



Kierunek: Metody ilościowe w ekonomii i systemy informacyjne
Specjalność: Modelowanie ekonometryczne w biznesie

Imię i nazwisko autora:

Justyna Zbiegień

Nr albumu: **107961**

Wpływ pandemii COVID-19 na sytuację finansową studentów w Polsce

Praca magisterska

Pod kierunkiem naukowym

dr Izabeli Grabowskiej

Zakład Demografii

Instytut Statystyki i Demografii

Warszawa 2023

Spis treści

Wstęp.....	2
Rozdział 1 - Problematyka badań w świetle literatury przedmiotu i ogólnodostępnych danych	3
Opis sytuacji pandemicznej uczelni wyższych w Polsce	3
Zdrowie psychiczne i fizyczne w dobie pandemii.....	4
Sytuacja finansowa studentów.....	4
Rozdział 2 – Metody użyte do zbadania determinant sytuacji finansowej studentów	7
Analiza głównych składowych	7
Klastrowanie metodą k-średnich.....	10
Uporządkowany model logitowy (ordinal logistic regression).....	11
Rozdział 3 - Analiza i interpretacja wyników badań.....	17
Wnioski.....	39
Bibliografia	40
Spis rysunków.....	41
Spis tabel.....	41
Streszczenie.....	42

Wstęp

W roku 2020 na świecie wybuchła pandemia COVID-19, ostrej zakaźnej choroby układu oddechowego wywoływanej przez wirusa SARS-CoV-2. Mimo, że pierwszy przypadek zakażenia datuje się na koniec roku 2019 w prowincji Wuhan (Chiny) WHO (Światowa Organizacja Zdrowia, z ang. World Health Organization) ogłosiło stan ogólnoświatowej pandemii 11.03.2020 roku¹. W Polsce pierwszy przypadek zanotowano niecały tydzień wcześniej w Zielonej Górze, od tamtej pory do końca maja 2022 roku w Polsce zachorowało prawie 2,9000 ludzi, z czego ponad 73 tys. zmarło². W Polsce pojawiło się kilka skupisk zachorowań na przestrzeni trwania całej pandemii, najbardziej ucierpiały duże miasta, województwa Śląskie, Mazowieckie czy Małopolskie. Wprowadzono stan pandemiczny i tymczasowo zamknięto wszystkie szkoły. Wprowadzono obostrzenia dotyczące noszenia maseczek w miejscach publicznych, ograniczenia liczby osób w pomieszczeniach zamkniętych, zalecano pozostawanie w domach, następnie zamknięto wszystkie lokale gastronomiczne i niektóre ośrodki kultury. Pandemia COVID-19 znacząco odbiła się na ekonomii gospodarki, służbie zdrowia, ale też edukacji. Uczelnie wyższe początkowo zrezygnowały z zajęć, a następnie powróciły do nauczania w trybie zdalnym.

Pandemia COVID-19 znacząco się odbiła na ludziach młodych studiujących na uczelniach wyższych. Znacząco zmieniono tryb nauczania, wprowadzono wiele obostrzeń w miejscach edukacji, co też wpłynęło bezpośrednio na sytuację zawodową studentów. Celem niniejszej pracy jest zbadanie wpływu pandemii COVID-19 na sytuację finansową studentów w Polsce. Dane zaczerpnięto z badania przeprowadzonego przez SKN Statystyki uczelni SGH w dniu 22.05.2021 roku pt. „WPLYW PANDEMII NA PLANOWANIE KARIERY I OSZCZĘDNOŚCI STUDENTÓW”. Dane zostały zebrane przez instytut badawczy Norstat na podstawie odpowiedzi 800 studentów z całej Polski. Badanie ankietowane pierwotnie zawierało ponad 100 pytań dotyczących zarówno sytuacji finansowej jak i zawodowej studenta. Na potrzeby niniejszej pracy do badania wzięto jedynie odpowiedzi z 14 pytań dla wszystkich obserwacji (tj. 800 studentów). Analizy statystyczne oraz modele przeprowadzono w języku R za pomocą RStudio.

¹ Dane pochodzą ze strony internetowej <https://www.medicover.pl/o-zdrowiu/pandemia-koronawirusa-na-swiecie-i-w-polsce-kalendarium,7252,n,192> [dostęp 23.12.2022 11:42]

² Dane pochodzą ze strony internetowej <https://koronawirusunas.pl/> [dostęp 23.12.2022 11:58]

Rozdział 1 - Problematyka badań w świetle literatury przedmiotu i ogólnodostępnych danych

Opis sytuacji pandemicznej uczelni wyższych w Polsce

Od wiosny 2020 roku pandemia COVID-19 i środki polityczne nałożone w celu złagodzenia jej szkodliwych skutków (w tym dwa tzw. lockdowny gospodarcze) zmieniły i ukształtowały życie ludzi w Polsce, w tym system szkolnictwa wyższego. Zamknięcie uniwersytetów i przejście na nauczanie online (zdalne) znacząco wpłynęło na zdrowie psychiczne i fizyczne studentów oraz ich postępy w nauce i wyniki nauczania³. Na świecie zaobserwowano ogólny kryzys gospodarczy i związany z nim spadek zatrudnienia, co również wpłynęło na zatrudnienie studentów i ich sytuację finansową.

Przyjęte przez władze państwowe oraz uczelnie regulacje prawne miały wpływ na funkcjonowanie instytucji edukacyjnej w reżimie sanitarnym. Obok ważnych zagadnień jak zapewnienie jakości kształcenia, rozbudzania kreatywności u studentów, a także zaangażowania wśród pracowników, pojawiła się konieczność zachowania szczególnej ostrożności i czujności w związku z COVID-19. Na uczelniach wyższych wprowadzono reżim sanitarny. Na korytarzach pojawiły się płyny do dezynfekcji, zajęcia prowadzono w maseczkach, stosowano dłuższe przerwy, podczas których wietrzono sale. Ograniczono również liczbę studentów przebywających fizycznie w sali, stosując się do zaleceń WHO dotyczących odpowiedniej liczby osób na m² oraz zminimalizowano jakikolwiek kontakt fizyczny. Kolejnymi podjętymi środkami ostrożności było nauczanie zdalne. Studenci oraz prowadzący uczestniczyli w zajęciach w formie online. Wykorzystano do tego wszelkie dostępne platformy a uczelnie postawiły na rozwój technologii i cyfryzację. Wraz z wprowadzeniem takiej formy zajęć pojawiły się nowe wyzwania takie, jak prowadzenie zajęć praktycznych czy przeprowadzanie egzaminów w nowej formie. Zdania co do nauczania zdalnego są podzielone wśród studentów i wykładowców. Pojawiły się rozwiązania wcześniejszych problemów takich, jak większy dostęp do materiałów czy to w postaci wirtualnych prezentacji czy nagrań wykładów. Jednak największym minusem zajęć w formie online jest zaangażowanie studentów, technologia znacząco ogranicza możliwość większej interakcji na zajęciach. Największy problem pojawił się na uczelniach o profilu czysto technicznym jak i medycznym. Studentom odwołano praktyki w placówkach poza uczelnią, a nie rzadko również te odbywające się wewnątrz uczelni. Jednak forma nauczania zdalnego zbliżyła studentów do uczelni i wykładowców, gdy wygodnie się siedzi w domu słuchając wykładu w znanym środowisku znacząco to

³“COVID-19 and College Academic Performance: A Longitudinal Analysis” N. Rodríguez-Planas, 2021

poprawia stosunek do edukacji, jednak jednocześnie oddaliła studentów od wykorzystywania wiedzy w praktyce, co stało się nowym wyzwaniem w XXI wieku⁴.

Zdrowie psychiczne i fizyczne w dobie pandemii

Globalna pandemia stworzyła wiele niesprzyjających warunków do optymalnego funkcjonowania większości ludzi. Izolacja i dystans społeczny stały się głównymi przyczynami zmiany jakości życia w zakresie zdrowia psychicznego i fizycznego. Szczególnie ludzie młodzi, w tym dzieci, młodzież oraz osoby kontynuujące naukę na uczelniach wyższych, zostali narażeni na intensywną restrukturyzację działań społecznych. Nagłe przejście edukacji na poziom komunikacji pośredniej za pomocą różnych środków komunikacji wywołało silny kryzys. Wiązało się to z utratą normalnych relacji z rówieśnikami i drastyczną zmianą poczucia osamotnienia i dobrego samopoczucia. Komunikację twarzą w twarz zastąpiła rozmowa osób przed ekranem komputera lub urządzeniem mobilnym. Życie środowiska studenckiego przeniosło się w dużej mierze do przestrzeni wirtualnej, a wiele spraw zaczęło się załatwiać przez Internet. Jeśli dochodziło do bezpośredniej komunikacji odbywała się to przy zachowaniu reżimu sanitarnego. Jednak czas pandemii nie wiązał się z negatywnymi zmianami życiowymi dla wszystkich. W obliczu narastającego kryzysu spowodowanego wybuchem pandemii Wydział Zdrowia Psychicznego i Uzależnień WHO opublikował istotne informacje wspierające dobrostan psychiczny w trudnej sytuacji globalnej. Ponadto WHO zareagowała na problemy związane z wirusem COVID-19, tworząc materiał informacyjny dotyczący zdrowia psychicznego i aspektów psychospołecznych, uwzględniający skutki pandemii i ujawniający możliwe stany, takie jak strach, bezradność, poczucie bezradności, ale także możliwość wzmożonej agresji wobec innych czy pogorszenia funkcjonowania społecznego⁵.

Sytuacja finansowa studentów

Finalnym celem kształcenia na uczelniach wyższych jest podjęcie pracy zawodowej przez studentów w trakcie procesu nauki, a przede wszystkim zatrudnienie absolwentów w zawodzie zgodnym z profilem ukończonego kierunku studiów. 20 marca 2020 roku w Polsce wprowadzono stan epidemii, pozbawiając wielu studentów dochodów z pracy zarobkowej. Wprowadzenie zajęć online nie obyło się bez negatywnych odczuć ze strony uczniów, którzy twierdzili, że nie jest to dopracowany system, a niektórzy, którzy stracili

⁴ „Zarządzanie szkołą wyższą w okresie pandemii koronawirusa SARS-COV-2”. POSTĘPY TECHNIKI przetwórstwa spożywczego, str. 206-212, prof. dr hab. S. Dawidziuk i dr J. Boguski, 2021

⁵ „Studenci Uniwersytetu Warszawskiego w dobie pandemii COVID-19 – aspekty prawne, kryminologiczne i społeczne”, Biuletyn Kryminologiczny, str. 95-122, P. Kawka, W. Matejuk i inni, 2021

pracę i pozostali w izolacji domowej, znaleźli się w trudnej sytuacji z powodu opłat – dale musieli płacić za uczelnię. Wielu studentów wróciło do domów rodzinnych na czas nieokreślony, ale ci, którzy zostali w domach studenckich, często zgłaszali potrzebę obniżenia czynszu⁶. W odpowiedzi na te prośby rektorzy niektórych uczelni, np. Uniwersytetu Warszawskiego, obniżyli ceny pokoi w akademikach na czas określony. Mając na uwadze powyższe wiele naukowców dostrzegło nowy problem badawczy w opisywanym obszarze, a mianowicie: czy okres pandemii COVID-19 - związane z tym obostrzenia, które z jednej strony wpłynęły na funkcjonowanie jednostek gospodarczych i możliwość kontaktów międzyludzkich, a z drugiej strony dały większą dyspozycyjność czasową studentom uczącym się w trybie zdalnym z dowolnego miejsca – miał wpływ na aktywność zawodową studentów i potencjalne możliwości zatrudnienia po ukończeniu uczelni w branży zgodnej z profilem studiów. Przeprowadzono wiele badań o tej tematyce, a jednym z nich jest badanie Politechniki Warszawskiej pt. „Współczesne problemy gospodarcze – gospodarki w czasach kryzysu. Część II” w 2021 roku. W tej publikacji poświęcono cały rozdział o tytule „Ocena wpływu pandemii COVID-19 na możliwość zatrudnienia w opinii studentów w państwowej szkole wyższej w Koninie” autorstwa K. Kuschek, K. Wodnickiej oraz P. Szczypy. Według przeprowadzonych badań, autorzy stwierdzili znaczący wpływ pandemii na sytuację finansową ludzi młodych. Część studentów przestała pracować w niepełnym wymiarze godzin z powodu pandemii, zmniejszyła się liczba zgłoszeń do pracy dorywczej lub zawodzie odpowiadającym kierunkowi wykształcenia, a część studentów nawet straciła pracę z powodu pandemii. Przez niepewne czasy w czasie pandemii pracodawcy znacznie ograniczyli zatrudnianie studentów do prac nieformalnych⁷. Kolejnym badaniem sprawdzającym wpływ pandemii COVID-19 na sytuację finansową studentów jest „The Financial Situation of Students during the COVID-19 Pandemic” autorstwa D. H. Meier, S. L. Thomsen oraz J. Trunzer z 2021 roku. Praca ta zbadała sytuację finansową studentów w Niemczech w różnych momentach pandemii. Autorzy podzielili okres trwania koronawirusa pięć faz. Fazy te charakteryzowały się nasileniem obostrzeń i zakażeń na terenie Niemiec. Konkluzją badań jest ogromny wpływ pandemii i idącej za nią obostrzeniami na sytuację finansową studentów. Badanie wykazało, że średni spadek całkowitego dochodu studentów w Niemczech wyniósł ok. 19% podczas pierwszej blokady (od marca do maja 2020 r.). Dekompozycja źródeł dochodu, którą dokonano podczas badania, ujawniła, że spadek ten wynikał z negatywnych konsekwencji dla pracy studentów (zwolnienia, urlopy bezpłatne, skrócony czas

