

PRACA DOMOWA NR 1 ANALIZA WIELOWYMIAROWA

Maciej Nasiński

TERMIN ODDANIA: 20.11.2021 23:59

Limit znaków ze spacjami: 10000

Przykład abstrakcyjny.

Właśnie zostaliście zatrudnieni w Ministerstwie Propagandy w roku 1984. Jako Towarzysz wyższego zaufania posiadacie dostęp do poufnych danych o wynagrodzeniach na obszarze całego kraju - plik `occup.csv` [68370 x 4]. Zwierzchnik jednak nie wie, że w głębi serca jesteś wojownikiem o prawdę. Twoim prawdziwym celem jest odnalezienie wszystkich społecznie istotnych odchyłeń od głównej linii narracji. danych. Proszę rozpocząć realizację zadania od zapoznania się ze strukturą zbioru danych, zwracając uwagę na charakterystykę zmiennych czy strukturę obserwacji brakujących (jeżeli takie występują).

Sformułowałeś kilka hipotez/pytań, poszukaj na nie odpowiedzi w danych:

1. (test t) Usłyszałem w telewizji że średnio kobiety zarabiają istotnie więcej od mężczyzn. Warto też sprawdzić czy na przestrzeni grup wariancja zarobków jest stała. Dodatkowo warto przeanalizować wielkość efektu.
2. (test t) Czujesz że dyrektorzy i politycy pomimo zapewnień nie wypłacają sobie średniego wynagrodzenia krajowego ('Przedstawiciele władz publicznych, wyżsi urzędnicy i kierownicy'). Pomocne będzie stworzenie dodatkowej zmiennej binarnej. Dodatkowo warto przeanalizować wielkość efektu.
3. (test χ^2) Pomimo zapewnień o absolutnej równości terytorialno zawodowej, masz wrażenie że zmienne `district` i `code NIE` są niezależne.
4. (Analiza Korespondencji) Chcesz sprawdzić czy powiązanie zmiennych `district` i `code`, gdy czujesz że kilka specyficznych grup zawodowych koncentruje się wokół kilku regionów. Jedną z nich są 'Przedstawiciele władz publicznych, wyżsi urzędnicy i kierownicy' w województwie ... oraz inne ...
5. (Wielokrotna Analiza korespondencji) Masz też wrażenie że w kraju wykształciły się specyficzne regiony biorąc pod uwagę zmienne `sex`, `code` i `district` :
 - W województwie Śląskim pracują głównie "Mężczyźni" w grupach zawodowych ...
 - 'Pracownicy wykonujący prace proste' to województwa ...
 - Województwo Mazowieckie to 3 grupy zawodowe 'Pracownicy usług i sprzedawcy', ...

Zadanie dodatkowe (Algebra liniowa - Optymalizacja kodu) Przeprowadzenie Analizy Korespondencji jest sprawne poprzez kod pythonowy, gdyż są to tajne kody podziemia. Proszę wyjaśnij dlaczego używane są metody `sklearn.utils.extmath.svd_flip` oraz

`scipy.sparse.diags` w metodzie `apply_svd` (klasa `CaBase`) znajdującej się w module `CAwne.CA` w folderze do zajęć 5.

Na koniec musisz przygotować podsumowanie ze wszystkimi wynikami swojego modelowania wraz z opisami/wnioskami. Dzięki tobie podziemie będzie miało materiały do swojej gazety.