

# Projekt 1

## Elementy statystyki opisowej i estymacji punktowej

Wynikiem wykonania projektu powinny być dwa pliki:

- plik `projekt1.html` (wygenerowany z odpowiedniego pliku `projekt1.Rmd`) zawierający raport z wykonania projektu
- plik `projekt1.zip` stanowiący skompresowany katalog zawierający pełne źródła raportu. W szczególności w katalogu tym musi znaleźć się plik `projekt1.Rmd` oraz wszelkie dane niezbędne do wygenerowania pliku `projekt1.html`.

Uwagi:

1. Oddają Państwo jeden zestaw plików na zespół (zespół definiuje kolumna `numer_zespolu` pliku `projekt1_par.csv`). Zestaw ten wgrywa na platformę Teams dokładnie jedna (dowolna) z osób należących do danego zespołu.
2. Po rozpakowaniu Państwa pliku `projekt1.zip` na serwerze lab09011 plik `projekt1.Rmd` powinien bezbłędnie kompilować się siłami pakietu knitr do pliku `projekt1.html`!
3. W pliku `projekt1.html` powinny być widoczne wszystkie fragmenty kodu.

**Zadanie 1.** (10p) Dla zadanych spółek<sup>1</sup> notowanych na WGPW, na podstawie ich notowań z drugiej połowy roku 2019 (z pliku `zad1mst.zip`),

- a. wyznacz procentowe zmiany cen zamknięcia tych spółek,
- b. zilustruj rozkłady ww. zmian (histogramy + wykresy pudełkowe),
- c. wyestymuj parametry rozkładów normalnych mogących modelować ww. rozkłady,
- d. porównaj graficznie rozkłady modelowe z danymi.

**Zadanie 2.** (10p) Na podstawie danych z października 2019 (z pliku `zad2csv_new.zip`) dotyczących zadanej spółki<sup>2</sup>:

- a. zilustruj jak wolumen transakcji rozkłada się pomiędzy 3 fazy notowań: otwarcie<sup>3</sup>, notowania ciągłe<sup>4</sup>, zamknięcie z dogrywką<sup>5</sup>.
- b. zilustruj jak wolumen transakcji rozkłada się w czasie notowań ciągłych,
- c. wyznacz dzień i jego 2-godzinny przedział czasu rozpoczynający się o pełnej godzinie między 10:00 a 14:00 włącznie, w którym jest najmniej sekund, w których przeprowadzane były transakcje na akcjach danej spółki<sup>6</sup>. Dla wyznaczonego przedziału czasu:
  - i) W każdej kolejnej minucie rozważanego przedziału wyznacz liczbę sekund, w których przeprowadzane były transakcje (otrzyma się w ten sposób 120 liczb z przedziału  $[0, 60]$ ),
  - ii) zamodeluj ww. liczby za pomocą rozkładu Poissona,
  - iii) porównaj rozkład modelowy z danymi.

---

<sup>1</sup>Kolumna `zad1ticker` pliku `projekt1_par.csv`; liczba spółek analizowanych w pierwszym zadaniu jest taka jak liczba osób w zespole.

<sup>2</sup>Kolumna `zad2ticker` pliku `projekt1_par.csv`.

<sup>3</sup>Wszystkie transakcje z godziny 9:00:00, które mają tę samą cenę co pierwsza transakcja z tej godziny.

<sup>4</sup>Wszystkie transakcje pomiędzy godzinami 9:00 a 16:51, które nie są transakcjami z fazy otwarcia

<sup>5</sup>Wszystkie transakcje po godzinie 16:59.

<sup>6</sup>Jeden wiersz pliku z danymi to jedna transakcja. Zwróć uwagę, że w jednej sekundzie mogło być wiele transakcji (taką sekundę zlicz tylko raz).