

INSTRUKCJA WYKONANIA ĆWICZENIA LABORATORYJNEGO Z TEMATU „PRZEGLĄDOWA ANALIZA DANYCH W JĘZYKU R”

Informacje ogólne

Zaplanowane w tym ćwiczeniu zadania z zakresu przeglądowej (eksploracyjnej) analizy danych należy wykonać na komputerach w laboratorium lub na własnych instalacjach języka R. Materiały pomocnicze do ćwiczenia zawierają rozdziały R-2 i R-3 wykładu. W szczególności podano tam odnośniki do pobrania bezpłatnych wersji środowiska R na popularne systemy operacyjne oraz bliżej omówiono zagadnienia potrzebne do zrealizowania ćwiczenia. Efektem wykonanej pracy ma być poznanie wybranych metod i narzędzi analizy przeglądowej, a stopień opanowania wiedzy zostanie sprawdzony w formie krótkiego testu.

Zadania do wykonania

1. Przygotować własne środowisko R (można ewentualnie wykorzystać już posiadaną instalację).
2. Uważnie przestudiować cały rozdział R-2 wykładu zawierający materiał pomocniczy potrzebny do realizacji ćwiczenia.
3. Uruchomić system R i wczytać plik csv zawierający dane o samochodach używanych. Odnośnik do tego pliku zamieszczono na stronie przedmiotu.
4. Starannie i bez pośpiechu wykonać wszystkie operacje z danymi o samochodach zaprezentowane w rozdziale R-3 wykładu.
5. Na zaliczenie ćwiczenia przygotować krótkie (4-6 stron) sprawozdanie zgodne z indywidualnie przydzielonym tematem (zob. tabela). W sprawozdaniu należy wykorzystać odnośny materiał z rozdziałów R2 i R3, który może być uzupełniony z wykorzystaniem dodatkowych źródeł. Sugeruje się także opracowanie własnego przykładu ilustrującego opisywane zagadnienie na niewielkim zbiorze danych pozyskanym np. z sieci.
6. Przygotowane opracowanie, wyłącznie w formie pliku pdf, należy przesłać na adres e-mail prowadzącego w terminie do 2-ch tygodni od zakończenia okresu przewidzianego na realizację ćwiczenia. Bardzo proszę o nadanie plikom jednolitej nazwy w formacie: DSC_<Gr>_<Nazwisko>.pdf. Na przykład plik zawierający opracowanie studenta o nazwisku Kowalski z grupy L1 o powinien mieć nazwę DSC_L1_Kowalski.pdf.

Tabela indywidualnego przydziału tematów

Grupa L1	Grupa L2	Przydzielony numer tematu wg numeracji w sekcji „Tematy sprawozdań”
167559	167584	1
158869	167585	2
167564	167586	3
167565	167587	4
167567	163912	5
167572	167591	6
167575	167592	7
163897	164025	8
167578	167593	9
167582	167594	1
167583	167595	2
	167596	3

Tematy sprawozdań

1. Wizualizacja zmiennych numerycznych – wykresy pudełkowe

2. Wizualizacja zmiennych numerycznych – histogramy
3. Wariancja i odchylenie standardowe.
4. Reguła 68-95-99.7
5. Poznawanie zmiennych kategorialnych
6. Wartość modalna dla danych kategorialnych i dla danych numerycznych
7. Badanie zależności pomiędzy zmiennymi i wykres punktowy
8. Współczynnik korelacji
9. Tablice skrócone dwukierunkowe