⁶ „Studenci Uniwersytetu Warszawskiego w dobie pandemii COVID-19 – aspekty prawne, kryminologiczne i społeczne”, Biuletyn Kryminologiczny, str. 95-122, P. Kawka, W. Matejuk i inni, 2021

⁷ „Współczesne problemy gospodarcze – gospodarki w czasach kryzysu. Część II” rozdział IV pt. „Ocena wpływu pandemii COVID-19 na możliwość zatrudnienia w opinii studentów w państwowej szkole wyższej w Koninie”, K. Kuschek i inni, 2021

pracy) z powodu nałożonych ograniczeń ekonomicznych. Dla około jednej trzeciej studentów dotkniętych ograniczeniami w pracy, konsekwencje te były daleko idące. Badanie wykazało, że dochód studencki z pracy był średnio o około 66% niższy niż w fazie przed pandemią. Studenci (częściowo) rekompensowali spadek dochodów z pracy poprzez zwiększenie finansowania pożyczek. Studenci spodziewali się, że ich sytuacja finansowa poprawi się po drugim lockdownie w zakresie dochodów ogółem (w porównaniu do fazy przed pandemią), głównie ze względu na zwiększenie oczekiwań dotyczących wsparcia ze strony rodziców oraz dochodów z pracy⁸. Zagadnieniem zmiany sytuacji materialnej studentów w trakcie pandemii COVID-19 zajęli się również badacze spoza Europy. Następne badanie przeprowadzono wśród studentów CUNY – Uniwersytetu Nowojorskiego, USA. Nosi ono tytuł „The Impact of the COVID-19 Pandemic on College Students’ Health and Financial Stability in New York City: Findings from a Population-Based Sample of City University of New York (CUNY) Students” autorstwa Heidi E. Jones i innych. Artykuł ten opublikowano w czasopiśmie *Urban Health* na początku 2021 roku. Amerykańscy naukowcy w porównaniu do niemieckiego badania przyjrzeni się bliżej wydatkom studentów, zamiast ich dochodom. Według badania większość uczniów (81,1%) podała, że oni i/lub ktoś inny w ich gospodarstwie domowym utracili dochody w wyniku pandemii. Prawie połowa (47,8%) podała, że ich tygodniowe wydatki domowe wzrosły, natomiast 21,2% zgłosiło spadek wydatków. Największym wzrostem wydatków domowych było zwiększenie kosztów środków czystości (72,9%) oraz kosztów żywności (71,4%). Połowa studentów (49,8%) podała, że jest bardzo lub nieco zaniepokojona utratą mieszkania w wyniku pandemii, a około połowa zgłosiła, że doświadcza niektórych form braku bezpieczeństwa żywnościowego w ciągu 2 tygodni poprzedzających⁹.

Patrząc na powyższe przykłady badań można przyjąć hipotezę, że pandemia koronawirusa znacząco wpłynęła na sytuację finansową studentów w Polsce – zarówno na wydatki jak i dochody. Brak możliwości zatrudnienia, utrata pracy, zwiększenie kosztów życia i podstawowych artykułów domowych, zaciąganie zobowiązań finansowych takich, jak pożyczki ekspresowe czy rezygnacja z wynajmu w miejscu uczelni, są jedynie kilkom z wielu czynników, które znacząco wpłynęły na zmianę sytuacji materialnej młodych ludzi z uczelni wyższych.

⁸ „The Financial Situation of Students during the COVID-19 Pandemic”, Insitute of Labor Economics, D. H. Meier, S. L. Thomsen, J. Trunzer, 2022

⁹ “The Impact of the COVID-19 Pandemic on College Students’ Health and Financial Stability in New York City: Findings from a Population-Based Sample of City University of New York (CUNY) Students”, *Urban Health* str. 187-196, Heidi E. Jones i inni, 2021

Rozdział 2 – Metody użyte do zbadania determinant sytuacji finansowej studentów

W celu zbadania wpływu pandemii COVID-19 na sytuację finansową studentów przeprowadzono szereg działań opartych na danych zapożyczonych z badania pt. „WPLYW PANDEMII NA PLANOWANIE KARIERY I OSZCZĘDNOŚCI STUDENTÓW” z 2022 roku. Pierwszym krokiem jest przeprowadzenie analizy głównych składowych i stworzenia linowych kombinacji pierwotnych zmiennych. Z tak przeprowadzonej analizy głównych składowych każdej obserwacji zostały przypisane indywidualne koordynaty. Na ich podstawie przeprowadzono klastrowanie metodą k-średnich i przypisano każdą obserwację (studenta) do danej grupy. Następnie stworzono uporządkowany model logitowy, gdzie zmiennymi niezależnymi są nie tylko zmienne uzyskane w badaniu, ale też zmienna oznaczająca przynależność studenta do danej grupy.

Analiza głównych składowych

Analiza głównych składowych (z ang. Principal Component Analysis czyli PCA) to jedna z metoda analizy czynnikowej i zarazem najbardziej popularny algorytm do redukcji liczby wymiarów. Polega na stworzeniu nowych zmiennych, tzw. głównych składowych, będących kombinacją liniową zmiennych pierwotnych wziętych do analizy. Nowe zmienne tworzy się w taki sposób, aby zachować pierwotnych charakter czynników oraz aby zbiór uczący zachował jak największą wartość wariancji. Idea tworzenia kolejnych składowych polega na tym, że kolejne składowe nie są skorelowane ze sobą oraz mają na celu zmaksymalizować zmienność, która nie została wyjaśniona przez pierwotne zmienne. Ważnym jest, że liczba głównych składowych jest równa lub niższa liczbie zmiennych pierwotnych (w ten sposób dochodzi do redukcji wymiarów).

Niech p oznacza liczbę zmiennych pierwotnych. Zgodnie z powyższymi założeniami główne składowe można wyrazić następującym wzorem¹⁰:

$$\begin{aligned} Z_1 &= a_{11}X_1 + a_{12}X_{12} + \dots + a_{1p}X_p \\ &\vdots \\ Z_m &= a_{m1}X_1 + a_{m2}X_{12} + \dots + a_{mp}X_p \end{aligned} \tag{1}$$

¹⁰ „Analiza porównawcza efektywności metod redukcji zmiennych - analiza składowych głównych i analiza czynnikowa” A. Czopek, 2013

gdzie:

m – liczba głównych składowych

p – liczbę pierwotnych zmiennych

a – współczynniki wyznaczone przez pierwotne zmienne

Aby zachować możliwie jak najwięcej zmienności danych, należy wyznaczyć składową główną o maksymalnej wariancji. Stąd wynika, że celem analizy składowych głównych jest wyznaczenie takich wartości współczynników a , aby wariancja Z była jak największa. W języku geometrii polega to na poszukiwaniu linii prostej, która jest najlepiej dopasowana do grupy punktów w przestrzeni. Po przekształceniach matematycznych, postawiony problem sprowadza się do rozwiązania układu p równań, który w zapisie macierzowym przyjmuje następującą postać:

$$(S - \lambda \cdot I)a_{ip} = 0 \quad (2)$$

gdzie:

S – macierz kowariancji zmiennych X

λ – wartość własna macierzy S

I – macierz identycznościowa

i – i -ta główna składowa

Niezerowe rozwiązanie powyższego równania istnieje wówczas, gdy macierz ta nie jest macierzą odwracalną, czyli w sytuacji, gdy jej wyznacznik jest równy zero. Można również udowodnić, że wariancja składowej głównej jest równa wartości własnej λ (czyli $Var(Z) = \lambda$). Zatem problem sprowadza się do znalezienia największej wartości własnej macierzy kowariancji. Współczynniki odpowiadające jej wektora własnego to poszukiwane współczynniki a głównej składowej. W rzeczywistości wyznacza się kilka głównych składowych, gdyż jedna nie jest w stanie wytłumaczyć całej zmienności zmiennych pierwotnych, dlatego każda kolejna składowa maksymalizuje zmienność, która nie została wyjaśniona przez poprzednią składową. Suma zmienności wszystkich składowych głównych powinna się sumować do wartości zmienności zmiennych pierwotnych, a zmienność poszczególnych zmiennych składowej można wyznaczyć ze wzoru:

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p} * 100\% \quad (3)$$

Analizę głównych składowych należy przeprowadzić dla danych mierzalnych osadzonych co najmniej na skali przedziałowej, gdzie występuje korelacja między zmiennymi. Aby sprawdzić, czy przeprowadzanie PCA dla wybranych zmiennych ma sens należy przeprowadzić test Bartletta, czyli test sprawdzenia równości wariancji dla zmiennych. Hipotezy tego testu oraz statystyka są następujące¹¹:

H0: wszystkie współczynniki korelacji są równe zero (macierz R obliczona z korelacji Pearsona jest równa macierzy jednostkowej)

H1: występują znaczące korelacje między zmiennymi

$$U = -(n - 1 - \frac{2p + 5}{6}) \sum_{i=1}^p \ln \lambda_i \quad (4)$$

gdzie:

n – liczba obserwacji

λ_i – i-ta wartość własna

Statystyka tego testu ma rozkład chi-kwadrat o $p(p-1)/2$ stopniach swobody. Pożądane jest odrzucenie hipotezy zerowej i przyjęcie alternatywnej, aby otrzymać ten wynik p-value musi być niższe niż przyjęta wartość krytyczna. Idąc dalej - założenie o normalności nie jest konieczne przy analizie składowych głównych, o ile opisujemy powiązania w dużym zbiorze danych. Jednak gdy włączamy wnioskowanie statystyczne do określenia istotności składowych, konieczne jest przyjęcie założenia o wielowymiarowej normalności. Aby wyniki były bardziej wiarygodne, musimy dysponować próbą o odpowiedniej liczebności. Absolutnym minimum jest próba o liczebności 50 obserwacji, jako odpowiednie przyjmuje się grupy o 100

¹¹ „Kryterialna ocena jakości działania systemu transportu miejskiego” Ł. Muślewski, Uniwersytet Technologiczno-Przyrodniczy w Bydgoszczy, 2015

obserwacjach. Również przed rozpoczęciem analizy należy wykryć przypadki odstające i je usunąć ze zbioru danych.

