Efekt 4, Redukcja wymiarów

Mateusz Nowak, Damian Okoń, Robert Zamiar

# Analiza składowych głównych – PCA

# Pytanie badawcze

Do jakiej minimalnej ilości wymiarów można zredukować predyktory, jednocześnie zachowując wysoki procent wyjaśnianej wariancji tych predyktorów? Co opisują te wymiary?

# Przygotowania do analizy

Jako pierwszą wykonano analizę składowych głównych. Najpierw jednak przeprowadzono testy, sprawdzające czy przeprowadzenie PCA jest w ogóle możliwe. Jako pierwszy zrobiono test sferyczności Bartletta. P-value tego testu wyszło bliskie zeru, co pozwala na przeprowadzenie PCA.

Obraz zawierający tekst

Opis wygenerowany automatycznie

Rysunek 1 Test sferyczności Bartletta

Następnie sprawdzono kryterium KMO, którego wynik ogólny (Overall MSA) większy od 0,5 również dopuszcza wykonanie PCA.

Obraz zawierający stół

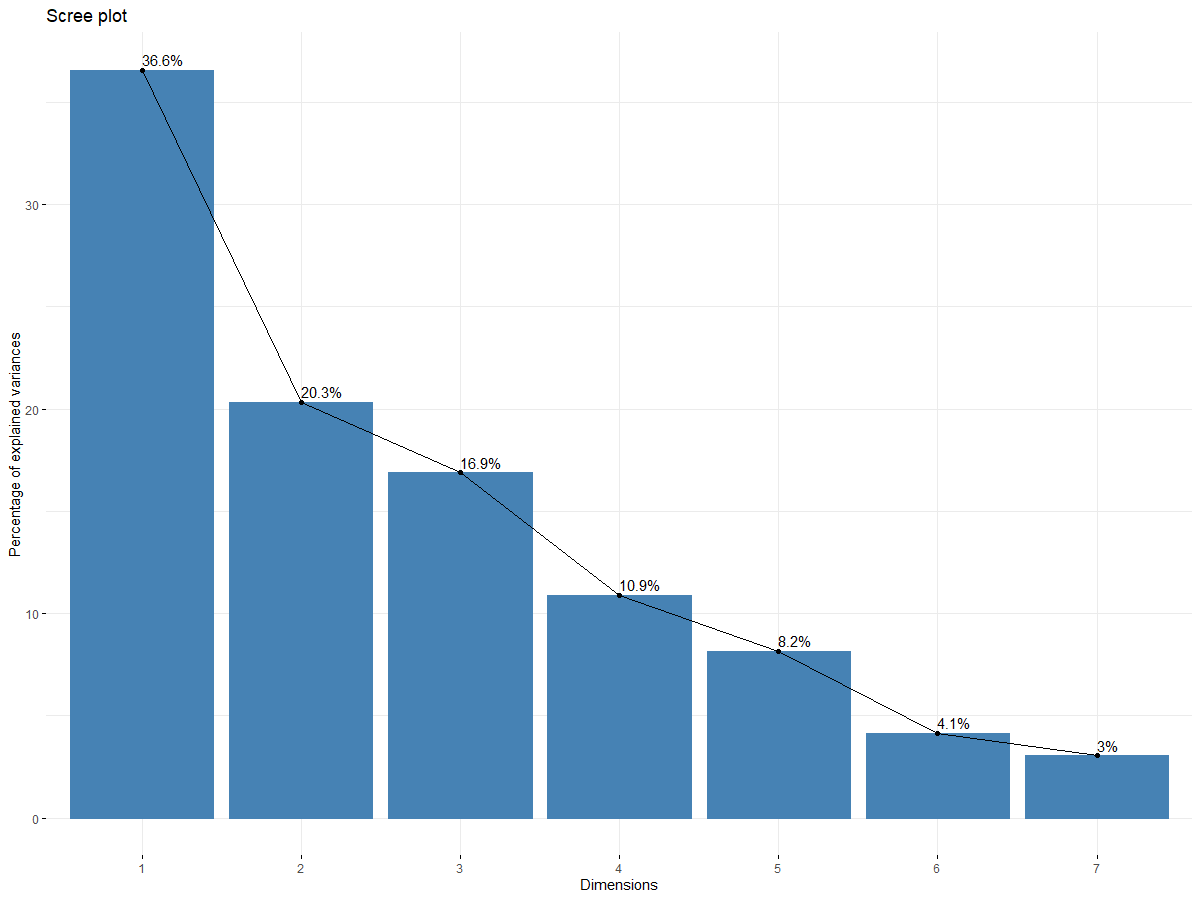
Opis wygenerowany automatycznie

Rysunek 2 Kryterium KMO

Na sam koniec przeprowadzono standaryzację zmiennych

# Właściwa analiza

Samą analizę PCA wykonano przy użyciu biblioteki FactoMiner w języku R. Najpierw przeprowadzono PCA dla liczby wymiarów równej liczbie predyktorów liczbowych, czyli dla 7 oraz stworzono wykres osypiska za pomocą funkcji fviz\_screeplot.



Rysunek 3 Wykres osypiska

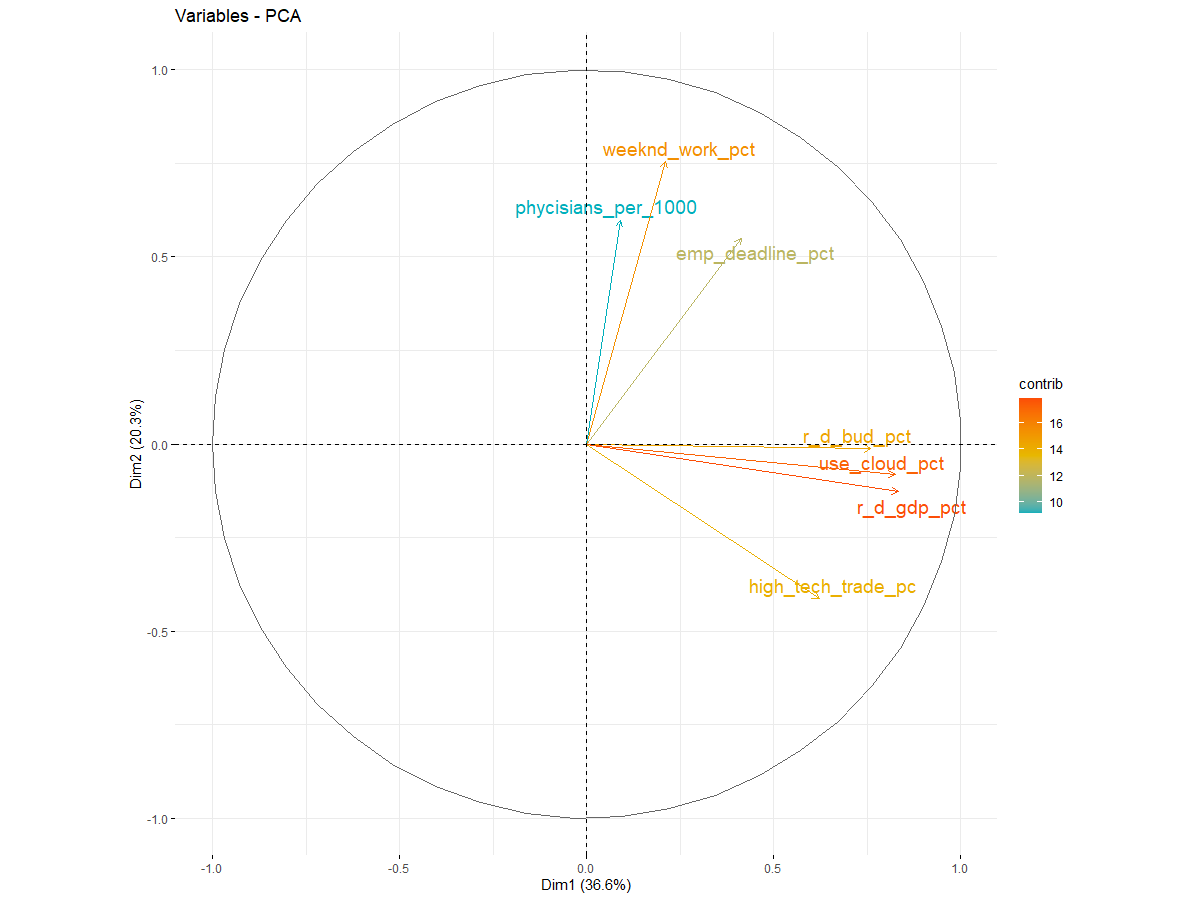
Spłaszczenie wykresu osypiska następuje po 2 wymiarach, które wyjaśniają 57% zmienności.

