

## Projekt 2

### Testowanie hipotez statystycznych

Wynikiem wykonania projektu powinny być dwa pliki:

- plik `projekt2.html` (wygenerowany z odpowiedniego pliku `projekt2.Rmd`) zawierający raport z wykonania projektu
- plik `projekt2.zip` stanowiący skompresowany katalog zawierający pełne źródła raportu. W szczególności w katalogu tym musi znaleźć się plik `projekt2.Rmd` oraz wszelkie dane niezbędne do wygenerowania pliku `projekt2.html`.

Uwagi:

1. Oddają Państwo jeden zestaw plików na zespół (zespół definiuje kolumna `numer_zespołu` pliku `projekt2_par.csv`). Zestaw ten wgrywa na platformę Teams dokładnie jedna (dowolna) z osób należących do danego zespołu.
2. Po rozpakowaniu Państwa pliku `projekt2.zip` na serwerze lab09011 plik `projekt2.Rmd` powinien bezbłędnie kompilować się siłami pakietu knitr do pliku `projekt2.html`!
3. W pliku `projekt2.html` powinny być widoczne wszystkie fragmenty kodu.

**Zadanie 1.** (15p) Korzystając z danych dostępnych w internecie<sup>1</sup>:

- a) Sprawdź, czy proces szczepień przeciw SARS-CoV-2 przebiega równie szybko we wszystkich województwach, tzn. czy liczba szczepień jest proporcjonalna do populacji tych województw.
- b) 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 w poszczególnych województwach.
- c) 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.

**Uwaga:** W zadaniu tym nie trzeba implementować funkcji do przeprowadzania testów (należy wykorzystać istniejące funkcje języka R). W punktach a) i b) należy wyznaczyć p-wartości odpowiednich testów oraz napisać jaką konkluzję zakończyłyby się te testy gdyby przeprowadzać je na poziomie istotności 0,05.

**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`.

---

<sup>1</sup>np. <http://bit.ly/covid19-poland>