Wrocław, 4.12.2022 r.

**PODSTAWY ANALIZY DANYCH – PRAKTYCZNE WARSZTATY**

**ZADANIA DO SAMODZIELNEGO PRZEĆWICZENIA, LISTA NR 4**

1. Na zajęciach poznaliśmy regresję krokową wsteczną (direction=”backward”) oraz regresję krokową pozwalającą zarówno na usuwaniem jak i dodawanie zmiennych do modelu (direction=”both”). Aby przeprowadzić tzw. regresję krokową postępującą (direction=„forward”), w przypadku której dodajemy zmienne do modelu, musimy zdefiniować najmniejszy, startowy model i wskazać zbiór zmiennych, które mogą znaleźć się w modelu.

Najpierw zbudujemy model startowy (nazwijmy go model0). Możemy przyjąć, że zawiera on tylko intercept. Przypomnijmy, że intercept w R oznaczamy jako „1”, zatem:

model0=lm(zmienna\_zależna~1, data=nazwa\_ramki\_danych).

Następnie wywołamy procedurę regresji krokowej postępującej:

stepAIC=(model0, direction=”forward”, scope=list(lower=model0, upper=~zmienna1+zmienna2+…+zmiennaN)) .

Proszę się pobawić plikiem state.x77, z którym pracowaliśmy na zajęciach i przeprowadzić regresję krokową postępującą dla różnych zestawów zmiennych. Proszę ograniczyć się do zmiennych o charakterze ilościowym, wprowadzanie zmiennych jakościowych do modelu omówimy w przyszłości.

UWAGA! U mnie powyższa instrukcja działa. Gdyby coś się zmieniło dla nowszej wersji programu i wystąpiły trudności, proszę o wiadomość.

2. Na MSTeams w zakładce „Pliki z danymi” umieściłam folder z czterema przykładami na regresję wieloraką. Proszę się pobawić oglądając dane, sprawdzając, czy zmienne „niezależne” rzeczywiście mogą być niezależne, budując różne modele, weryfikując je i porównując. Do wybranych przykładów wrócimy na zajęciach.

3. Dla dowolnych przykładów na model regresji liniowej prostej opisanych na liście do samodzielnego przećwiczenia nr 3 proszę sprawdzić testem Shapiro-Wilka, czy reszty surowe, reszty standaryzowane oraz reszty studentyzowane pochodzą z populacji charakteryzującej się rozkładem normalnym. Odpowiednia komenda ma postać: shapiro.test(*wektor\_danych*). Następnie proszę sprawdzić testem studenta dla jednej próby, czy wartość oczekiwana rozkładu tych reszt wynosi zero. Odpowiednia komenda ma postać: t.test(*wektor\_danych*, mu = *wartość\_którą\_sprawdzamy*). W naszym przypadku komenda ma postać: t.test(*wektor\_danych*, mu = *0*). Proszę zinterpretować wyniki dla poziomów istotności α = 0,05 oraz α = 0,01. Proszę pamiętać o sprawdzeniu, czy program wykonuje właściwy test😊, informacja na ten temat zostanie wyświetlona w pierwszej linijce komunikatu otrzymanego po wykonaniu testu.

4. Proszę zapoznać się z treścią pliku „PORÓWNANIE\_DWÓCH\_GRUP”, który znajdą Państwo w zakładce „Materiały do powtórzenia lub pogłębienia wiedzy teoretycznej” na MSTeams. Zajęcia z tego tematu rozpoczniemy od przykładów.