Po przeprowadzaniu analizy głównych składowych należy sprawdzić tzw. kryterium Kaisera-Guttmana. Jest to technika statystyczna używana do określenia, czy analiza czynnikowa jest odpowiednia dla danego zestawu danych. Kryterium to opiera się na założeniu, że struktura czynnikowa danych jest jasna i możliwa do zinterpretowania, i służy do określenia liczby czynników, które powinny być zachowane w analizie. Aby zastosować kryterium Kaisera-Guttmana, należy najpierw obliczyć wartości własne dla każdego z czynników w danych (wartość własna jest miarą wariancji w danych wyjaśnionej przez każdą główną składową). Kryterium Kaisera-Guttmana sugeruje, że należy zachować wszystkie czynniki o wartościach własnych większych niż 1.0, ponieważ te składowe są uważane za znaczące i wnoszą istotny wkład w wariancję danych. Główne składowe o wartościach własnych mniejszych niż 1.0 są uważane za mniej istotne i mogą zostać usunięte z analizy.

Klastrowanie metodą k-średnich

Klastrowanie jest techniką iteracyjną mającą na celu stworzenie grup (klastrow) obiektów sobie podobnych ze względu na podane cechy. Podzielenie obserwacji na grupy może służyć do wykrycia anomalii, dalszego prognozowania czy wykrycia relacji między grupami. Klastry można wydzielić za pomocą różnych metod. Metody te można podzielić na dwie grupy: algorytmy hierarchiczne i niehierarchiczne. Te dwie grupy algorytmów różnią się głównie tym, że w algorytmach hierarchicznych nie trzeba z góry podawać liczby klastrow, które chcemy stworzyć, natomiast algorytmom hierarchicznym należy z góry ustalić liczbę grup, na które chcemy podzielić obserwacje. Metoda k-średnich należy do grupy algorytmów niehierarchicznych. Metoda k-średnich tworzy k centroidów, czyli punktów, które są reprezentatywne dla każdego klastra. Algorytm k-średnich przypisuje każdy przychodzący punkt danych do jednego z klastrow, według przyjętej metody liczenia odległości obiektów od centroidów. Następnie polega na przenoszeniu obiektów z grupy do grupy tak długo, aż zostaną zoptymalizowane zmienności wewnątrz klastrow oraz pomiędzy nimi. Podobieństwo obserwacji wewnątrz klastra powinno być jak największe, natomiast grupy powinny się między sobą jak najbardziej różnić. Prościej mówiąc, aby przeprowadzić metodę k-średnich należy wykonać następujące kroki:

- 1. Ustalenie liczby klastrow**
- 2. Ustalanie wstępnych środków klastrow – centroidów**

Środki skupień tak zwane centroidy możemy dobrać na kilka sposobów: losowy wybór k obserwacji, wybór k pierwszych obserwacji, dobór w taki sposób, aby zmaksymalizować odległości skupień. Jedną z najczęściej stosowanych metod jest kilkakrotne uruchomienie algorytmu i wybór najlepszego modelu, gdy wstępnie środki skupień były wybierane losowo.

3. Obliczenie odległości obiektów od centroidów

Najczęściej stosowaną odległością jest odległość euklidesowa. Stosuje się również kwadrat tej odległości czy też odległość Czebyszewa.

4. Przypisanie obiektów do skupień

Dla danej obserwacji porównuje się odległości od wszystkich centroidów i przypisuje ją do grupy, do którego środka ma najbliżej.

5. Ustalenie nowych centroidów

Najczęściej nowym środkiem skupienia jest punkt, którego współrzędne są średnią arytmetyczną współrzędnych punktów należących do danego skupienia.

6. Wykonanie kroków 3,4,5 do czasu, aż warunek zatrzymania zostanie spełniony

Najczęściej stosowanym warunkiem stopu jest ilość iteracji zadana na początku lub brak przesunięć obiektów pomiędzy skupieniami.

Uporządkowany model logitowy (ordinal logistic regression)

Uporządkowane modele ekonometryczne często występują gdy zostały zaobserwowane pewne zjawiska niemierzalne. Najprostszym przykładem są badania ankietowe, gdzie respondentów pyta się o zdania subiektywne (np. preferencje, poglądy, poziom zadowolenia), niemierzalny ranking (np. ze względu na markę samochodu), niewygodne pytania (np. dochód) czy klasę wielkości (zmienne przedziałowe). W tych sytuacjach stosuje się zmienną porządkową y , która jest ograniczonym zapisem pewnej nieobserwowalnej zmiennej ciągłej y^* . y^* jest liniową funkcją zmiennych objaśniających zapisanych w wektorze x oraz nieznanymi parametrami w wektorze β ¹².

¹² Wzory pochodzą z książki „Mikroekonometria. Modele i analizy danych indywidualnych” (wyd. wydanie II rozszerzone) str. 134-139, redakcja naukowa Gruszczyński, 2012

$$\begin{aligned} y_i^* &= x_i' \beta + u_i \\ u &\sim iid \end{aligned} \quad (5)$$

gdzie:

i – i-ty obiekt

u - składniki losowy opisany funkcją gęstości $f(u)$ o zerowej wartości oczekiwanej i stałej wariancji

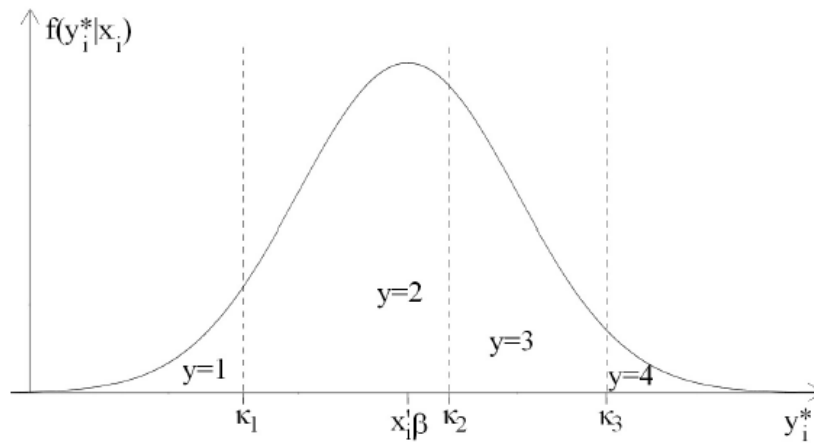
Uporządkowany model w porównaniu do większości modeli ekonometrycznych nie zawiera stałej. Zmienna y^* jest nieobserwowalna, ale można zaobserwować różne wartości zmiennej porządkowej y , która jest jej odzwierciedleniem. Zmienna obserwowalna y przyjmuje J wartości, które oznaczają się kolejnymi liczbami naturalnymi $j = 1, 2, 3, \dots, J$, w taki sposób, że im wyższa liczba oznacza daną kategorię, tym wyższej wartości y^* odpowiada dana kategoria (lub odwrotnie – ważne, aby odwzorowanie było monotoniczne). Przekształcenie zmiennej nieobserwowalnej y^* na zmienną obserwowalną y odbywa się w następujący sposób:

1. Należy podzielić zakres wartości zmiennej y^* na J przedziałów
2. Wyznaczyć $J+1$ wartości, które będą granicami (progami) – wektor tych wartości to κ a jego elementy to punkty odcięcia
3. Za κ_0 przyjąć $-\infty$, a za κ_J przyjąć $+\infty$

Zmienna obserwowalna y przyjmuje wartość j , gdy zmienna nieobserwowalna y^* przyjmuje wartość z j -tego przedziału, czyli spomiędzy κ_{j-1} a κ_j .

$$y_i = j \leftrightarrow \kappa_{j-1} < y_i^* \leq \kappa_j \quad (6)$$

Powyższy formalny zapis można przedstawić za pomocą wykresu (Rys.1):

Rysunek 1: Zmienna nieobserwowalna y^* a zmienna porządkowa y 

Źródło: „Mikroekonometria. Modele i metody analizy danych indywidualnych” Gruszczyński, 2012

Zatem formalna postać modelu uporządkowanego bez nieobserwowalnej y^* wygląda następująco:

$$\begin{aligned}
 y_i = 1 &\leftrightarrow -\infty < u_i^* \leq \kappa_1 - x_i' \beta \\
 y_i = 2 &\leftrightarrow \kappa_1 - x_i' \beta < u_i^* \leq \kappa_2 - x_i' \beta \\
 &\vdots \\
 y_i = J &\leftrightarrow \kappa_{J-1} - x_i' \beta < u_i^* \leq +\infty
 \end{aligned} \tag{7}$$

Parametry uporządkowanego modelu ekonometrycznego szacuje się za pomocą metody największej wiarygodności, jednak zanim się przystąpi do tej procedury należy znać wiarygodność pojedynczej obserwacji, określoną jako funkcja nieznanych parametrów i zmiennych objaśniających. Niech p_{ij} oznacza prawdopodobieństwo, że dla i -tego obiektu zmienna y przyjmie wartość j , a d_{ij} to wskaźnik oznaczający, że w próbie zmienna y_i przyjęła faktycznie wartość j , czyli:

$$\begin{aligned}
 p_{ij} &= P(y_i = j) \\
 d_{ij} &= \begin{cases} 1, & y_i = j \\ 0, & \text{w p.p.} \end{cases}
 \end{aligned} \tag{8}$$

Wtedy wiarygodność danej obserwacji $L_i(y_i | x_i)$ to prawdopodobieństwo przyjęcia przez zmienną objaśnianą w jej przypadku dokładnie tej wartości, która faktycznie wystąpiła w próbie. Czyli $L_i(y_i | x_i) = p_{ij}$, gdy $y_i = j$. Inaczej zapisując:

$$L_i(y_i|x_i) = p_{i1}^{d_{i1}} * p_{i2}^{d_{i2}} * ... * p_{ij}^{d_{ij}} = \prod_{j=1}^J p_{ij}^{d_{ij}} \quad (9)$$

Wskaźniki d_{ij} można wyliczyć z próby, ale za to prawdopodobieństwa p_{ij} nie znane. Należy zatem wskazać formułę na p_{ij} zależną od nieznanych parametrów i wartości zmiennych objaśniających. Wykorzystując specyfikę modelu, wzór na prawdopodobieństwo oraz wz. 8 można zapisać wzór na p_{ij} jako:

$$p_{ij} = F(\kappa_j - x_i' \beta) - F(\kappa_{j-1} - x_i' \beta) \quad (9)$$

gdzie F jest dystrybuantą składnika losowego u odpowiadającą funkcji gęstości f . Mając wybrany rozkład składnika losowego $f(u)$ i korzystając z powyższych wzorów, wiarygodność każdej obserwacji można przedstawić jako funkcję zmiennych objaśniających i nieznanych parametrów. Zakładając niezależność obserwacji można funkcję wiarygodności przedstawić jako iloczyn pojedynczych wiarygodności, a następnie zlogarytmizować, wtedy wzór na $\ln L$ jest równy:

$$L(y|x; \beta, \kappa_1, \kappa_2, \dots, \kappa_{J-1}) = \sum_{i=1}^n \sum_{j=1}^J d_{ij} \ln [F(\kappa_j - x_i' \beta) - F(\kappa_{j-1} - x_i' \beta)] \quad (10)$$

Z tak otrzymanego wzoru należy przeprowadzić metodę największej wiarygodności i oszacować parametry.

