

Oferta seminarium licencjackiego 2019/2020
*Analiza nieklasycznych źródeł danych z wykorzystaniem pakietów
statystycznych*

dr Maciej Beręsewicz
Katedra Statystyki
maciej.beresewicz@ue.poznan.pl, [Strona Internetowa](#),
[Google Scholar \(publikacje\)](#), [Baza Wiedzy UEP](#), [Baza Osiągnięć UEP](#)

Kierunek studiów: Informatyka i Ekonometria, **Specjalność:** wszystkie specjalności.

Słowa kluczowe: R, pakiety statystyczne, statystyka, statystyka publiczna, modelowanie, otwarte dane, big data, internetowe źródła danych, data science.

Sposób zapisów: 1) kontakt mailowy w celu umówienia spotkania, 2) rozmowa kwalifikacyjna, 3) decyzja po odbyciu wszystkich rozmów.

Opis seminarium

Seminarium przeznaczone jest dla studentów zainteresowanych metodami ilościowymi, nowymi (nieklasycznymi) źródłami danych¹ oraz nauką profesjonalnych narzędzi do analizy danych. Szukam studentów, którym nie jest obce pojęcie `data science` i którzy wolą pisać własne programy zamiast klikać. Podczas seminarium poruszana będzie tematyka wykorzystania nowych źródeł danych na potrzeby statystyki publicznej oraz wykorzystania oprogramowania typu open source.

Przykładowa tematyka prac:

- Analiza portali ogłoszeniowych (np. oferty pracy, nieruchomości),
- Analiza jakości i deduplikacja danych portali ogłoszeniowych,
- Prognozowanie z wykorzystaniem Google Trends,
- Wykorzystanie otwartych danych,
- Porównanie pakietów statystycznych (np. R, Python::Scipy oraz Julia),
- Analiza porównawcza method imputacji danych,
- Inne: problematyka pomiaru cudzoziemców w Polsce.

Dodatkowe informacje na temat seminarium można uzyskać podczas konsultacji.

Wymagania

- Dobra znajomość statystyki oraz ekonometrii.
- Dobra znajomość języka angielskiego (umożliwiająca czytanie tekstów angielskojęzycznych).
- Podstawowa znajomość następujących narzędzi: **R** (np. pakiety [tidyverse](#)), Python (np. [pandas](#)), [Julia](#) lub [SAS \(University Edition\)](#). Mile widziany ukończony kurs [Pogromcy Danych](#) lub inny kurs poświęcony profesjonalnym narzędziom do analizy danych.
- Mile widziana znajomość narzędzi big data, systemu linux, systemów kontroli wersji (np. git + github), Markdown, [L^AT_EX](#) i [B_IB_TE_X](#).

¹Klasyczne (statystyczne) źródła danych to: spisy powszechnie, badania częściowe czy sprawozdawczość. Nieklasyczne (niestatystyczne) to głównie: rejestry administracyjne, dane pochodzące z Internetu czy tzw. big data.

Kryteria przyjęć

- Rozmowa kwalifikacyjna.
- Oceny ze studiów.
- Zainteresowanie tematyką oraz chęć zgłębiania wiedzy.

Zasady współpracy

- Praca może być przygotowywana w j. polskim lub j. angielskim.
- Praca powinna być przygotowana w \LaTeX lub Markdown (preferowane połączenie z **R** przez pakiet [bookdown](#)).
- Analiza danych wyłącznie w oprogramowaniu typu open source (np. R, Python, Julia).
- Finalna wersja pracy powinna być oddana miesiąc przed obroną.

Proponowane zaliczenie seminarium

- Udział w comiesięcznym seminarium, na którym będziemy dyskutować o analizie danych oraz uczyć się profesjonalnych narzędzi statystycznych.
- Sugerowane zaliczenie poszczególnych semestrów (według roku akademickiego; przykład dla osób, które przeprowadzą analizę danych w R; ostateczne zasady mogą ulec zmianie):
 - **semestr V**
 - * ukończenie kursu [Pogromcy Danych](#) i rozwiązanie wszystkich zadań (do połowy semestru),
 - * przerobienie książki [Język R](#) i rozwiązanie wszystkich zadań (do końca semestru),
 - * przygotowanie krótkiego dokumentu w Markdown i \LaTeX (do połowy semestru),
 - * przedstawienie tematu i planu pracy (do końca pierwszego miesiąca),
 - * przedstawienie minimum jednego rozdziału pracy (do końca semestru).
 - **semestr VI**
 - * przerobienie I i II rozdziału z książki [Analiza danych z programem R](#) (do połowy semestru),
 - * przerobienie książki [Programowanie w języku R \(wyd. 2\)](#) i rozwiązanie wszystkich zadań (do końca semestru),
 - * przygotowanie krótkiej prezentacji z wynikami w Beamer (\LaTeX) (do połowy semestru),
 - * przedstawienie pierwszej (pełnej, zawierającej wszystkie rozdziały) wersji pracy w Markdown lub \LaTeX (do połowy semestru).
 - * oddanie pełnej pracy na miesiąc przed obroną.

Informacje dodatkowe

- Czym jest R – [https://pl.wikipedia.org/wiki/R_\(język_programowania\)](https://pl.wikipedia.org/wiki/R_(język_programowania)) i jego popularność <https://www.tiobe.com/tiobe-index/r/>
- Czym jest Julia – <https://julialang.org>
- Czym jest \LaTeX – <http://www.ptm.org.pl/latex-ksiazka-kucharska>
- Czym jest \BIBTeX – https://pl.wikibooks.org/wiki/LaTeX/Zarządzanie_bibliografią
- Czym jest Markdown – <https://pl.wikipedia.org/wiki/Markdown>