

Zaawansowane metody i techniki analizy danych

OPIS PROJEKTU

Celem projektu jest zastosowanie nabytej wiedzy do samodzielnie przeprowadzonej analizy różnorodnych danych oraz wyciągania uogólnionych wniosków na ich podstawie.

Termin oddania projektu podaje prowadzący i jest on nieprzekraczalny!

Dane użyte w projekcie nie mogą być danymi używanymi podczas laboratoriów!

Część 2

1. Znajdź odpowiednie do analizy dane, w którym możemy wyróżnić przynajmniej 3 populacje np. cena akcji różnych spółek przez 3 dni, skuteczność przynajmniej 3 leków testowana na pewnej grupie osób itp. Wielkość każdej populacji jest dowolna, pamiętaj jednak, że wielkość populacji może determinować wybór właściwego testu statystycznego
2. Określ parametry rozkładu poszczególnych populacji: średnia, odchylenie standardowe, minimum, maksimum, mediana, pierwszy kwartył i trzeci kwartył.
3. Sprawdź równość badanej cechy we wszystkich populacjach odpowiednim testem. Pamiętaj o sprawdzeniu czy zmienne są powiązane, równości wariancji, rozkładu badanej cechy. Jeśli jest potrzeba wykonaj odpowiednią analizę post-hoc.
4. Sprawdź, w której populacji badana cecha ma największą średnią/medianę. Wykonaj kilka testów parami: populacja 1 z 2, 1 z 3, 2 z 3. Wybierz do tego zadania odpowiedni test, pamiętając o sprawdzeniu czy zmienne są powiązane, równości wariancji, rozkładu badanej cechy.
5. Sprawdź odpowiednim testem czy średnia/mediana w każdej populacji jest mniejsza niż $\text{średnia/mediana} + 10\% \cdot \text{średnia/mediana}$ wyznaczona w pkt. 2.
6. Wyznacz (jeśli to możliwe lub uzasadnij, dlaczego nie można) 95% i 99% przedział ufności dla średniej oraz wariancji dla każdej populacji
1. Uzyskane wyniki przedstaw w czytelnej formie tabel i/lub wykresów. Postaw właściwe hipotezy, przeanalizuj uzyskane wyniki, podejmij próbę ich wyjaśnienia, przedyskutuj wynikające z nich wnioski, uzasadnij wybór testów. Napisany kod opisz stosownymi komentarzami.