Aby oszacować model uporządkowany, trzeba wybrać rozkład składnika losowego $f(u)$, którego dystrybuenta występuje we wzorze na wiarygodność próby. Dwa typowe wybory to rozkład normalny i rozkład logistyczny, które prowadzą odpowiednio do uporządkowanego modelu probitowego i uporządkowanego modelu logitowego. Dystrybuenta rozkładu logistycznego wyraża się wzorem:

$$F(u) = \frac{\exp(u)}{1 + \exp(u)} \quad (11)$$

Dla uporządkowanego modelu logitowego wzór (9) ma zatem postać:

$$p_{ij} = \frac{\exp(\kappa_j - x_i' \beta)}{1 - \exp(\kappa_j - x_i' \beta)} - \frac{\exp(\kappa_{j-1} - x_i' \beta)}{1 - \exp(\kappa_{j-1} - x_i' \beta)} \quad (12)$$

Tak oszacowane parametry są nieinterpretowalne. W celu ich zinterpretowania należy obliczyć iloraz szans. Najpierw należy uzyskać proporcjonalne szanse. Założenie proporcjonalnych szans nie polega na tym, że szanse są takie same, ale że współczynniki szans są takie same dla wszystkich kategorii. Te współczynniki szans można uzyskać przez wykładnicze współczynniki, ale interpretacja jest nieco nieoczekiwana. Należy pamiętać, że współczynnik reprezentuje jednostkową zmianę logarytmicznych szans dla zmiennej zależnej y przy zmianie x_i o jedną jednostkę¹³:

$$\text{logit}(P(Y \leq j | x_i = 1)) - \text{logit}(P(Y \leq j | x_i = 0)) = \beta_i \quad (13)$$

Ponieważ wykładnik jest funkcją odwrotną do logarytmu, można skorzystać z własności $\log\left(\frac{b}{a}\right) = \log b - \log a$, wtedy:

$$\frac{\frac{P(Y \leq j | x_i = 1)}{P(Y > j | x_i = 1)}}{\frac{P(Y \leq j | x_i = 0)}{P(Y > j | x_i = 0)}} = \exp(\beta_i) \quad (14)$$

Aby uprościć powyższe równanie można założyć, że:

$$\frac{P(Y \leq j | x_i = 1)}{P(Y > j | x_i = 1)} = p_1 / (1 - p_1) \quad \frac{P(Y \leq j | x_i = 0)}{P(Y > j | x_i = 0)} = p_0 / (1 - p_0) \quad (15)$$

Wtedy iloraz szans można zapisać w postaci:

$$\frac{p_1 / (1 - p_1)}{p_0 / (1 - p_0)} = \exp(\beta_i) \quad (16)$$

¹³ „Analysis of categorical data with R” Bilder, C. R., & Loughin, T. M., 2014

Tak uzyskane ilorazy szans dla każdej zmiennej niezależnej x można interpretować i następnie sprawdzić dopasowanie modelu. W tym celu można użyć kryteriów informacyjnych Akaike oraz Bayesa-Schwarza. Kryterium informacyjne Akaike AIC (z ang. Akaike Information Criterion) oraz Bayesowskie kryterium informacyjne Schwartz BIC (z ang. Bayes Information Criterion) są ze sobą ściśle powiązane i opierają się na funkcji prawdopodobieństwa – do modelu można dodawać zmienne powiększając prawdopodobieństwo dopasowania, jednak może to prowadzić do jego sztucznego podwyższania, aby do tego nie dopuścić kontroluje się właśnie AIC oraz BIC. W modelu najbardziej jest pożądane, aby oba kryteria były jak najmniejsze. Wzory dla oby kryteriów wyglądają następująco¹⁴:

$$AIC = \log\left(\frac{e'e}{2}\right) + \frac{2K}{N} \quad (17)$$

$$BIC = \log\left(\frac{e'e}{2}\right) + \frac{K \log(N)}{N} \quad (18)$$

gdzie:

$e'e$ - suma kwadratów reszt

K – liczba zmiennych

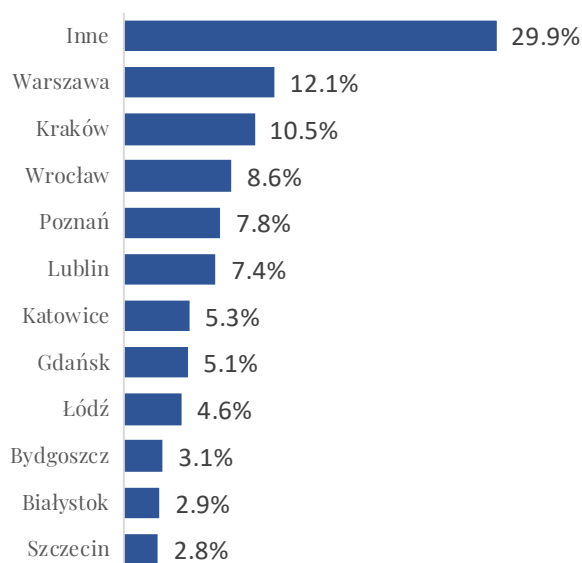
N – liczba obserwacji

Kolejnym sprawdzeniem dopasowania modelu do danych jest wyznaczenie pierwiastek błędu średniokwadratowego – RMSE (ang. Root mean square error). Interpretuje się go podobnie jak błąd średni absolutny błąd, ale jest bardziej czuły na wartości skrajne. Przedstawia się go następującym wzorem:

$$RMSE = \sqrt{\frac{1}{s} \sum_{\tau=1}^s (y_{\tau} - y_{\tau}^p)} \quad (19)$$

¹⁴ J. Mycielski (2010) „Skrypt. Rozdział 10 Metodologia testowania hipotez”, UW, s. 6

Rysunek 3: Wykres przedstawiający miasta studiowania respondentów w %



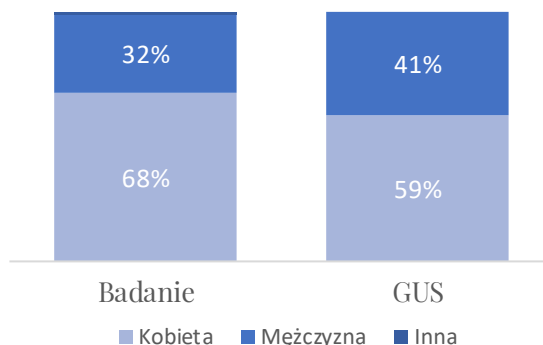
Źródło: Badanie „Wpływ pandemii na planowanie kariery i oszczędności studentów”, SKN Statystyki SGH, 2022

Powyższe rysunki (Rys.2 i Rys.3) przedstawiają rozkład geograficzny studentów ze względu na miasto, w którym obecnie studiują podaną w procentach. W badanej próbie największe odchyły względem liczby ludności osób w wieku 20-29 występują w województwie podkarpackim (za dużo o 3,1%) oraz w województwie wielkopolskim (za mało o 1,9%). Pozostałe województwa odchylają się od populacji młodych o mniej niż 1,5%¹⁵, także można stwierdzić że próbka jest reprezentatywną pod względem geograficznym. Natomiast struktura płci wśród respondentów znacznie bardziej się różni od rzeczywistej, poniższy wykres obrazuje to porównanie¹⁶:

¹⁵ Dane porównane z Banku Danych Lokalnych GUS z roku 2021

¹⁶ Dane pochodzą ze sprawozdań o studiach wyższych składanych przez uczelnie na potrzeby statystyki publicznej dla GUS (na dzień 31.12.2021). Sprawozdania są generowane w systemie POL-on i opierają się na metodyce opracowanej przez GUS we współpracy z Ośrodkiem Przetwarzania Informacji – Państwowym Instytutem Badawczym.

Rysunek 4: Porównanie struktury respondentów z GUS na rok 2021 według płci



Źródło: Opracowanie własne

Jak widać na powyższym rysunku (Rys.4) w badaniu wzięło udział niemal o 10% mniej mężczyzn niż według statystyk GUS. Około 80% studentów zaznaczyło, że studiuje stacjonarnie, na studiach pierwszego stopnia 45%, inżynierskich 20%, a na studiach drugiego stopnia 33%. Największym obszarem studiów, z którego przebadano studentów jest biznes, administracja i prawo (około 17% respondentów). Badanym przedstawiono ankietę w postaci CAWI składającą się z kilkudziesięciu pytań, jednak na potrzeby tej pracy do dalszych analiz przyjęto jedynie 13 (poza metryczką, tj. płeć, stopień studiów, realizowany semestr, wielkość miejscowości pochodzenia). Poniższa tabela przedstawia te pytania, przekodowaną nazwę zmiennej, typ zmiennej i przyjmowane dla niej wartości:

Tabela 1: Opis danych przyjętych z badania

Nr pytania	Treść pytania	Zmienna	Typ zmiennej	Wartości
1	Czy obawiasz się o swoją przyszłość finansową w związku z pandemią?	obawa	Kategoryczna	1. Zdecydowanie obawiam się. 2. Raczej się obawiam. 3. Raczej się nie obawiam 4. Zdecydowanie się nie obawiam 5. Trudno powiedzieć

2	Jak określasz swoją aktualną sytuację finansową w porównaniu z okresem sprzed pandemii?	synt_fin_do_p	Kategoryczna	1. Moja sytuacja finansowa polepszyła się 2. Moja sytuacja finansowa pogorszyła się 3. Moja sytuacja finansowa się nie zmieniła
3	Porównując z okresem sprzed pandemii, jak zmieniły się Twoje wydatki na poszczególne elementy budżetu? – Artykuły spożywcze	art_spoz	Kategoryczna	1. Wydaję na to znacznie mniejszą sumą 2. Wydaję na to raczej mniejszą sumą 3. Wydaję na to tyle samo 4. Wydaję na to raczej większą sumę 5. Wydaję na znacznie to więcej 6. To nie jest częścią mojego budżetu
4	Ile wynosi Twój średni miesięczny dochód netto w PLN?	dochod	Ciągła	Od 0
5	Jaki jest Twój miesięczny koszt utrzymania w PLN? (np. czynsz i rachunki)	koszt	Ciągła	Od 0

6	Jak często kontrolujesz stan swojego konta bankowego oraz zasobu gotówki w porównaniu do okresu sprzed pandemii?	kontrola	Kategoryczna	<ol style="list-style-type: none"> 1. Zdecydowanie częściej niż przed pandemią 2. Raczej częściej niż przed pandemią 3. Tak samo często jak przed pandemią 4. Raczej rzadziej niż przed pandemią 5. Zdecydowanie rzadziej niż przed pandemią 6. Trudno powiedzieć
7	Uśredniając i biorąc pod uwagę całą kwotę, jaką dysponujesz miesięcznie i wszystkie wydatki, które ponosisz w ciągu miesiąca – jaka jest Twoja sytuacja finansowa na koniec miesiąca?	syt_fin	Kategoryczna	<ol style="list-style-type: none"> 1. Zdarza się, że nie starcza mi na podstawowe wydatki i zapożyczam się 2. Wystarcza mi tylko na podstawowe, bieżące wydatki, bez możliwości oszczędzania 3. Mogę sobie pozwolić na odłożenie jakiejś kwoty pieniędzy, w zależności od możliwości 4. Odkładam co miesiąc stałą, z góry założoną kwotę

8	Czy masz określone zobowiązania finansowe wobec innych osób/banków w znaczącej kwocie?	zobowiazania	Kategoryczna	<ol style="list-style-type: none"> 1. Kredyt gotówkowy w banku 2. Debet/limit na koncie 3. Kredyt hipoteczny 4. Dług na karcie kredytowej 5. Pożyczka tzw. chwilówka 6. Zakup na raty 7. Długi u rodziny/znajomych 8. Inne 9. Nie mam żadnego zobowiązania finansowego
9	Mój budżet znacząco się zmniejszył i potrzebuję wsparcia finansowego - Z jakiego powodu zdecydowałeś/aś się na zaciągnięcie zobowiązań finansowych?	zobo_powod	Kategoryczna	<ol style="list-style-type: none"> 1. Ze względu na niską stopę procentową 2. Mój budżet znacząco się zmniejszył i potrzebuję wsparcia finansowego 3. Korzystne ceny nieruchomości 4. Utrata źródła utrzymania 5. Ze względu na zwiększenie wydatków podczas pandemii 6. Atrakcyjna oferta zakupu na raty 7. Inne

10	Czy posiadasz oszczędności?	oszcz	Binarna	1. Tak 2. Nie
11	Ile średnio miesięcznie oszczędzasz w PLN?	oszczPLN	Ciągła	Od 0
12	Jak zmieniła się suma oszczędzanych przez Ciebie pieniędzy w trakcie pandemii?	oszcz_zm_w_p	Kategoryczna	1. Przeznaczam zdecydowanie większą część swojego dochodu na oszczędności niż przed pandemią 2. Przeznaczam taką samą część swojego dochodu na oszczędności jak przed pandemią 3. Przeznaczam mniejszą część swojego dochodu na oszczędności niż przed pandemią
13	Jaki wpływ w Twoim odczuciu ma pandemia na Twoje oszczędności?	wplyw_p_oszcz	Kategoryczna	1. Zdecydowanie pozytywny 2. Raczej pozytywny 3. Neutralny 4. Raczej negatywny 5. Zdecydowanie negatywny

Źródło: Opracowanie własne

Na wzór literatury przyjęto do badań zmienne opisujące zarówno dochody jak i wydatki studentów. Przejęto, że najważniejsze wydatki są na podstawowe artykuły spożywcze, jak pandemia wpłynęła na oszczędności oraz subiektywne obawy dotyczące własnej sytuacji finansowej. Wzięto również

pod uwagę zaciągnięte zobowiązania oraz ich powód. Większość zmiennych przyjętych do analizy są zmiennymi kategorycznymi (Tab.1), jedynie pytania o koszty i dochód są zmiennymi liczbowymi ciągłymi. W danych znajdują się braki lub odmowy odpowiedzi na niektóre pytania, jednak ze względu na małą liczbę braków do dalszych analiz przyjęto wszystkie obserwacje. Wykryto również niewiele obserwacji odstających (występują głównie w zmiennych dotyczących dochodów) i tak samo jak, obserwacje z brakującymi danymi, uznano, że nie zaburzą znacząco struktury danych i postanowiono nie usuwać ich z modelu.

