

**SAD 2023Z**  
**Część 2 Projektu**  
**Termin oddania: 27 stycznia 2024**

Efektem projektu powinien być raport w formacie pdf. Dla każdego problemu powinien on zawierać:

- a) Metodę rozwiązania problemu
- b) Otrzymane wyniki zaprezentowane w przejrzystej formie
- c) Interpretacja wyników i/lub komentarze i/lub wnioski

Wraz z raportem należy też przekazać kody źródłowe.

**Problem 1 – weryfikacja hipotez związanych z wypadkami drogowymi**

Na podstawie danych z wybranego roku dostępnych pod adresem

<https://www.ciekawestatystyki.pl/2023/05/wypadki-drogowe-w-polsce-i-w-europie.html>

proszę zbadać następujące dwie hipotezy:

1. Wypadki drogowe rozkładają się równomiernie na wszystkie dni tygodnia
2. Wypadki drogowe rozkładają się „równomiernie” na wszystkie miesiące w ciągu roku (cudzystów, bo należy uwzględnić to, że różne miesiące w roku mają różne liczby dni).

Przyjąć poziom istotności testu np. 0.05.

Uwaga. Zamiast liczby wypadków można zbadać np. liczbę rannych lub liczbę zabitych.

**Problem 2 – badanie testów**

Wybrać dwie metody sprawdzenia normalności rozkładu – można zerknąć np. na stronę

[https://en.wikipedia.org/wiki/Normality\\_test](https://en.wikipedia.org/wiki/Normality_test) (polecam wybranie takich metod, które są zaimplementowane w pakiecie R i gotowe do użycia). Wybrać poziom istotności testu (np. 0.05) i rozmiar próby losowej (np.  $n=20$ ).

Następnie przeprowadzić symulacje komputerowe (w pętli wielokrotnie generować dane, przeprowadzać oba testy i zliczać ich wyniki) w celu oszacowania:

1. Czy poziom istotności zgadza się z założonym;  
wskazówka: generacja danych zgodnych z  $H_0$  (funkcja `rnorm`) i zliczanie fałszywych decyzji za  $H_1$ .
2. Sprawdzenie, który z wybranych testów ma większą moc w przypadku danych z rozkładu  $t$ -Studenta. Badania przeprowadzić dla 3 wybranych wartości liczb stopni swobody rozkładu  $t$ . (tzn. chodzi o czułość, zdolność testów do wykrycia tego, że podane dane nie są z rozkładu normalnego).  
wskazówka: generacja danych zgodnie z  $H_1$  (funkcja `rt`) i zliczanie prawidłowych decyzji  $H_1$

### Problem 3 – estymacja

W problemie 2 projektu 1 był badany rozkład *odstępów między impulsami* w detektorze promieniowania gamma (każdy impuls odpowiadał interakcji fotonu gamma z detektorem – błysk światła, który następnie był wzmacniany w fotopowielaczu i rejestrowany). Z teorii wiadomo, że rozkład ten powinien być *wykładniczy*. Teraz proszę wyestymować wartość parametru  $\lambda$  tego rozkładu wykładniczego, a następnie w celu prostego sprawdzenia poprawności nałożyć na siebie odpowiednio przeskalowany histogram i funkcję gęstości prawdopodobieństwa rozkładu wykładniczego z wyestymowanym parametrem, tzn.  $\text{Exp}(\hat{\lambda})$ .

**Uwaga.** Problem wcale nie jest prosty, ponieważ nie jesteśmy w stanie odzyskać z sygnału informacji o impulsach położonych bardzo blisko siebie (bliżej niż wynosi długość impulsu, tzn. impulsów nałożonych). Należy więc opracować nową (być może nie istniejącą nigdzie w literaturze) metodę estymacji rozkładu wykładniczego na podstawie niepełnych „ocenzurowanych” danych.

Uwaga: Jest to kontynuacja problemu 2 z projektu 1, więc należy bazować na tych samych danych `signal_50MHz.bin` (i pewnie częściowo również na tych samych kodach).

**Wskazówka:** W celu wyznaczenia estymatora można:

1. Określić jak wygląda funkcja gęstości prawdopodobieństwa (fgp) uciętego rozkładu wykładniczego
  - a. Próg ucięcia oznaczyć np. przez  $x_0$  i zapisać wórk
  - b. Upewnić się, że fgp całkuje się do 1
2. Wyznaczyć estymator największej wiarygodności
  - a. Zapisać funkcję wiarygodności
  - b. Następnie logarytmiczną funkcję wiarygodności
  - c. Obliczyć jej pochodną i wyznaczyć wartość parametru  $\lambda$ , dla której pochodna ta = 0