Obraz zawierający tekst

Opis wygenerowany automatycznie

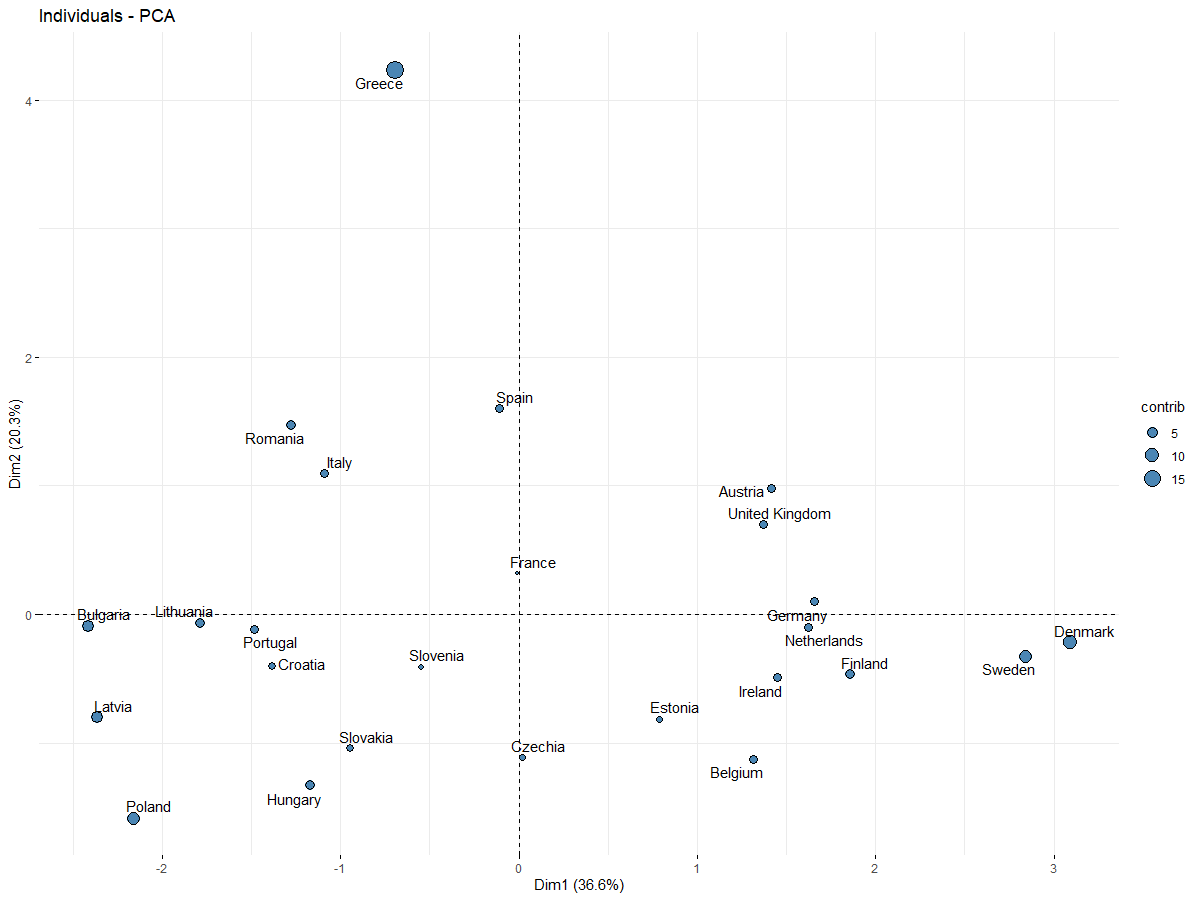
Rysunek 4 Badanie wymiarów

Pierwszy wymiar wyjaśnia handel wysokimi technologiami per capita w 62%, wydatki na badania i rozwój jako procent PKB i procent budżetu kolejno w 83% i 76% oraz procent ludzi korzystających z chmury w 82%. Drugi wymiar z kolei, pokrywa się z procentem ludzi pracujących w weekendy w 75% oraz z pracownikami mającymi deadline w 55%. Ostatnia zmienna, czyli liczba lekarzy na tysiąc mieszkańców w jest wyjaśniana przez drugi wymiar w 60%. Następnie wykonano PCA dla 2 wymiarów. Wyniki lepiej widoczne są na biplocie.