Dla tak przyjętych danych zdecydowano się przejść do kolejnego kroku – analizy głównych składowych. W pierwszej kolejności przeprowadzono analizę i za pomocą funkcji *pca()* i oceniono kumulatywną wariancję tłumaczoną przez kolejne wymiary:

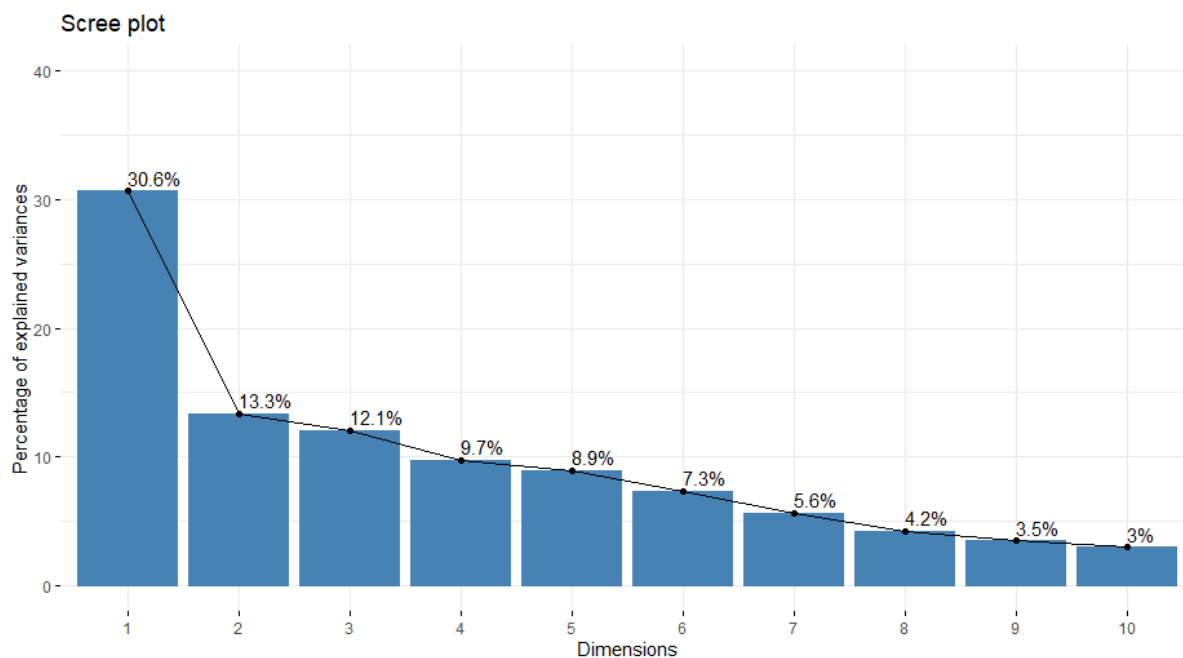
Tabela 2: Analiza wymiarów w PCA

Numer wymiaru	1	2	3	4	5	6	7	8	9	10	11	12
Wariancja	3.68	1.60	1.45	1.17	1.07	0.88	0.67	0.51	0.42	0.36	0.14	0.08
% wariancji	30.63	13.31	12.08	9.74	8.92	7.30	5.60	4.21	3.47	2.97	1.12	0.65
Kumulatywny % wariancji	30.63	43.95	56.02	65.77	74.68	81.98	87.58	91.80	95.26	98.23	99.35	100
Wartość własna	3.68	1.60	1.45	1.17	1.07	0.88	0.67	0.51	0.42	0.36	0.13	0.08

Źródło: Opracowanie własne

Z powyższej tabeli (Tab.2) wynika, że około 75% wariancji jest wyjaśnionej przez 5 wymiarów. Poniżej znajduje się wykres (z ang. scree plot) pokazujący wytłumaczoną zmienność danych pierwotnych przez konkretny wymiar:

Rysunek 5: Wykres tłumaczenia % wariancji przez wymiary

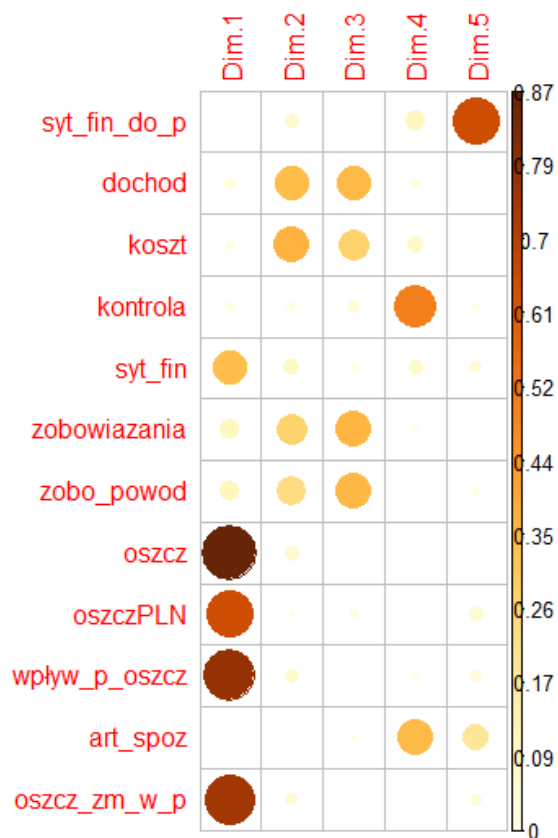


Źródło: Opracowanie własne

Patrząc na wartość własną, widać, że jest ona powyżej 1.0 dla pierwszych pięciu wymiarów, wymiar szósty już posiada ten wyznacznik na poziomie 0.88. Biorąc pod uwagę procent wyjaśnionej wariancji (Rys.5) oraz kryterium Kaisera-Guttmana do dalszej analizy zdecydowano się przyjąć jedynie pięć pierwszych wymiarów. W ten sposób zredukowano liczbę trzynastu zmiennych do pięciu.

Poniżej znajduje się wykres korelacji pomiędzy zmiennymi pierwotnymi, a uzyskanymi wymiarami:

Rysunek 6: Wykres korelacji między zmiennymi a wymiarami

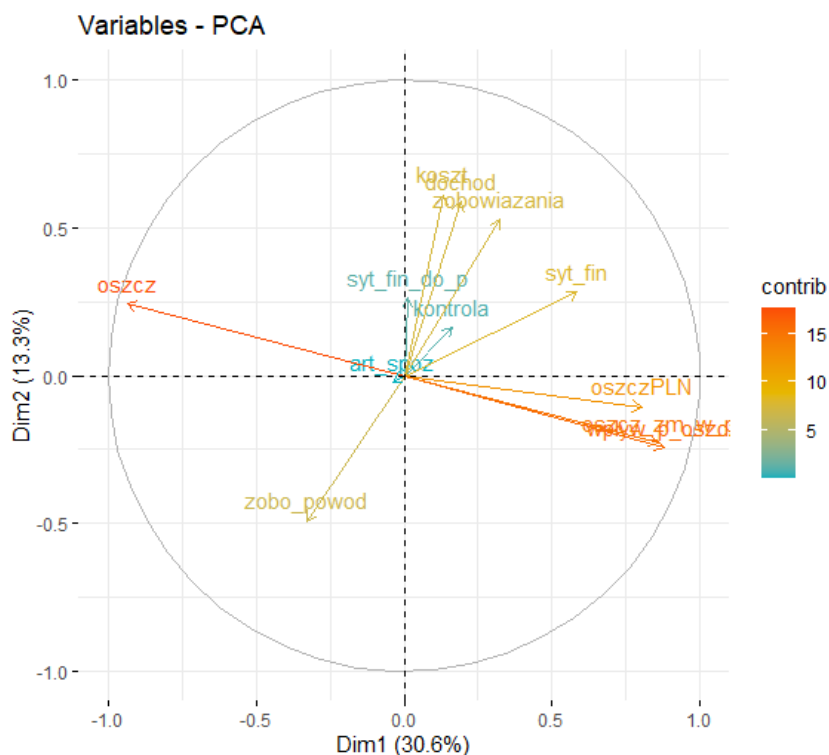


Źródło: Opracowanie własne

Jak widać z macierzy korelacji (Rys.6) z wymiarem pierwszym są najbardziej skorelowane zmienne *oszcz*, *oszczPLN*, *wpływ_p_oszcz* oraz *oszcz_zm_w_p*. Są to zmienne wskazujące na poziom oszczędności studenta przed i po pandemii oraz w jaki sposób się zmieniły. Wymiar drugi i trzeci są głównie skorelowane ze zmiennymi *dochod*, *koszt*, *zobowiazania*, *zobo_powod*. Te zmienne wskazują na poziom dochodu i kosztów jakie ponosi student w codziennym życiu. Również mówią o zaciągniętych zobowiązaniach, tj. kredyt bankowy. Wymiar czwarty jest głównie skorelowany ze zmienną *kontrola*, mówiąca o tym jak często student sprawdza stan swojego konta bankowego oraz zmienna *art_spoz* mówiąca o tym, jak bardzo zmieniły się wydatki na artykuły spożywcze przez pandemię. Ostatni wymiar jest głównie skorelowany ze zmienną *syt_fin_do_p*, która ogólnie mówi o porównaniu sytuacji finansowej respondenta w porównaniu do okresu sprzed pandemii. Kolejnym sposobem na pokazanie korelacji między zmiennymi a dwoma głównymi wymiarami (wymiar 1 i wymiar 2) są poniższe wykresy. Pokazują one również

znaczenie składowej głównej dla wektora przeskalowanej wersji korelacji kwadratowej między zmiennymi a osiami składowymi (contrib). Zmienna ta służy do oceny jakości reprezentacji zmiennych składowej głównej.

Rysunek 7: Wykres przedstawiający zmienne contrib



Źródło: Opracowanie własne

Powyższy wykresy (Rys. 8) można interpretować następująco:

- Zmienne dodatnio skorelowane są zgrupowane razem.
- Zmienne negatywnie skorelowane są umieszczone po przeciwnych stronach (przeciwnie kwadranty).
- Odległość między zmiennymi a początkiem mierzy jakość zmiennych na mapie czynnikowej.
- Zmienne, które są oddalone od początku, są dobrze reprezentowane na mapie czynnikowej.
- Kolor zmiennych dotyczy ich znaczenia i kontrybucji w wymiarach 1 i 2.

W pierwszych dwóch wymiarach największe znaczenie mają zmienne *oszcz*, *oszczPLN*, *wplyw_p_oszcz* oraz *oszcz_zm_w_p*, o czym świadczy długość strzałek oraz intensywny kolor. Najmniej istotne są zmienne *syt_fin_do_p*, *kontrola*, *art_spoz*, ponieważ są one uwzględnione w ostatnich wymiarach (wymiar 4 i 5).

Poniższe tabele (Tab.3 i Tab.4) pokazują kombinacje liniowe pierwotnych zmiennych dla nowych wymiarów ze względu na znaczenie głównej składowej dla wektora contrib.

