

KARTA KURSU

Nazwa	Programy do analizy danych
Nazwa w j. ang.	Software for data analysis

Koordynator	dr hab. prof. UKEN Łukasz Binkowski	Zespół dydaktyczny
Punktacja ECTS*	4	dr hab. prof. UKEN Łukasz Binkowski

Opis kursu (cele kształcenia)

Przedstawione zostaną podstawy analizy danych obejmujące gromadzenie danych, budowanie ich bazy, analizę oraz przygotowanie raportu. W ramach kursu studenci poznają najpopularniejsze programy komputerowe stosowane w analizie danych, w tym arkusz kalkulacyjny Microsoft Excel z dodatkiem Power Pivot, Microsoft Access, StatSoft Statistica, Microsoft Power BI oraz R. Opis funkcji programów (szczególnie Microsoft Excel) będzie dostosowany do wiedzy posiadanej już przez studentów. Omówione zostaną zasady tworzenia bazy danych w arkuszu kalkulacyjnym oraz programie do zarządzania bazami danych. Studentom przedstawione także będzie zastosowanie języka SQL, VBA oraz składni DAX. Praca w programach będzie odbywała się w oparciu o proste analizy statystyczne (analiza wariancji, analiza korelacji, EDA). Studenci także będą poznawać funkcje edytora tekstu pomocne w redagowaniu rozbudowanych opracowań i raportów z analiz. Dodatkowo w kursie zaprezentowane będzie kilka rozwiązań IT ułatwiających codzienną pracę z danymi (VPN, skrypty automatyzacyjne, digitalizery wykresów, itp.).

Umiejscowienie kursu w planie studiów względem jego głównej tematyki

Semestr	Kurs
1	Matematyka
1	Wprowadzenie do statystyki
1	Wstęp do programowania
1	Rachunek prawdopodobieństwa
2	Programy do analizy danych
2	Programowanie obiektowe
2	Algorytmy i struktury danych
3	Statystyczna analiza danych
3	Specjalizowane języki programowania
3	Modele matematyczne nauk przyrodniczych
4	Bazy danych
4	Grafika komputerowa i wizualizacja danych
5	Metodyka badań naukowych
5	Grafika inżynierska
5	Modelowanie procesów biologicznych
6	Techniki prezentacji danych
7	Analiza danych wysokoprzepustowych

Warunki wstępne

Wiedza	Znajomość: podstawowych procesów biologicznych i ekologicznych, podstaw statystyki.
Umiejętności	Obsługa programów w środowisku Windows, Linux lub macOS. Podstawowa obsługa pakietu biurowego Microsoft Office. Czytanie tekstów w języku angielskim.
Kursy	Matematyka, Wprowadzenie do statystyki, Wstęp do programowania, Rachunek prawdopodobieństwa.

Efekty uczenia się (wiedza)

	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
Wiedza	W01: Zna zastosowanie wybranego oprogramowania do analizy danych biologicznych	K_W03, K_W13
	W02: Zna główne założenia baz danych w arkuszach kalkulacyjnych	K_W09, K_W11
	W03: Zna podstawowe ograniczenia liczbowej i graficznej analizy danych	K_W03, K_W09, K_W18
	W04: Zna podstawowe zasady bezpieczeństwa i higieny pracy podczas analizy danych	K_W20
	W05: rozróżnia i objaśnia różnice pomiędzy różnymi systemami operacyjnymi i typami plików	K_W11
	W06: definiuje pojęcia z zakresu edycji dokumentów elektronicznych oraz komputerowego składu tekstu	K_W03
	W07: wymienia sposoby wyszukiwania i selekcjonowania informacji pochodzących z różnych źródeł (baz)	K_W01, K_W03
	W08: posiada podstawowe informacje o obrazie cyfrowym	K_W03
	W09: zna narzędzia wykorzystywane w prostym modelowaniu i symulacji danych biologicznych	K_W03, K_W12

Efekty uczenia się (umiejętności)