Rysunek 5 Analiza PCA – biplot

Zgodnie z wcześniejszymi obserwacjami cztery zmienne układają się zgodnie z wymiarem pierwszym (oś X), trzy zmienne natomiast układają się wzdłuż wymiaru drugiego (oś Y). Można zatem przyjąć, że wymiar pierwszy opisuje nakłady pieniężne na technologię oraz stopień rozwoju technologicznego danego państwa. Drugi wymiar natomiast opisuje kraje bardziej pod kątem społecznym. Ostatnim etapem wykonanej analizy PCA jest naniesienie danych na wykres.



Rysunek 6 Wykres obserwacji

Zgodnie z wykresem przedstawionym na Rysunku 6, kraje można podzielić na 4 grupy zgodnie z kwadrantami, w których się znajdują. Pierwszy kwadrant zawiera Austrię, Wielką Brytanię oraz Niemcy. Są to państwa wysoko rozwinięte oraz dość silnie inwestujące w nowe technologie. Charakteryzują się również wysoką wartością przynajmniej jednego czynnika opisującego wymiar drugi. W przypadku Austrii jest to liczba lekarzy na tysiąc mieszkańców, z kolei w przypadku Wielkiej Brytanii jest to liczba pracowników pracujących w ścisłych terminach. Mimo tego, że Niemcy należą do tego kwadrantu, w żadnej z jego składowych nie odstają mocno od średniej. W kolejnym kwadrancie znalazły się kraje skandynawskie, kraje byłego Beneluksu, Irlandia, Czechy oraz Estonia. Podobnie jak państwa z poprzedniego kwadrantu, kraje te również charakteryzują wyższym nakładem inwestycyjnym na nowe technologie niż pozostałe państwa. Wyróżnia je natomiast niższy wynik w wymiarze drugim. Przeciętnie mniej mieszkańców tych krajów pracuje w weekendy oraz mniejszy ich odsetek pracuje z deadline’ami. Ponadto mogą mieć oni dostęp do mniejszej ilości lekarzy w przeliczeniu na tysiąc mieszkańców. Następny kwadrant stanowią głównie kraje postsowieckie, w tym Polska. Państwa te przeznaczają znacznie mniej środków na rozwój nowoczesnych technologii od swoich poprzedników. Są jednak stosunkowo podobne do krajów z kwadrantu drugiego pod względem wymiaru społecznego. Polska w tej grupie jest państwem najbardziej odstającym pod względem niskiego wyniku w wymiarze drugim. Na taki rezultat wpływ ma najniższa w zestawieniu liczba lekarzy na tysiąc mieszkańców. Do ostatniego kwadrantu należą Rumunia oraz kraje śródziemnomorskie, takie jak Grecja, Włochy, Hiszpania i Francja. Państwa te również przeznaczają mniejszą część budżetu i PKB na rozwój. Wyróżnia je natomiast przeciętnie większa liczba lekarzy na tysiąc mieszkańców, większy odsetek obywateli pracujących w weekendy oraz z deadline’ami. Na tle tych krajów wyraźnie wyróżnia się Grecja. Państwo to przoduje w liczbie pracowników weekendowych, liczbie lekarzy na tysiąc mieszkańców, a także jest drugie w ilości pracowników, pracujących w ścisłych terminach.

# Odpowiedź na pytanie badawcze

Maksymalnie można zredukować liczbę predyktorów do dwóch, jednocześnie zachowując 57% zmienności. Pierwszy wymiar opisuje poziom nakładów pieniężnych na technologie oraz stopień rozwoju technologicznego danego państwa. Drugi wymiar z kolei opisuje aspekt społeczny tych krajów, głównie pod kątem sposobu pracy ich obywateli, a także dostępu do wykwalifikowanej opieki medycznej.