

Nazwa przedmiotu	Analiza wielowymiarowa
Kod przedmiotu	2400-M1iEPKAW
Profil studiów	ogólnoakademicki
Stopień studiów	drugi
Kierunek studiów	Informatyka i Ekonometria
Specjalność (jeśli dotyczy)	-
Forma studiów	studia stacjonarne oraz niestacjonarne wieczorowe
Rok studiów	I lub II
Semestr	zimowy
Liczba godzin	30h
Forma/typ zajęć	konwersatorium
Punkty ECTS	4
Rodzaj przedmiotu (przedmiot obowiązkowy / kierunkowy do wyboru / fakultatywny / seminarium)	kierunkowy do wyboru
Sposób realizacji przedmiotu (stacjonarny / zdalny / kurs internetowy / kurs realizowany metodą blended learning)	stacjonarny
Język wykładowy	polski
Wymagania formalne	brak
Założenia wstępne	Algebra liniowa: podstawowe operacje rachunku macierzowego. Rachunek prawdopodobieństwa: Pojęcia wartości oczekiwanej, wariancji wektora losowego, macierzy wariancji–kowariancji. Statystyka: Średnia z próby, wariancja z próby, odchylenie standardowe z próby, kowariancja empiryczna, współczynnik korelacji empirycznej, histogram, częstość empiryczna, tablica krzyżowa. Weryfikacja hipotez: hipoteza zerowa i hipoteza alternatywna, poziom istotności, błąd I i II rodzaju, wartość p.
Skrócony opis przedmiotu	Celem zajęć jest zapoznanie uczestników z najważniejszymi metodami statystyki wielowymiarowej. Metody te służą do analizy zjawisk opisywanych przez więcej niż jedną zmienną objaśnianą. Znajdują zastosowanie przy analizie dużych zbiorów danych, umożliwiając wydobycie z danych najbardziej istotnych informacji. Stanowią podstawę współczesnych metod uczenia maszynowego. Celem dodatkowym jest przygotowanie uczestników do praktycznego stosowania poznanych procedur statystycznych oraz samodzielnej interpretacji uzyskanych wyników. Uczestnicy poznają najbardziej popularne i przydatne metody stosowane w badaniach społecznych i badaniach rynku. Zajęcia są prowadzone w środowisku Stata/Python.
Treści kształcenia dla przedmiotu	Celem zajęć jest zapoznanie uczestników z najważniejszymi metodami statystyki wielowymiarowej. Metody te służą do analizy zjawisk opisywanych przez więcej niż jedną zmienną objaśnianą. Znajdują zastosowanie przy analizie dużych zbiorów danych, umożliwiając wydobycie z danych najbardziej istotnych informacji. Stanowią podstawę współczesnych metod uczenia maszynowego. Zajęcia są podzielone na bloki tematyczne. Pierwszym blokiem

	trwającym 1 jednostkę zajęciową jest wprowadzenie do zasad i tematyki zajęć. Blok 2 zawiera powtórzenie i utrwalenie wiedzy z rachunku prawdopodobieństwa i statystyki. Jest rozszerzony o proste modele ilościowe. Kolejność bloków tematycznych od 6 do ostatniej jednostki zajęciowej (w zależności od liczby jednostek zajęciowych w semestrze) zostanie ogłoszona podczas pierwszych zajęć. Bloki merytoryczne składają się z części teoretycznej i części praktycznej oraz proponowanych ćwiczeń do samodzielnego wykonania. Blok 1: Wprowadzenie (1 zajęcia) [1] - sprawy organizacyjne: zasady zaliczenia i wymagania formalne - wprowadzenie do analizy wielowymiarowej Blok 2: Podstawy, które warto znać (4 zajęcia) [2-5] Statystyczne podstawy - skale pomiarowe - statystyczne własności rozkładów - przydatne statystyki - tabele częstości Testy parametryczne i nieparametryczne - porównanie średnich - porównanie wariancji - tabelaryczny opis danych - statystyczne porównanie rozkładów zmiennych Współzależność zjawisk - analiza korelacji, miary współzależności dwóch cech - analizy zróżnicowania: jednoczynnikowa analiza wariancji, wieloczynnikowa analiza wariancji Skalowanie wielowymiarowe i obserwacje nietypowe - skalowanie klasyczne - skalowanie metryczne - obserwacje odstające - obserwacje błędne - winsoryzacja Blok 3: Analiza czynnikowa (2 zajęcia) [6-7] - analiza czynnikowa - metody rotacji - analiza składowych głównych Blok 4: Analiza dyskryminacji (1 zajęcia) [8] - opisowa i predykcyjna analiza dyskryminacyjna - liniowa analiza dyskryminacyjna - kwadratowa analiza dyskryminacyjna - logistyczna analiza dyskryminacyjna - metoda najbliższych sąsiadów Blok 5: Analiza skupień (grupowania, segmentacja) (4 zajęcia) [9-12] Klasyczne metody segmentacji - klasyfikacja - niehierarchiczna analiza skupień (optymalizacja i wybór liczby skupień) - kryteria podziału na grupy, algorytm PAM - hierarchiczna analiza skupień (optymalizacja i wybór liczby skupień) - metody aglomeracyjne, dendrogram - metody podziału, algorytm DIANA - metryka a rezultaty analizy skupień Współczesne techniki segmentacji - Mieszanki rozkładów gausowskich - DBSCAN - inne metody Blok 6: Powtórzenie i zaliczenie (2 zajęcia) [13-14] -konsultacje dla studentów i zaliczenie
--	--

