

Rachunek Prawdopodobieństwa i Statystyka

Rozwiązanie zadania z kolokwium nr 2

Wojciech Adamiec, Michał Mikołajczyk

2 maja 2020

Treść zadania:

Dokonaj analizy danych dotyczących zarażeń wirusem COVID-19 opublikowanych przez WHO pod kątem korelacji zachodzącej między przebiegami pandemii w różnych krajach. Znajdź takie przesunięcie czasowe, aby współczynnik korelacji był możliwie największy.

Rozwiązanie:

Policzenie współczynnika korelacji nie stanowi żadnego problemu, funkcja go obliczająca jest dostępna w niemal każdym arkuszu kalkulacyjnym. Dokładny sposób liczenia owego współczynnika można znaleźć np. tu: Współczynnik Korelacji.

Do naszej analizy zdecydowaliśmy się wybrać następujące kraje: Niemcy, Wielka Brytania, Francja, Rosja, Polska oraz Szwecja. Będziemy chcieli między każdymi dwoma z tych krajów dokonać krótkiej analizy możliwych współczynników korelacji. Aby wyniki były bardziej wiarygodne za pierwszy dzień zarażenia będziemy uznawać dzień, który rozpoczyna ciągłe wzrosty ilości zarażonych - to znaczy nie będziemy rozpatrywać bardzo fragmentarycznych danych początkowych. Przez przesunięcie czasowe o n dni będziemy rozumieli takie sparowanie danych o zarażeniach z dwóch krajów, że w jednej parze x, y będą dane z dnia d oraz $d + n$. Przesunięcie czasowe powoduje niestety, że nie będziemy mogli skorzystać ze wszystkich aktualnych danych.

Warto oczywiście spojrzeć na maksymalne współczynniki korelacji, między przebiegami pandemii w różnych krajach, jakie udało nam się osiągnąć.

Maksymalne współczynniki korelacji						
Kraj	Niemcy	Wlk. Bryt	Francja	Rosja	Polska	Szwecja
Niemcy	-	0.86	-0.79	0.9	-0.85	-0.88
Wlk. Bryt	0.86	-	0.84	0.97	0.91	0.92
Francja	-0.79	0.84	-	0.92	0.78	0.75
Rosja	0.9	0.97	0.92	-	0.89	0.93
Polska	-0.85	0.91	0.78	0.89	-	0.86
Szwecja	-0.88	0.92	0.75	0.93	0.86	-

Ze względu na mnogość danych uznaliśmy, że najlepiej będzie do dokumentacji dołączyć arkusz kalkulacyjny, które zawiera wszystkie przeprowadzone przez nas obliczenia. Można w nim znaleźć współczynniki korelacji dla wszystkich rozpatrywanych przez nas przesunięć czasowych oraz konfiguracje dające największe współczynniki.

Wydaje nam się, że warto pochylić się na chwilę nad kilkoma spostrzeżeniami:

1. Maksymalny współczynnik korelacji między Polską a Wielką Brytanią jest wyliczany dla przesunięcia zerowego.
2. Odnotowujemy duże współczynniki korelacji między Niemcami a Rosją, gdy przesunięcie w czasie jest po stronie Niemieckiej - w oczywisty sposób taki współczynnik nie może być wiarygodny, bo wiemy że w Rosji epidemia zaczęła się zdecydowanie później.
3. Największy odnotowany współczynnik korelacji jest między Wielką Brytanią a Rosją, dla przesunięcia Rosji o 15 dni do przodu.