	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
Umiejętności	U01: Konstruuje bazy danych do analiz wyników i identyfikuje problemy podczas tej pracy	K_U01, K_U07, K_U12
	U02: Planuje i wykonuje analizy z użyciem oprogramowania	K_U03, K_U05, K_U06, K_U08
	U03: Planuje makra i skrypty programowe w analizie danych	K_U12, K_U13
	U04: sporządza dokumenty tekstowe o rozbudowanej strukturze wykorzystując narzędzia automatyzujące pracę	K_U07
	U05: przygotowuje dokument wielostronicowy do druku, w tym: stosuje style, spisy, podpisy dla obiektów, menadżer bibliografii i inne odwołania	K_U01, K_U0, K_U08
	U06: korzysta z narzędzi pracy zespołowej (rejestracja zmian w dokumencie, komentarze, konspekt)	K_U08, K_U10
	U07: poprawnie dobiera funkcje w arkuszu kalkulacyjnym do rozwiązywanych problemów: matematyczne, statystyczne, logiczne, wyszukiwania, daty i czasu oraz tekstowe,	K_U03, K_U06, K_U12, K_U13
	U08: przeprowadza proste symulacje oraz eksportuje i wizualizuje dane	K_U06
	U09: wyszukuje i pobiera dane w naukowych bazach danych i w literaturowych bazach danych	K_U01, K_U06, K_U08

	(m.in. Scopus, Web of Science, Mendeley, Google Scholar)	
	U10: w sposób automatyczny wykonuje zmiany plików tekstowych z danymi oraz tworzy wykresy o zadanych parametrach	K_U02, K_U06

Efekty uczenia się (kompetencje społeczne)

Kompetencje społeczne	Efekt uczenia się dla kursu	Odniesienie do efektów kierunkowych
	K01: Korzysta z technik kształcenia zdalnego i rozumie potrzebę samokształcenia	K_K01, K_K06
	K02: Postępuje z powierzonym sprzętem zgodnie z obowiązującymi procedurami	K_K05
	K03: Sprawnie planuje swoją pracę w ramach kursu wraz z wyborem zadań priorytetowych i tych, których wykonanie można zaplanować na dalsze zajęcia	K_K02, K_K03, K_K06
	K04: pracuje zespołowo i rozumie konieczność systematycznej pracy nad wieloetapowymi projektami	K_K02
	K05: korzysta z różnych źródeł informacji do poszerzania własnej wiedzy i zdobywania nowych umiejętności	K_K06

Organizacja

Forma zajęć	W	A	K	L	S	P	E
Liczba godzin			45				
Forma zaliczenia			zal.				

Opis metod prowadzenia zajęć

Zajęcia prowadzone przez platformę Microsoft Teams z wykorzystaniem komputerów studentów. Każdy student w porozumieniu i wg instrukcji prowadzącego pracuje samodzielnie na komputerze. Studenci także otrzymują polecenia wykonania zadań domowych.

Cześć pracy przewidziana jest w oparciu o materiały i aktywności zamieszczone w kursie na platformie moodle <https://elearning.up.krakow.pl>, kurs: PDAD. Szacowana liczba godzin przeznaczona na pracę studentów z platformą elearningową ~ 20.

Pliki robocze do analiz znajdują się na dysku sieciowym, do którego dostęp studenci uzyskają po rozpoczęciu kursu.

Formy sprawdzania efektów uczenia się

	E-learning	Gry dydaktyczne	Ćwiczenia w szkole	Zajęcia terenowe	Praca laboratoryjna	Projekt indywidualny	Projekt grupowy	Udział w dyskusji	Referat	Praca pisemna (esej)	Egzamin ustny	Egzamin pisemny	Inne
W01					x	x		x				x	
W02					x	x		x				x	
W03					x	x		x				x	
W04					x	x		x					
W05					x	x		x				x	
W06					x	x		x				x	
W07					x	x		x				x	

W08					x	x		x					x	
W09					x	x		x					x	
U01					x	x		x						
U02					x			x						
U03					x	x		x						
U04					x	x		x						
U05					x	x		x						
U06					x	x		x						
U07					x	x		x						
U08					x	x		x						
U09					x	x		x						
U10					x	x		x						
K01								x						
K02					x	x								
K03					x	x		x						
K04					x			x						
K05					x	x		x						

Kryteria oceny

Aktywny udział w zajęciach, praca nad projektami (zadaniami domowymi) oraz zaliczenie testu zaliczeniowego. Zadania i test prowadzone są przez platformę moodle.

Uwagi

Kurs prowadzony w języku polskim.

Studenci w trakcie kursu pracują na własnych komputerach, na których instalują wymagane oprogramowanie wskazane przez prowadzącego (Microsoft Office lub jego bezpłatna alternatywa), Microsoft Power BI (oprogramowanie bezpłatne do celów edukacyjnych), StatSoft Statistica (licencja udostępniona przez UP), R, R Studio, Mendeley Reference Manager (oprogramowanie bezpłatne).

