

## Projekt 2

### Testowanie hipotez statystycznych

Sprawozdanie z projektu powinno składać się z dwóch plików:

**projekt2.html** zawierającego raport z wykonania projektu (plik **projekt2.html** należy wygenerować z pliku **projekt2.Rmd** siłami pakietu knitr),

**projekt2.zip** zawierającego skompresowany katalog z pełnymi źródłami raportu (w tym plikiem **projekt2.Rmd** oraz wszelkimi wykorzystywanymi do jego wygenerowania danymi).

Po rozpakowaniu na serwerze lab09011 pliku **projekt2.zip**, plik **projekt2.Rmd** powinien bezbłędnie kompilować się siłami pakietu knitr do pliku **projekt2.html**.

**Zadanie 1.** (15p) Korzystając z danych dostępnych w internecie (patrz uwagi poniżej):

- Sprawdź, czy wirus SARS-CoV-2 równomiernie rozprzestrzenił się na terenie Polski, tzn. czy liczba osób zarażonych wirusem w poszczególnych województwach jest proporcjonalna do populacji tych województw.
- Sprawdź, czy skuteczność leczenia osób zarażonych wirusem SARS-CoV-2 jest taka sama na terenie całej Polski, tzn. czy liczba przypadków śmiertelnych jest proporcjonalna do liczby osób zarażonych przy podziale na województwa.
- W każdym z powyższych przypadków wskaż ewentualnie, które województwa istotnie się od siebie różnią pod kątem rozważanych statystyk.

Zestaw województw do przeanalizowania dla każdego zespołu zostanie podany w osobnym pliku.

**Zadanie 2.** (15p) Sprawdź tezę mówiącą o tym, że tzw. atut własnego boiska w ligowym meczu piłkarskim nie zależy od kraju, w którym toczą się dane rozgrywki ligowe. W tym celu, zweryfikuj hipotezę o niezależności wyniku meczu, rozumianego jako zwycięstwo, porażkę lub remis gospodarzy, od kraju rozgrywek. Dane do przeprowadzenia testu można znaleźć na stronie <https://pl.fcstats.com>. Dane te znajdują się w zakładce „Porównanie lig” / „Zwycięzcy meczów” w tabeli, w której w poszczególnych kolumnach „1”, „x” i „2” podane są, odpowiednio, liczby zwycięstw, remisów i porażek gospodarzy. Zestaw lig do przeanalizowania dla każdego zespołu zostanie podany w osobnym pliku.

W ramach rozwiązania zadania zaimplementuj odpowiednie dla testu niezależności wyrażenia, na podstawie których obliczane są: statystyka testowa, wartość krytyczna oraz p-wartość. Przeprowadź wnioskowanie porównując wartość otrzymanej statystyki testowej z wartością krytyczną, przy poziomie istotności 0,05. Czy obliczona p-wartość potwierdza wynik wnioskowania? Sprawdź uzyskane wyniki wykorzystując funkcję `chisq.test`.

#### Uwagi:

- W zadaniu 1 nie trzeba implementować funkcji do przeprowadzania testów (należy wykorzystać istniejące funkcje języka R).
- Dane do zadania 1 można znaleźć np. pod adresami:
  - <https://koronawirus-w-polsce.pl>
  - <https://warszawa.twoje-miasto.pl/koronawirus>
  - [https://docs.google.com/spreadsheets/u/1/d/1ierEhD6gcq51HAm433knjnVwey4ZE5DCnu1bW7PRG3E/htmlview?fbclid=IwAR0oqN4iikk1f1MvFbb0ecLOKGpkHo7LCYj75N6Z1vc4n\\_kuhtI214dqN5Ho](https://docs.google.com/spreadsheets/u/1/d/1ierEhD6gcq51HAm433knjnVwey4ZE5DCnu1bW7PRG3E/htmlview?fbclid=IwAR0oqN4iikk1f1MvFbb0ecLOKGpkHo7LCYj75N6Z1vc4n_kuhtI214dqN5Ho)
- W punktach a) i b) zadania 1 należy wyznaczyć p-wartości odpowiednich testów oraz napisać jaką konkluzję zakończyłyby się te testy gdyby przeprowadzać je na poziomie istotności 0,05.