

Lista 11

Python: lista dodatkowa

1. Wczytaj dane `data_health`. Przygotuj je do analizy. Zmień typ zmiennych `Groups` i `Sex` na kategoriyczny, zrób to w jednym poleceniu używając `apply`.
2. Stwórz tabelę krzyżową, z której będzie można odczytać ilu pacjentów danej płci choruje na każdą z chorób. Następnie zmodyfikuj tabelę tak aby dla każdej z podgrup została wyliczona średnia `hsCRP`.
3. Stwórz model, który pozwoli na przewidzenie wieku na podstawie poziomu hemoglobiny i `GFR`. Następnie wyznacz reszty studentyzowane.
4. Usuń z danych większa dla których wartość bezwzględna reszty studentyzowanej jest większa od 2. Następnie stwórz po raz drugi model. Jak usunięcie reszt wpłynęło na dopasowanie modelu?
5. Sprawdź za pomocą testu i wykresu czy reszty w modelu pochodzą z rozkładu normalnego.