Treści merytoryczne (wykaz tematów)

1. Systemy operacyjne i ich implikacje z uwagi na analizę danych.
2. Bazy danych w analizie danych – wprowadzenie do języka SQL i relacyjnych baz danych.
3. Przegląd popularnych platform analitycznych: Microsoft Power BI i Tableau.
4. Zaawansowane funkcje programu Microsoft Excel: wybrane funkcje, dodatki Power Query, Power Pivot oraz język VBA i składnia DAX.
5. Wprowadzenie do języka programowania statystycznego R i programu R Studio.
6. Praca w środowisku Statsoft Statistica.
7. Przygotowanie danych i analiza danych w programie R Studio.
8. Eksport wyników analiz i elementów graficznych.
9. Analizy obrazu (Plot Digitizer)
10. Naukowe bazy danych i bazy bibliograficzne.
11. Menadżery bibliografii.
12. Aplikacje usprawniające pracę nad analizą danych.
13. Składnia markdown i jej zastosowanie.
14. Skrypty, makra i automatyzacja w analizie danych.
15. Zaawansowane funkcje edytora tekstu mające znaczenie dla analityków danych (m.in. śledzenie zmian, style, integracja z menadżerem bibliografii, VBA).

Wykaz literatury podstawowej

1. Internetowy podręcznik statystyki Statsoft <http://www.statsoft.pl/textbook/stathome.html>
2. Biecek P. 2011. Przewodnik po pakiecie R, Oficyna Wydawnicza Gis, Wrocław.
3. Internetowa strona projektu R <https://cran.r-project.org>
4. Systemy pomocy, w które wyposażone są wykorzystywane na zajęciach narzędzia, programy i aplikacje komputerowe, w tym np. pomoc techniczna dla pakietu Office <https://support.office.com>

5. Publikacja naukowa koordynatora zawierająca szczegółowy opis analizy statystycznej „Binkowski Ł.J., Meissner W. 2013. Levels of metals in blood samples from Mallards (*Anas platyrhynchos*) from urban areas in Poland. *Environmental Pollution* 178, 336–342”.
6. Publikacja naukowa koordynatora zawierająca szczegółowy opis zaawansowanej analizy statystycznej „Binkowski Ł.J., Merta, D., Przystupińska A., Sołtysiak Z., Pacoń J., Stawarz R. 2016. Levels of metals in kidney, liver and muscle tissue and their relation to the occurrence of parasites in the red fox in the Lower Silesian Forest in Europe. *Chemosphere* 149, 161–167”.
7. Publikacja naukowa koordynatora zawierająca szczegółowy opis analizy statystycznej „Kucharska K., Binkowski Ł.J., Dudzik K. 2021. Spatial and temporal trends in mercury levels in the down of black stork chicks in central Europe. *Environmental Pollution* 274: 116571. DOI: 10.1016/j.envpol.2021.116571”.
8. Publikacja naukowa koordynatora zawierająca szczegółowy opis analizy statystycznej „Binkowski L.J., Fort J., Brault-Favrou M., Gallien F., Le Guillou G., Chastel O., Bustamante P. 2021. Foraging ecology drives mercury contamination in chick gulls from the English Channel. *Chemosphere* 267: 128622. DOI: 10.1016/j.chemosphere.2020.128622”.

Wykaz literatury uzupełniającej

1. Fora internetowe poświęcone R, Microsoft Excel, makrom VBA i analizom statystycznym.

Bilans godzinowy zgodny z CNPS (Całkowity Nakład Pracy Studenta)

Liczba godzin w kontakcie z prowadzącymi	Wykład	0
	Konwersatorium (ćwiczenia, laboratorium itd.)	30
	E-learning	0
	Pozostałe godziny kontaktu studenta z prowadzącym	5
liczba godzin pracy studenta bez kontaktu z prowadzącymi	Lektura w ramach przygotowania do zajęć	15
	Przygotowanie krótkiej pracy pisemnej lub referatu po zapoznaniu się z niezbędną literaturą przedmiotu	25
	Przygotowanie projektu lub prezentacji na podany temat	15
	Przygotowanie do egzaminu/zaliczenia	10
Ogółem bilans czasu pracy		100
Liczba punktów ECTS w zależności od przyjętego przelicznika		4