Tabela 3: Kombinacja liniowa pierwotnych zmiennych jako nowe wymiary ze względu na contrib

CONTRIB	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
<i>syt_fin_do_p</i>	0.0041	4.2621	0.0641	8.8126	60.7429
<i>dochod</i>	1.0122	21.6623	24.1327	2.1851	0.2911
<i>koszt</i>	0.4966	23.3925	19.2068	7.0535	0.0058
<i>kontrola</i>	0.7181	1.6652	3.0046	43.1041	1.6541
<i>syt_fin</i>	9.2768	4.9580	1.1515	5.8406	4.1826
<i>zobowiazania</i>	2.8705	17.4652	25.1205	1.2010	0.0423
<i>zobo_powod</i>	2.9470	15.1633	24.4875	0.0136	1.7201
<i>oszcz</i>	23.7742	3.6626	0.0720	0.1961	0.0001
<i>oszczPLN</i>	17.6271	0.7170	1.2511	0.0620	5.1286
<i>wplyw_p_oszcz</i>	21.0173	3.7045	0.2881	0.8529	3.5723
<i>art_spoz</i>	0.0414	0.0306	1.0025	30.0844	18.7200
<i>oszcz_zm_w_p</i>	20.2148	3.3167	0.2186	0.5941	3.9399

Źródło: Opracowanie własne

Inaczej tabela (Tab.3) przedstawia dane w następujący sposób:

$$\begin{aligned} \text{Dim.1} = & 0.004 * \text{syt_fin_do_p} + 1.012 * \text{dochod} + 0.497 * \text{koszt} + 0.718 * \text{kontrola} + 9.277 * \text{syt_fin} + \\ & 2.87 * \text{zobowiazania} + 2.947 * \text{zobo_powod} + 23.774 * \text{oszcz} + 17.627 * \text{oszczPLN} + \\ & 21.017 * \text{wplyw_p_oszcz} + 0.041 * \text{art_spoz} + 20.215 * \text{oszcz_zm_w_p} \end{aligned}$$

$$\begin{aligned} \text{Dim.2} = & 4.262 * \text{syt_fin_do_p} + 21.662 * \text{dochod} + 23.392 * \text{koszt} + 1.665 * \text{kontrola} + 4.958 * \text{syt_fin} \\ & + 17.465 * \text{zobowiazania} + 15.163 * \text{zobo_powod} + 3.663 * \text{oszcz} + 0.717 * \text{oszczPLN} + \\ & 3.705 * \text{wplyw_p_oszcz} + 0.031 * \text{art_spoz} + 3.317 * \text{oszcz_zm_w_p} \end{aligned}$$

$$\text{Dim.3} = 0.064 * \text{syf_fin_do_p} + 24.133 * \text{dochod} + 19.207 * \text{koszt} + 3.005 * \text{kontrola} + 1.151 * \text{syf_fin} \\ + 25.121 * \text{zobowiazania} + 24.487 * \text{zobo_powod} + 0.072 * \text{oszcz} + 1.251 * \text{oszczPLN} + \\ 0.288 * \text{wpływ_p_oszcz} + 1.002 * \text{art_spoz} + 0.219 * \text{oszcz_zm_w_p}$$

$$\text{Dim.4} = 8.813 * \text{syf_fin_do_p} + 2.185 * \text{dochod} + 7.054 * \text{koszt} + 43.104 * \text{kontrola} + 5.841 * \text{syf_fin} + \\ 1.201 * \text{zobowiazania} + 0.014 * \text{zobo_powod} + 0.196 * \text{oszcz} + 0.062 * \text{oszczPLN} + \\ 0.853 * \text{wpływ_p_oszcz} + 30.084 * \text{art_spoz} + 0.594 * \text{oszcz_zm_w_p}$$

$$\text{Dim.5} = 60.743 * \text{syf_fin_do_p} + 0.291 * \text{dochod} + 0.006 * \text{koszt} + 1.654 * \text{kontrola} + 4.183 * \text{syf_fin} + \\ 0.042 * \text{zobowiazania} + 1.72 * \text{zobo_powod} + 0.0001 * \text{oszcz} + 5.129 * \text{oszczPLN} + \\ 3.572 * \text{wpływ_p_oszcz} + 18.72 * \text{art_spoz} + 3.94 * \text{oszcz_zm_w_p}$$

Jak widać w powyższych równaniach im wyższy współczynnik przy danej zmiennej, tym bardziej jest ona istotna w danym wymiarze, np. dla wymiaru pierwszego (Dim.1) najwyższy współczynnik posiada zmienna *oszcz* – jest on równy 23.7742. Porównując go z macierzą korelacji (Rys.6), gdzie korelacja pomiędzy wymiarem pierwszym a tą zmienną wynosiła około 0.9 oraz z wykresem przedstawiającym contrib dla zmiennych (Rys.7) widać konsekwentność. Kolejnymi najistotniejszymi zmiennymi w wymiarze pierwszym są: *wpływ_p_oszcz*, *oszcz_zm_w_p*, *oszczPLN*. Zatem można powiedzieć, że wymiar pierwszy reprezentuje oszczędności studenta w pandemii. Wymiar drugi głównie reprezentuje kombinację zmiennych: *dochod*, *koszt*, *zobowiazania*, *zobo_powod*, czyli wymiar drugi przedstawia finanse studenta z naciskiem na jego dochody i koszty. Wymiar trzeci również przedstawia finanse studenta, ale z naciskiem na jego zobowiązania finansowe, typu pożyczki. W wymiarze czwartym najistotniejszymi zmiennymi są *kontrola* oraz *art_spoz*, zatem można powiedzieć że Dim.4 reprezentuje wydatki na podstawowe artykuły oraz obawę przed skończeniem się środków finansowych. Wymiar piąty skupia się głównie na jednej zmiennej - *syf_fin_do_p*, czyli porównuje sytuację finansową studenta z przed pandemii do dnia obecnego.

Po przeprowadzeniu analizy głównych składowych następnie zastosowano klastrowanie obserwacji za pomocą metody k-średnich na podstawie indywidualnych koordynatów

poszczególnych obserwacji ww. wcześniej przeprowadzonej PCA. Zdecydowano się na stworzenie pięciu klastrow co jest spójne z wzięciem do analizy pięciu wymiarów. Poniższy wykres przedstawia liczebności poszczególnych klastrow, w sumie łącznie 800 obserwacji:

Rysunek 8: Liczebności klastrow



Źródło: Opracowanie własne

Jak widać na wykresie (Rys.8) najliczniejszymi są klastry numer 1 oraz 2, a najmniej obserwacji przynależy do klastra nr 3, ponieważ tylko 77. Każdy klaster przedstawia grupę studentów sobie podobnych ze względu na odpowiedzi, jakich udzielili w kwestionariuszu na powyższe 13 pytań oraz jak one zostały sklasyfikowane do poszczególnych wymiarów. Dzięki przeprowadzonemu grupowaniu do każdej obserwacji można przyporządkować kolejną zmienną kategorię – numer klastra. Ma ona pięć kategorii:

- k1- klaster pierwszy
- k2 - klaster drugi
- k3 – klaster trzeci
- k4 - klaster czwarty
- k5 – klaster piąty

Każda kategoria jest zmienną binarną. Jednak aby zbudować model ekonometryczny, usunięto zmienną k5, ponieważ gdy dla pozostałych zmiennych, tj. k1 -k4 wartości przyjmą 0, oznacza to, że student znajduje się w klastrze piątym.

Poniższa tabela pokazuje środki (średnie wartości) poszczególnych zmiennych (w tym wypadku wymiarów) dla każdego klastra.

Tabela 4: Korelacja klastrów z wymiarami

Nr Klastra	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5
1	-3.437	-1.579	1.921	0.107	0.475
2	-2.402	0.858	-0.474	0.066	-0.097
3	1.748	1.132	0.649	-0.118	0.309
4	1.150	-0.533	-0.895	-0.050	-0.175
5	0.667	-1.554	1.281	0.136	-0.082

Źródło: Opracowanie własne

Tak otrzymane wyniki można zinterpretować następująco:

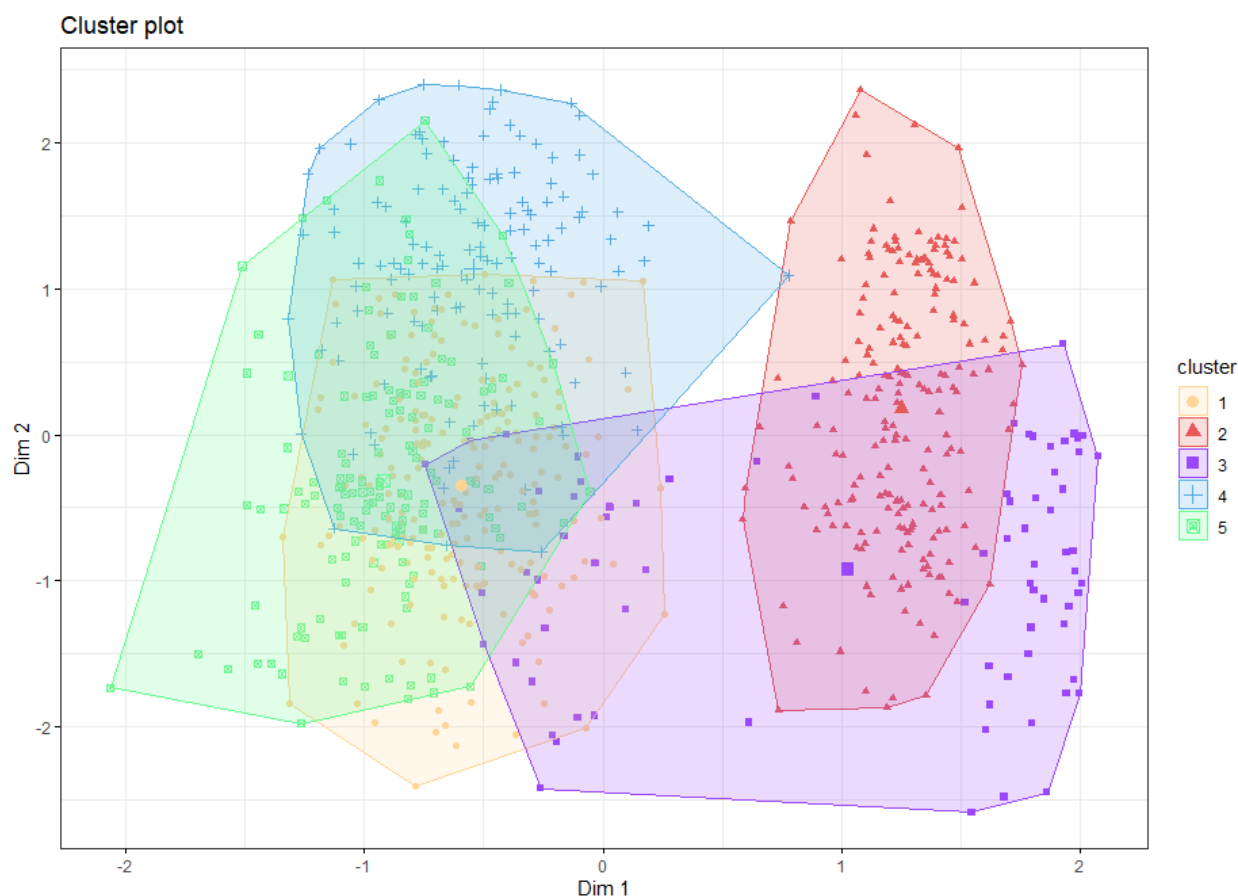
- W klastrze pierwszy znaleźli się studenci, którzy posiadają niskie oszczędności, a pandemia nie miała większego wpływu na ich zmianę. Sytuacja finansowa tych studentów jest na niezbyt dobrym poziomie, ale ogólnie poprawiła się w stosunku do czasu sprzed pandemii.
- W klastrze drugim znaleźli się studenci o podobnych cechach do klastra pierwszego, jedynie z tą różnicą, że pandemia nie miała ogólnego wpływu na ich sytuację finansową.
- W klastrze trzecim znaleźli się głównie studenci, którzy odstają pod względem dobrej sytuacji finansowej, z dużymi oszczędnościami i pozytywnym wpływem pandemii na ich sytuację finansową.
- W klastrze czwartym znaleźli się studenci z niskim dochodem, dużymi zobowiązaniami i proporcjonalnie dużymi oszczędnościami.
- W klastrze piątym znalazły się osoby „średnie”, posiadające dochód i zobowiązania na relatywnie średnim poziomie tak, samo jak oszczędności, wydatki na artykuły spożywcze i sytuację finansową.

