Efekt 3. Wstępna analiza danych

Damian Okoń, Mateusz Nowak, Robert Zamiar

# Opis zbioru danych, statystyki opisowe

W efekcie przekształceń danych otrzymano następujące zmienne:

Tabela 1. Zmienne z opisami

|  |  |  |
| --- | --- | --- |
| zmienna | wyjaśnienie | typ zmiennej |
| country | kraj | etykieta |
| GDP\_pc | Produkt krajowy brutto per capita w euro | zmienna numeryczna |
| high\_tech\_trade\_pc | Handel wysokimi technologiami per capita w euro | zmienna numeryczna |
| r&d\_gdp\_pct | Wydatki na badania i rozwój jako procent PKB | zmienna numeryczna |
| r&d\_bud\_pct | Wydatki na badania i rozwój jako procent budżetu | zmienna numeryczna |
| use\_cloud\_pct | Procent ludzi korzystających z chmury | zmienna numeryczna |
| weeknd\_work\_pct | Procent ludzi pracujących w weekendy (15-64) | zmienna numeryczna |
| emp\_deadline\_pct | Procent ludzi pracujących intensywnie i z deadlinem (15-64) | zmienna numeryczna |
| working\_pop\_pct | Procent populacji w wieku produkcyjnym (15-64) | zmienna numeryczna |
| phycisians\_per\_1000 | lekarze na 1000 mieszkańców | zmienna numeryczna |
| sea\_access | dostęp do morza | zmienna kategoryczna (binarna) |
| joined\_EU | dekada dołączenia do Unii Europejskiej | zmienna kategoryczna |
| is\_euro\_currency | Czy walutą jest euro? | zmienna kategoryczna (binarna) |
| nuclear\_electricity | Czy kraj posiada elektorwnie atomowe? | zmienna kategoryczna (binarna) |

Dane numeryczne jak i etykietę z nazwami krajów(nazwy w języku angielskim) jak i zmienne numeryczne zaczerpnięto z serwisu Eurostatu, natomiast zmienne kategoryczne zostały utworzone przez zespół na podstawie źródeł z serwisów geopolitycznych. Zmienne te mogą być wykazane w przyszłych, bardziej wnikliwych analizach – analizach głównych składowych czy analizach skupień.

Wszystkie zmienne numeryczne podane są w ujęciu procentowym, na 1000 mieszkańców, bądź per capita. Dzięki takiemu zabiegowi zniwelowany został wpływ wielkości poszczególnych krajów na badane wartości. Przekształcenia te pozwalają również zniwelować ilość obserwacji odstających w zbiorze.

Zastosowane formuły przy łączeniu zbiorów można podejrzeć w arkuszu dane

Jedyną wartością brakującą w zbiorze była ilość lekarzy dla Czech.

Brak ten w finalnym zbiorze uzupełniono obliczając średnią wszystkich wartości wskaźnika liczba lekarzy na 1000 mieszkańców

## Obliczanie zmienności

Przygotowane przez nas dane są przekrojowe – pochodzą z roku 2015. więc nie są szeregiem czasowym. Z powodu przyszłego wykorzystania danych do analizy regresji obliczono współczynnik zmienności będący ilorazem odchylenia standardowego i średniej danych zmiennych – wskaźnik ten obliczono zarówno dla zmiennych numerycznych jak i binarnych. Tabela z wynikami na następnej stronie:

|  |  |  |  |
| --- | --- | --- | --- |
| **Nazwa Zmiennej** | **Średnia** | **Odchylenie standardowe** | **Współczynnik zmienności** |
| GDP\_pc | 23677 | 13148,34391 | 55,53% |
| high\_tech\_trade\_pc | 863 | 567,7546457 | 65,81% |
| r&d\_gdp\_pct | 2 | 0,839291765 | 49,30% |
| r&d\_bud\_pct | 1 | 0,42803816 | 36,21% |
| use\_cloud\_pct | 25 | 9,006849246 | 36,09% |
| weeknd\_work\_pct | 24 | 7,983405706 | 32,68% |
| emp\_deadline\_pct | 43 | 8,675188759 | 20,07% |
| working\_pop\_pct | 65 | 1,762087543 | 2,70% |
| phycisians\_per\_1000 | 4 | 0,811189088 | 22,29% |
| sea\_access | 1 | 0,374165739 | 44,54% |
| is\_euro\_currency | 1 | 0,489897949 | 76,55% |
| nuclear\_electricity | 1 | 0,506622805 | 90,47% |

Jak możemy zauważyć – największą zmiennością charakteryzują się zmienne związane z handlem wysokimi technologiami jak i PKB per capita. Zmienne te posiadają dosyć wysokie odchylenie standardowe z powodu zróżnicowania tych wartości w poszczególnych krajach Unii Europejskiej. Zmienne binarne, o najwyższych współczynnikach zmienności z kolei charakteryzują dwa popularne dzisiaj trendy: energię atomową oraz posiadanie przez kraj waluty Euro – czyli czynniki, które zależą od przyjętej przez państwo polityki nuklearnej oraz pieniężnej.