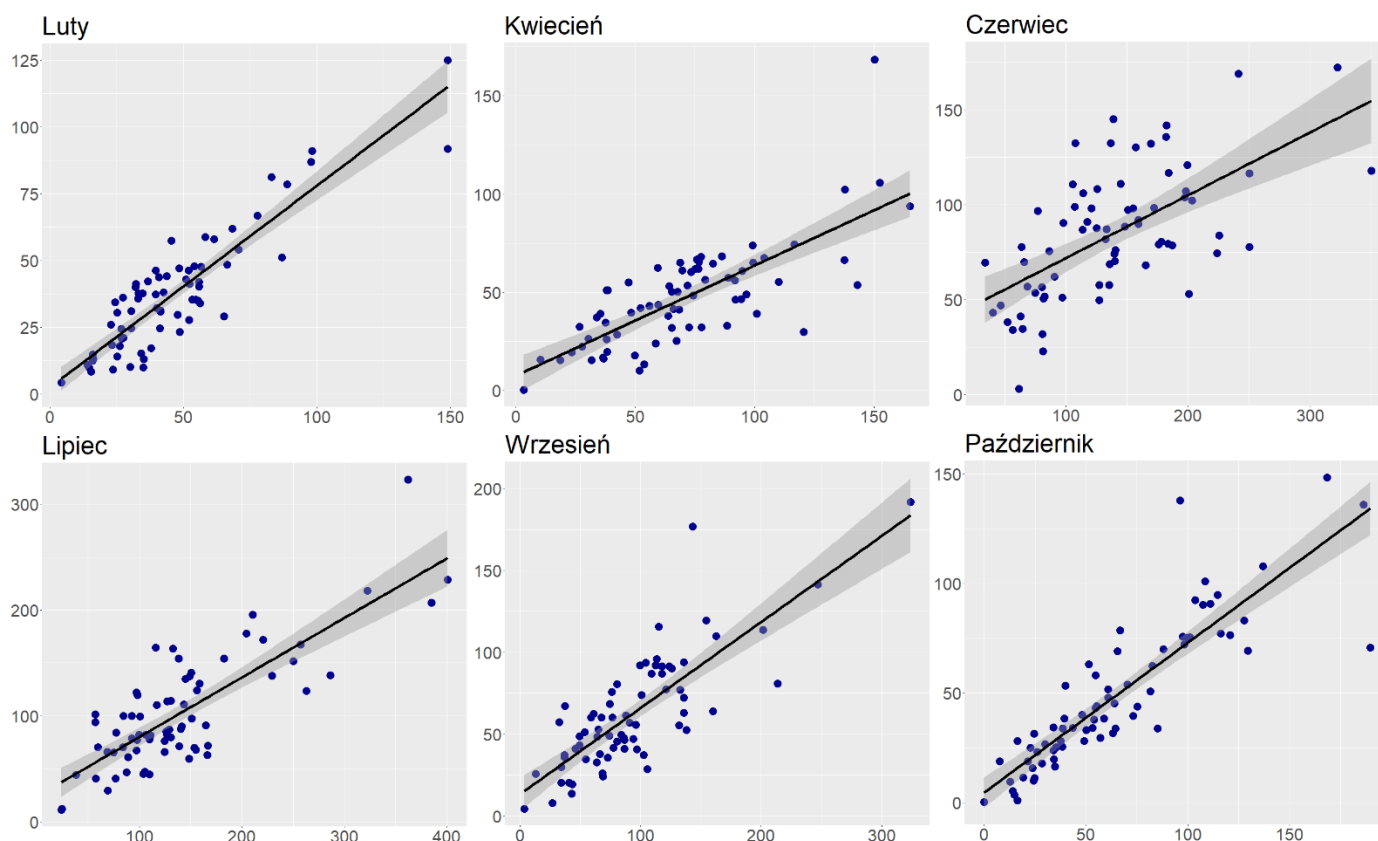
Interpretacja współczynnika korelacji rang Spearmana jest „praktycznie pełna” dla wszystkich miesięcy.

Tabela 2 Wartości współczynników korelacji dla średnich temperatur w Bielsko-Białej i Katowicach.

Średnie temperatury powietrza				
Miesiąc	Współczynnik korelacji Pearsona	Istotność statyczna korelacji Pearsona	Współczynnik korelacji rang Spearmana	Interpretacja współczynnika korelacji rang Spearmana
Luty	0.9932721	istotna	0.9899851	praktycznie pełna
Kwiecień	0.9897824	istotna	0.9846438	praktycznie pełna
Czerwiec	0.9808014	istotna	0.9809071	praktycznie pełna
Lipiec	0.9770537	istotna	0.9779465	praktycznie pełna
Wrzesień	0.9807653	istotna	0.9798773	praktycznie pełna
Październik	0.9807286	istotna	0.9800624	praktycznie pełna

Porównując sumy opadów w Bielsko Białej z sumami opadów w Katowicach dla tych samych sześciu miesięcy (rys. 7) również można zauważyć korelację.

Najmniejsza korelacja występuje w czerwcu.



Rysunek 7 Wykresy rozrzutu z dodaną prostą regresji pokazujące związek pomiędzy sumami opadów w Bielsko-Białej, a w Katowicach.

Analizując wartości współczynników korelacji dla tych opadów (tab. 3) okazuje się, że najsilniejsza korelacja (praktycznie pełna) zachodzi w październiku, a najsłabsza (choć wciąż wysoka) w czerwcu.

Tabela 3 Wartości współczynników korelacji dla sum opadów w Bielsko-Białej i Katowicach.

Miesięczne sumy opadów				
Miesiąc	Współczynnik korelacji Pearsona	Istotność statyczna korelacji Pearsona	Współczynnik korelacji rang Spearmana	Interpretacja współczynnika korelacji rang Spearmana
Luty	0.8950933	istotna	0.8120331	bardzo wysoka
Kwiecień	0.7535306	istotna	0.7314145	wysoka
Czerwiec	0.6197297	istotna	0.6097804	wysoka
Lipiec	0.8079424	istotna	0.6738041	wysoka
Wrzesień	0.7992651	istotna	0.7528317	wysoka
Październik	0.8745265	istotna	0.9164214	praktycznie pełna

Silniejsza korelacja występuje przy średnich miesięcznych temperaturach (tab. 2), niż miesięcznych sumach opadów (tab. 3) w obu miastach. Może to być spowodowane położeniem Bielsko-Białej tuż obok łańcucha górskiego, co wpływa na ilości opadów w tym mieście.

5. Podsumowanie

Miesięczna suma opadów w Bielsko-Białej charakteryzuje się zmiennością w skali roku. W ciągu ostatnich 70 lat najczęściej najwięcej opadów było w lipcu, a najmniej w styczniu.

Najwięcej opadów w Bielsko-Białej występuje w miesiącach letnich, gdy jest najwyższa temperatura powietrza.

Dla większości miesięcy można zaobserwować lekki trend spadkowy sumy opadów na przestrzeni kolejnych lat.

Współczynniki korelacji są znacznie mniejsze przy porównywaniu sum opadów i średnich temperatur w jednym mieście, niż przy porównywaniu tego samego typu danych w dwóch najbliższych miastach.

5. Literatura

https://meteomodel.pl/dane/srednie-miesieczne/?imgwid=349190600&par=tm&max_empty=2 – dane meteorologiczne

<https://towardsdatascience.com/clearly-explained-pearson-v-s-spearman-correlation-coefficient-ada2f473b8> - Pearson vs Spearman