Przedmiotowe efekty uczenia się	Po ukończeniu przedmiotu, student: W ZAKRESIE WIEDZY: • Zna i rozumie w pogłębionym stopniu metody analizy wielowymiarowej oraz ich zastosowanie w modelowaniu złożonych zjawisk ekonomicznych i społecznych. • Zna zaawansowane metody statystyki wielowymiarowej oraz rozumie ich założenia, ograniczenia i konsekwencje ich stosowania w badaniach empirycznych. • Zna nowoczesne narzędzia analityczne i źródła danych, w tym środowiska Stata i Python, wykorzystywane do pozyskiwania i analizy danych ilościowych i jakościowych. W ZAKRESIE UMIEJĘTNOŚCI: • Potrafi formułować i rozwiązywać złożone problemy ekonomiczne z wykorzystaniem wielowymiarowych zbiorów danych, dobierając adekwatne metody analizy i narzędzia informatyczne. • Potrafi wykonywać zaawansowane analizy statystyczne i ekonometryczne, w tym analizę czynnikową, dyskryminacyjną i analizę skupień, oraz krytycznie interpretować uzyskane wyniki. • Potrafi opracowywać i prezentować wyniki badań empirycznych, przygotowując profesjonalne raporty oraz współpracując i pełniąc różne role w zespołach projektowych. W ZAKRESIE KOMPETENCJI: • Jest gotów do krytycznej oceny posiadanej wiedzy i kompetencji analitycznych oraz ciągłego ich rozwijania w odpowiedzi na zmieniające się wymagania nauki i praktyki gospodarczej. • Jest gotów do odpowiedzialnego i etycznego wykonywania zadań analitycznych, z poszanowaniem zasad rzetelności badań, ochrony danych i standardów zawodowych. • Jest gotów do aktywnego uczestnictwa w zespołach analitycznych, także w środowisku zróżnicowanym kulturowo i międzynarodowym.
Powiązanie efektów przedmiotowych z efektami kierunkowymi/specjalnościowym i (oznaczonymi kodami z programu nauczania)	kierunek Informatyka i Ekonometria: K_W02, K_W03, K_W04, K_U02, K_K01, K_K03
Nakład pracy studenta	Szacunkowy nakład pracy studenta: 4ECTS x 25h = 100h (K) - godziny kontaktowe (S) - godziny pracy samodzielnej zajęcia: 30h (K) 0h (S) konsultacje: 5h (K) 0h (S) przygotowanie do zajęć, praca z literaturą: 0h (K) 30h (S) przygotowanie do zaliczenia: 0h (K) 35h (S) Razem: 35h (K) + 65h (S) = 100h
Metody i kryteria oceniania	Student jest obowiązany do obecności podczas zajęć. Zgodnie z zapisem paragrafu 33 Regulaminu studiów na Uniwersytecie Warszawskim studenci nieobecni podczas zajęć kierują do prowadzącego prośbę o usprawiedliwienie nieobecności bez zbędnej zwłoki. Nieobecności w liczbie przekraczającej 3 nie są usprawiedliwiane i powodują brak klasyfikacji studenta. Test zaliczeniowy (100%). Uczestnicy zajęć mogą zdobyć dodatkowe punkty wykonując prace domowe. Skala ocen: [0%-50%) – ndst

	[50%-60%) – dst [60%-70%) – dst + [70%-80%) – db [80%-90%) – db+ [90%-100%] – bdb
Literatura	Podstawowa: 1. Francuz P., Mackiewicz L., Liczby nie wiedzą skąd pochodzą, KUL 2007. 2. Kaufman L., Rousseeuw P., Finding Groups in data. An introduction to cluster analysis, JohnWiley & Sons, 1990. 3. Rabe-Hesketh S., Everitt B., A Handbook of Statistical Analyses using Stata”, Chapman and Hall, 2007 4. Xu, D., Tian, Y. A Comprehensive Survey of Clustering Algorithms. Ann. Data. Sci. 2, 165–193, 2015. 5. https://doi.org/10.1007/s40745-015-0040-1 Dodatkowa: 1. Everitt B., Landau S., Leese M. Stahl D., Cluster analysis, John Wiley and Sons, Nowy Jork, 2011. 2. Gatnar E., Walesiak M. (red), Metody statystycznej analizy wielowymiarowej w badaniach marketingowych, Wydawnictwo AE we Wrocławiu, Wrocław 2004. 3. Herczyński J., Strawiński P., Postawy zawodowe nauczycieli: próba typologii, Edukacja 2014, nr 3, str. 22-37. 4. Jajuga K., Statystyczna analiza wielowymiarowa, PWN, 1993 5. Larose D.T. Odkrywanie wiedzy z danych. Wprowadzenie do eksploracji danych, Wydawnictwo Naukowe PWN, Warszawa 2006 6. Morrison D.F., Wielowymiarowa analiza statystyczna, PWN, Warszawa 1990 7. Panek T., Statystyczne metody wielowymiarowej analizy porównawczej, SGH, Warszawa 2009 8. Rószkiewicz M., Zarys metod statystyki wielowymiarowej z wykorzystaniem programów komputerowych, SGH, Warszawa 1990.
Metody dydaktyczne	• wykład problemowy • dyskusja kierowana • ćwiczenia praktyczne w środowiskach Stata i Python • metoda problemowa (rozwiązywanie zadań analitycznych) • metoda projektowa (zadania indywidualne i zespołowe) • samodzielna praca studenta (prace domowe)
Oprogramowanie	Stata, Python