Studenci w obrębie jednego klastra są do siebie podobni na poziomie 59.2%, co oznacza na dość duże zróżnicowanie wewnątrz jednej grupy. Gdyby wszystkie obserwacje w ramach grupy znajdowały się dokładnie w tym samym punkcie przestrzeni n-wymiarowej, wówczas osiągnięto by dokładnie 100%.

Poniższy rysunek przedstawia pogrupowane 800 obserwacji pokazane na dwuwymiarowej płaszczyźnie – Dim.1 oraz Dim2. Pamiętając, że wymiar pierwszy głównie pokazuje zmienne związane z oszczędzaniem, a drugi z kosztami o dochodami, można dojść do następujących wniosków:

- Klastry 1,4 oraz 5 są do siebie bardzo podobne zwłaszcza pod względem wymiaru 1. Są to głównie studenci, z obiektywnie niskimi oszczędnościami, a ich większe różnice widać dopiero w dalszych wymiarach.
- Klaster drugi jest najbardziej wyraźny, ponieważ są tam studenci o podobnych cechach z grupy 1, dlatego w osi OY nie widać większej różnicy, natomiast zmiana w osi OX sugeruje małą zmianę w oszczędnościach ze względu na pandemię.
- Klaster trzeci pokazuje głównie obserwacje odstające, które odstają od każdej innej grupy.

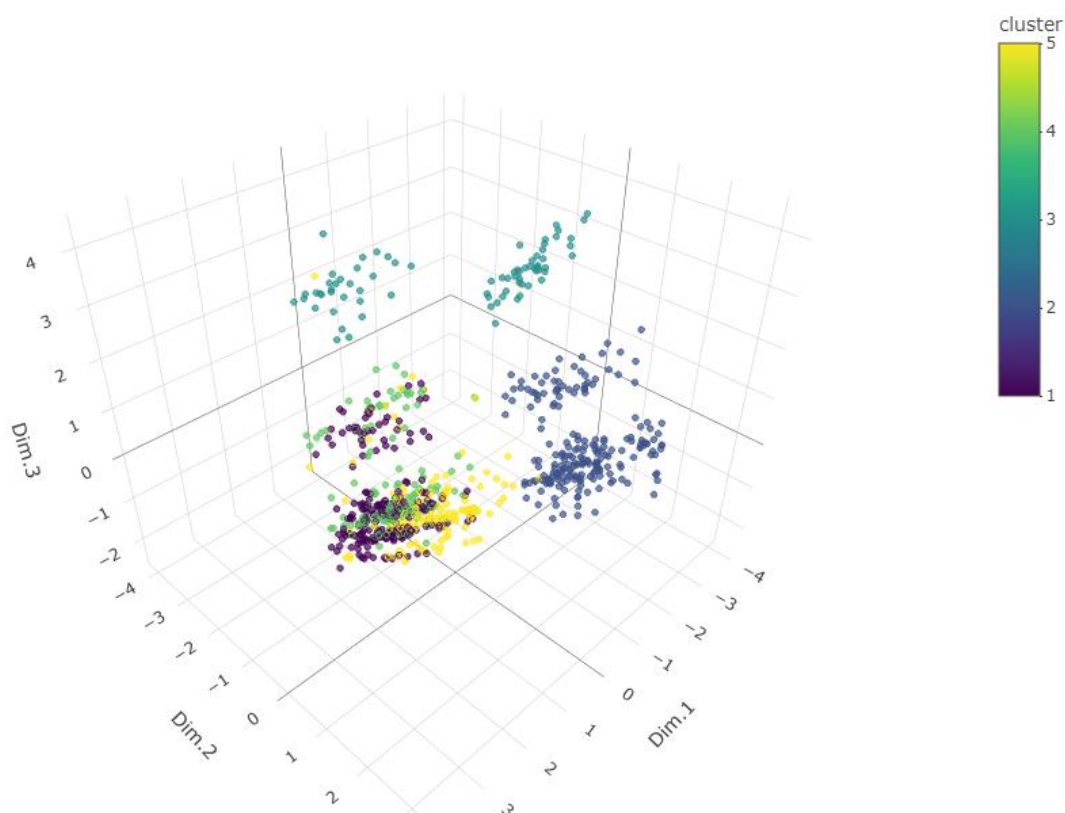
Rysunek 9: Klastry w dwuwymiarowej przestrzeni



Źródło: Opracowanie własne

Poniżej znajduje się rysunek opisujący trzy pierwsze wymiary zamiast dwa. Na nim dużo lepiej widać obserwacje odstające czyli klaster numer 3. Klaster 2 oraz 1 są na tym poziomie w osiach OZ i OY natomiast różnią się tylko w osi OX. Studenci z grupy 5 znajdują się na wszystkich poziomach we wszystkich trzech kierunkach – są to studenci o „średnich” wartościach zmiennych. Klaster czwarty dość dużo pokrywa się z klastrami 5 oraz 4.

Rysunek 10: Klaster w trzywymiarowej przestrzeni



Źródło: Opracowanie własne

Dla tak przygotowanych zmiennych zbudowano model ekonometryczny. Wzorując się na badaniu „The Financial Situation of Students during the COVID-19 Pandemic” (Rozdział 1), gdzie stworzono model logarytmiczny, zdecydowano się na przyjęcie uporządkowanego modelu logitowego. Jako zmienną zależną przyjęto obawę o swoją przyszłość finansową w związku z pandemią - zmienna *obawa*. Jest to zmienna kategoryczna, dlatego też zdecydowano się na podjęcie modelu ze zmienną uporządkowaną zamiast klasycznego modelu logarytmicznego, gdzie

zmienną zależną jest zmienna ciągła. Pierwotnie w kwestionariuszu studenci mieli do wyboru pięć wartości tej zmiennej:

1. Zdecydowanie obawiam się
2. Raczej się obawiam.
3. Raczej się nie obawiam
4. Zdecydowanie się nie obawiam
5. Trudno powiedzieć

Jednak na potrzeby dalszych analiz przekodowano wartości tak, aby zmienna przyjmowała trzy wartości. Zmienna przyjmuje wartość 1 gdy student odpowiedział, że zdecydowanie się obawia o swoją przyszłość finansową oraz, że się raczej obawia. Jako 2 oznaczono odpowiedzi „Trudno powiedzieć”, natomiast 3, gdy student zaznaczył, że raczej się nie obawia lub zdecydowanie się nie obawia. Pozostałymi zmiennymi przyjętymi do modelu jako zmienne zależne są:

- *plec* – oznacza płeć studenta
- *stopien* – oznacza stopień realizowanych studiów (np. stopień licencjacki)
- *semestr* – oznacza obecnie realizowany semestr przez studenta
- *miescowosc* – oznacza wielkość miejscowości pochodzenia studenta (np. miasto powyżej 500 tys.)

Dla tak przygotowanych zmiennych model przyjmuje postać:

$$obawa_i = \beta_1 plec_i + \beta_2 stopien_i + \beta_3 semestr_i + \beta_4 miescowosc_i + \beta_5 k1_i + \beta_6 k2_i + \beta_7 k3_i + \beta_8 k4_i$$

Poniższa tabela przedstawia oszacowane parametry w modelu wraz z ich odchyleniem standardowym i p-value (parametrem wyznaczającym istotność oszacowanego parametru):

Tabela 5: Oszacowane parametry w modelu

Zmienna	Oszacowany parametr	Odchylenie standardowe	p-value
<i>plec</i>	-0.427	0.145	0.003**
<i>stopien</i>	0.114	0.089	0.200
<i>semestr</i>	-0.094	0.061	0.126
<i>miescowosc</i>	0.034	0.047	0.472

$k1$	0.018	0.200	0.930
$k2$	-0.213	0.203	0.294
$k3$	-0.847	0.2896	0.003***
$k4$	0.303	0.227	0.183

Źródło: Opracowanie własne

Z powyższej tabeli wynika, że model można przedstawić w postaci:

$$obawa_i = -0.427plec_i + 0.114stopien_i - 0.094semstr_i + 0.034miejscowosc_i + 0.018k1_i - 0.213k2_i - 0.847k3_i + 0.303k4_i$$

Odchylenia standardowe są dość wysokie zwłaszcza dla zmiennych oznaczających numer grupy, do której należy student. Szczególnie duże jest dla grupy $k1$, gdzie odchylenie jest większe od szacowanego parametru. Wartość p-value wyznacza istotność danego parametru. Symbol gwiazdki * oznacza jak bardzo jest istotna dana zmienna w porównaniu z wartością krytyczną równą 0.05. Im niższe p-value, czyli im niższa wartość od wartości krytycznej, tym bardziej istotny jest dany parametry. Najistotniejsze okazały się być płeć (p-value = 0.003) oraz klaster numer 3 (p-value = 0.003). Najniższą istotnością charakteryzuje się klaster pierwszy, co też pokazało wysokie odchylenie standardowe. Poniższa tabela pokazuje wyniki dla zmiennej zależnej *obawa* odnoszące się do niższej grupy: 2 w porównaniu do 1 oraz 3 w porównaniu do 2. Wyniki te służą do interpretacji oszacowanych parametrów w dalszej części pracy, w tzw. ilorazie szans:

Tabela 6: Oszacowane parametry w modelu dla szans zmiennej zależnej

Zmienna	Oszacowany parametr	Odchylenie standardowe	p-value
1/2	-0.927	0.360	(0.010) *
2/3	-0.501	0.359	(0.164)

Źródło: Opracowanie własne

Kolejna tabela przedstawia kryteria opisujące poprawność dopasowania modelu do danych. Kryteria informacyjne Akaike i Bayesa-Schwarza są na umiarkowanym poziomie – odpowiednio 1513 oraz 1559.8. Im niższe wartości dla tych kryteriów, tym lepiej dopasowany model. Kolejnym

jest RMSE (ang. Root mean square error), czyli pierwiastek błędu średniokwadratowego, zwany też standardowym błędem regresji. Jest on wyrażany w jednostkach wyjściowych, czyli im mniejszy tym lepiej dopasowany model. Wyznacza on również przedziały ufności dla zmiennej zależnej (*obawa*-RMSE; *obawa*+RSME).

Tabela 7: Kryteria dopasowania modelu do danych

AIC	1513.0
BIC	1559.8
RMSE	1.91

Źródło: Opracowanie własne

Aby poprawnie zinterpretować oszacowane parametry należy najpierw obliczyć iloraz szans. Poniższa tabela przedstawia otrzymane wyniki wraz z przedziałami ufności:

Tabela 8: Iloraz szans wraz z przedziałami ufności

zmienna	kategoria	OR	5.0%	95.0%
<i>plec</i>				
	Kobieta	0.674	-0.698	-0.089
	Mężczyzna	.	.	.
<i>miejscowosc</i>				
	Miasto do 50 tys.	0.776	-0.642	0.136
	Miasto od 150 tys. do 500 tys.	1.112	-0.302	0.514
	Miasto od 50 tys. do 150 tys.	0.783	-0.676	0.187
	Miasto powyżej 500 tys.	1.117	-0.319	0.541
	Wieś	.	.	.
<i>stopien</i>				
	Doktoranckie	4.468	0.245	2.75
	Inżynierskie	1.423	-0.089	0.796
	Licencjackie	0.988	-0.416	0.392
	Magisterskie	.	.	.
<i>semestr</i>				
	1	1.174	-0.377	0.696
	2	1.103	-0.417	0.613

3	1.074	-0.438	0.58
4	0.993	-0.547	0.533
5	.	.	.

klaster

k1	1.045	-0.361	0.448
k2	0.827	-0.598	0.217
k3	0.414	-1.453	-0.314
k4	1.372	-0.139	0.771
k5	.	.	.

