

Statystyka matematyczna i ekonometria, lato 2023/24
Wytyczne do projektu zaliczeniowego – laboratorium

Projekt oddają Państwo w formie elektronicznej do dnia 20.06.2024 przesyłając na adres:
anna.skowronska-szmer@pwr.edu.pl

Projekt należy wykonać z wykorzystaniem dowolnej bazy danych, może być również baza danych z pakietu R. Niektóre etapy projektu nie będą wymagały wykorzystania tej bazy danych, a wygenerowania próby losowej wybranego rozkładu (punkt V).

Projekt powinien zawierać stronę tytułową: imiona i nazwiska oraz nr indeksów osób tworzących projekt, data powstania projektu.

Kolejne etapy projektu powinny zawierać informacje:

1. Co robię – krótki opis co na danym etapie będę robić i co chcę otrzymać, dlaczego to robię;
2. Jak to robię – wklejony kod programu w R;
3. Jaki jest efekt działania – zrzut ekranu wyników z programu w R (jeżeli jest długa tabela należy umieścić kilka pierwszych wierszy, lub tych które są ważne).
4. Interpretacja otrzymanych wyników – krótka analiza wyników, najważniejsze wnioski.

Wytyczne na ocenę dostateczną, każda rzecz zrobiona więcej podnosi ocenę.

I. Opis bazy danych

Projekt należy rozpocząć od opisanie i przedstawienie bazy danych. Proszę podać źródło skąd baza danych pochodzi. Należy opisać z czego składa się baza danych (liczba zmiennych, obserwacji, jakie to są zmienne, itp.) – wykorzystaj jak najwięcej funkcji umożliwiających jej przedstawienie.

II. Wyliczenie podstawowych statystyk

W tej sekcji należy wyliczyć statystyki dla wybranych zmiennych. Dla minimum dwóch zmiennych, minimum 6 statystyk. Zbuduj z nich ramkę danych.

III. Wykresy

Dla wybranych zmiennych narysuj minimum:

- histogram,
- wykres liniowy,
- wykres słupkowy,
- wykres pudełko – wąsy,
- wykres gęstości.

IV. Obserwacje odstające

Przeanalizuj wybrane zmienne pod względem występowania obserwacji odstających. Zastosuj minimum dwie różne metody sprawdzania występowania obserwacji odstających.

V. Wyliczanie prawdopodobieństwa dla zmiennej.

Wygeneruj próbę losową dla wybranego rozkładu dyskretnego i ciągłego. Oblicz prawdopodobieństwo punktowe i przedziałowe dla wygenerowanej próby losowej dla dowolnych wartości x . Narysuj gęstość i dystrybuantę dla wygenerowanej próby losowej. Narysuj obszar określony prawdopodobieństwem, dla którego wykonujesz obliczenia dla zmiennej losowej ciągłej.

VI. Z dowolnych trzech zmiennych z bazy danych zbuduj macierz. Dla zbudowanej macierzy przedstaw jak najwięcej jej parametrów używając funkcji R.

VII. Wylicz po dwa dowolne przedziały ufności dla zmiennej numerycznej i jakościowej.

VIII. Przetestuj dwie dowolne hipotezy dla testów parametrycznych i dwie dla testów nieparametrycznych.

IX. Zastosuj prostą regresję liniową dla dowolnej zmiennej/ zmiennych.