Źródło: Opracowanie własne

Powyższe wyniki można zinterpretować następująco:

- Mężczyźni mają większe prawdopodobieństwo o ok. 37% w porównaniu do kobiet być w wyższej grupie y, to znaczy mniej się obawiać o swoją przyszłość finansową w związku z pandemią.
- Studenci mieszkający w miastach do 150 tys. mieszkańców w porównaniu do studentów mieszkających na wsi mają ponad 20% wyższą szansę na większe obawianie się o swoje finanse w związku z pandemią, natomiast studenci mieszkający w większych miastach, tj. powyżej 150 tys. mieszkańców, mają ponad 10% większe prawdopodobieństwo na mniejsze obawy związane z finansami w porównaniu ze studentami mieszkającymi na wsi.
- Studenci studiujący na poziomach licencjackim oraz magisterskim mają podobne szanse na obawianie się o swoją przyszłość finansową, studenci studiujący na stopniu inżynierskim mają ponad 40% większą szansę na niższe obawy, a studenci realizujący studia doktoranckie mają niemal 4.5 krotnie większą szansę na niższą obawę o sytuację finansową w związku z pandemią.
- Studenci studiujący na trzecim, czwartym lub piątym semestrze mają podobne szanse na posiadanie obawy o swoją przyszłość finansową na podobnym poziomie, w przeciwieństwie do studentów studiujących na pierwszych dwóch semestrach, gdzie szansa na posiadanie niższych obaw jest kolejno wyższa o około 10% oraz 17%.
- Studenci, którzy znaleźli się w pierwszym i piątym klastrze mają podobne szanse na posiadanie obaw o finanse w związku z pandemią na podobnym poziomie. Studenci z

klastra drugiego mają szansę na większą obawę o przyszłość finansową o niemal 20%, a studenci z klastra trzeciego nawet o 60% w porównaniu do studentów z klastra piątego. Studenci znajdujący się w klastrze czwartym mają największą szansę na posiadanie mniejszych obaw o swoją przyszłość finansową, ponieważ aż o 37% większą niż studenci z klastra piątego.

Iloraz szans (ang. Odds ratio - OR) jest miarą związku między cechą (zmienną niezależną) a zmienną zależną. Porównuje on szanse wystąpienia wyniku (zmiennej zależnej) w grupie posiadającej cechę do szans wystąpienia wyniku w grupie przeciwnej. OR większy niż 1 wskazuje, że nacechowanie jest związane ze zwiększonym prawdopodobieństwem wystąpienia danej zmiennej zależnej, natomiast OR mniejszy niż 1 wskazuje, że nacechowanie jest związane ze zmniejszonym prawdopodobieństwem wystąpienia zmiennej y . OR równy 1 wskazuje, że ekspozycja nie ma wpływu na szanse zmiennej zależnej. Podczas gdy duży OR może sugerować silny związek między dwoma zmiennymi, samo ryzyko może być małe w wartościach bezwzględnych, a na precyzję OR może wpływać również rozpowszechnienie cechy, dlatego istotne jest spojrzenie na przedziały ufności. Podczas interpretacji wyników przedziałów ufności, należy przede wszystkim zwrócić uwagę na to, czy zawiera on wartość 1. Jeśli nie zawiera 1, oznacza to, że prawdziwy OR prawdopodobnie różni się od 1, co z kolei wskazuje na istnienie statystycznie istotnego związku między badanymi zmiennymi. Kierunek związku (tj. czy OR jest większy czy mniejszy niż 1) można określić na podstawie oszacowania punktowego (tj. OR). Jeśli przedział ufności zawiera wartość 1, oznacza to, że prawdziwy OR prawdopodobnie wynosi około 1 i że nie ma wystarczających dowodów, aby stwierdzić, że istnieje związek między zmiennymi. W tym przypadku OR nie jest statystycznie istotny. Dodatkowo, szerokość przedziału ufności może również dostarczyć informacji o precyzji oszacowania OR. Węższy przedział wskazuje, że oszacowanie OR jest bardziej precyzyjne, podczas gdy szerszy przedział wskazuje, że oszacowanie OR jest mniej dokładne. Według powyższej tabeli (tab. 8) jedynie dla jednej zmiennej przedział ufności zawiera 1, jest to zmienna *stopień – Doktoranckie*, oznacza to, że studenci studiujący na poziomie doktoranckim są najmniej istotną kategorią. Pozostałe zmienne nie zawierają 1 w przedziale ufności, zatem można uznać resztę zmiennych za znacząco istotne i słusznie przyjęte do modelu.

Wnioski

Uzyskane wyniki w badaniu pokazały, że duża część studentów obawia się o swoją przyszłość finansową w związku z pandemią koronawirusa. Płeć studenta okazała się mieć szczególne znaczenie, czego się nie spodziewano w fazie początkowej. Mężczyźni wykazują znacznie większe obawy w porównaniu do kobiet. W przeciwieństwie do badań zamieszczonych w artykule pt. „The Impact of the COVID-19 Pandemic on College Students’ Health and Financial Stability in New York City: Findings from a Population-Based Sample of City University of New York (CUNY) Students” z 2021 roku, gdzie to kobiety okazały się mieć większe skłonności do depresji oraz obaw związanych z pandemią. Natomiast w artykule „The Financial Situation of Students during the COVID-19 Pandemic” z 2022 roku wykazał nieznaczące różnice pomiędzy płciami. Miejscowość pochodzenia studenta okazała się nie mieć wielkiego znaczenia, jest to zmienna mało istotna. Jednak jeśli weźmie się ją pod uwagę, można powiedzieć, że studenci pochodzący z większych miast czują mniejsze obawy związane z pandemią pod kątem finansowym. Można to próbować wytłumaczyć większym wsparciem ze strony rodziny, czy większymi możliwościami zatrudnienia. Wiek studentów pokazany jako zmienne *semestr* i *stopień*, czyli poziom edukacji jaki już student otrzymał, również ma mały wpływ na subiektywne odczucie obawy finansowej, jednak studenci starszych lat odczuwają mniejsze obawy w stosunku do młodszych rówieśników. Wpływ dochodów, wydatków czy zaciągniętych zobowiązań, ukryte w danych grupach przynależności studentów mają umiarkowany wpływ. Niektóre grupy takie, jak klaster 3, mają dużo większą istotność statystyczną w porównaniu do innych grup takich, jak klaster 1. Jednak zbiorczo można stwierdzić, że dochody i wydatki studenta, a także zaciągnięte przez niego zobowiązania (w postaci choćby pożyczek chwilowych) ma znaczenia na subiektywne postrzeganie sytuacji finansowej związanej z COVID-19.

Bibliografia

- Czopek, A. (2013). *Analiza porównawcza efektywności metod redukcji zmiennych - analiza składowych głównych i analiza czynnikowa*. Studia Ekonomiczne Uniwersytetu Ekonomicznego w Katowicach.
- D. H. Meier, S. L. (luty 2022). *The Financial Situation of Students during the COVID-19 Pandemic*. Bonn: Insitute of Labor Economics.
- Górnjak, J. (1998, lipiec). Analiza czynnikowa i analiza głównych składowych. *ASK: Research and methods*, strony 83-102.
- Heidi E. Jones, M. M. (2021, Luty 11). The Impact of the COVID-19 Pandemic on College Students' Health and Financial Stability in New York City: Findings from a Population-Based Sample of City University of New York (CUNY) Students. *Urban Health*, strony 187-196.
- K. Szuper, K. H. (2021). *Współczesne problemy gospodarcze – gospodarki w czasach kryzysu. Część II*. Płock: Politechnika Warszawska Kolegium Nauk Ekonomiczno-Społecznych.
- Monika Bazyl, M. G. (2012). *Mikroekonometria. Modele i analizy danych indywidualnych* (wyd. wydanie II rozszerzone). (r. n. Gruszczyński, Red.) Warszawa: Wolters Kluwer Polska Sp. z o.o.
- Muślewski, Ł. (2015). *Kryterialna ocena jakości działania systemu transportu miejskiego*. Uniwersytet Technologiczno-Przyrodniczy w Bydgoszczy.
- P. Kawka, W. M. (2021). Studenci Uniwersytetu Warszawskiego w dobie pandemii COVID-19 – aspekty prawne, kryminologiczne i społeczne. *Biuletyn Kryminologiczny*, 95-122.
- Prof. dr hab. Stanisław Dawidziuk, d. J. (2021). Zarządzanie szkołą wyższą w okresie pandemii koronawirusa SARS-COV-2. *POSTĘPY TECHNIKI przetwórstwa spożywczego*, 206-212.
- Richard A. Settersten Jr., L. B. (2020, wrzesień). Understanding the effects of Covid-19 through a life course lens. *Advances in Life Course Research*.
- Sarah Negash, N. K.-P. (2021). *Worsened Financial Situation During the COVID-19 Pandemic Was Associated With Depressive Symptomatology Among University Students in Germany: Results of the COVID-19 International Student Well-Being Study*. *frontiers in Psychiatry*.
- WeWork. (2020). *The Impact of COVID-19 on the University Student Experience*. WeWork.
- Zbigniew Gołaś, I. K. (2014). Zastosowanie uporządkowanego modelu logitowego w analizie rentowności branż przemysłu spożywczego. *Zagadnienia Ekonomiki Rolnej*, strony 78-96.

Spis rysunków

Rysunek 1: Zmienna nieobserwowalna y^* a zmienna porządkowa y	13
Rysunek 2: Rozkład geograficzny respondentów w postaci mapy pokazana w %	17
Rysunek 3: Wykres przedstawiający miasta studiowania respondentów w %	18
Rysunek 4: Porównanie struktury respondentów z GUS na rok 2021 według płci	19
Rysunek 5: Wykres tłumaczenia % wariancji przez wymiary	25
Rysunek 6: Wykres korelacji między zmiennymi a wymiarami	26
Rysunek 7: Wykres przedstawiający zmienne contrib	27
Rysunek 8: Liczebności klastrów	30
Rysunek 9: Klastry w dwuwymiarowej przestrzeni	32
Rysunek 10: Klastry w trzywymiarowej przestrzeni	33

Spis tabel

Tabela 1: Opis danych przyjętych z badania	19
Tabela 2: Analiza wymiarów w PCA	24
Tabela 3: Kombinacja liniowa pierwotnych zmiennych jako nowe wymiary ze względu na contrib	28
Tabela 4: Korelacja klastrów z wymiarami	31
Tabela 5: Oszacowane parametry w modelu	34
Tabela 6: Oszacowane parametry w modelu dla szans zmiennej zależnej	35
Tabela 7: Kryteria dopasowania modelu do danych	36
Tabela 8: Iloraz szans wraz z przedziałami ufności	36

Streszczenie

Pandemia COVID-19 wywołana w 2020 roku przez wirusa SARS-CoV-2 znacząco odbiła się na młodych ludziach nie tylko pod względem edukacji i zdrowia psychicznego, ale też pod względem finansowym. Wiele studentów straciło pracę, borykało się ze znalezieniem zatrudnienia, walczyła z rosnącymi cenami produktów i czynszów. Zajęcia prowadzące w trybie zdalnym uniemożliwiły prowadzenie zajęć praktycznych i przyczyniła się do powrotu studentów do rodzinnych miast. Niniejsza praca opisuje związek wielu czynników takich, jak dochody, wydatki, ceny artykułów spożywczych, zaciągnięte zobowiązania przez studentów, zmianę oszczędności i też płeć, stopień studiów i wielkość rodzinnej miejscowości z subiektywną obawą studenta dotyczącą przyszłości finansowej z powodu pandemii. Do badań użyto danych opublikowanych w badaniu studenckim SGH, koła naukowego statystyki pt. „WPLYW PANDEMII NA PLANOWANIE KARIERY I OSZCZĘDNOŚCI STUDENTÓW” z 2021 roku. W analizie użyto takich metod, jak analiza głównych składowych, klastrowanie metodą k-średnich oraz uporządkowany model logistyczny, a obliczenia przeprowadzono w języku R.