

WYDZIAŁ MATEMATYKI**KARTA PRZEDMIOTU**

| | |
|---------------------------------------|---|
| Nazwa przedmiotu w języku polskim: | Analiza Danych Ankietych |
| Nazwa przedmiotu w języku angielskim: | Categorical Data Analysis |
| Kierunek studiów: | Matematyka, Matematyka i Analiza Danych |
| Specjalność: | Uczenie maszynowe i inżynieria danych (na kierunku Matematyka) |
| Stopień studiów i forma: | I stopień, stacjonarna |
| Rodzaj przedmiotu: | Wybieralny |
| Kod przedmiotu: | |
| Grupa kursów: | TAK |

| | Wykład | Ćwiczenia | Laboratorium | Projekt | Seminarium |
|---|----------------------------|-----------|--------------|---------|------------|
| Liczba godzin zajęć zorganizowanych w Uczelni (ZZU) | 30 | | 15 | | |
| Liczba godzin całkowitego nakładu pracy studenta (CNPS) | 90 | | 60 | | |
| Forma zaliczenia | zaliczenie na ocenę | | | | |
| Dla grupy kursów zaznaczyć kurs końcowy | X | | | | |
| Liczba punktów ECTS | 3 | | 2 | | |
| w tym liczba punktów odpowiadająca zajęciom o charakterze praktycznym (P) | | | 2 | | |
| w tym liczba punktów ECTS odpowiadająca zajęciom wymagającym bezpośredniego udziału nauczycieli lub innych osób prowadzących zajęcia (BU) | 1,1 | | 0,6 | | |

WYMAGANIA WSTĘPNE W ZAKRESIE WIEDZY, UMIEJĘTNOŚCI I INNYCH KOMPETENCJI

Zna podstawowe pojęcia i twierdzenia rachunku prawdopodobieństwa i statystyki matematycznej takie jak: zmienna losowa, rozkład prawdopodobieństwa, zbieżność rozkładów, prawa wielkich liczb, centralne twierdzenie graniczne, model statystyczny, statystyka, estymator, test statystyczny.

CELE PRZEDMIOTU

- C1 Przedstawienie rodzajów badań statystycznych i rodzajów danych ankietych.
- C2 Przedstawienie metod konstrukcji przedziałów ufności dla prawdopodobieństwa sukcesu.
- C3 Wyrobienie umiejętności wyznaczania przedziałów ufności dla prawdopodobieństwa sukcesu.
- C4 Przedstawienie testów stosowanych w analizie danych ankietych, w szczególności testów niezależności.
- C5 Wyrobienie umiejętności weryfikowania hipotez w analizie danych ankietych.
- C6 Przedstawienie modeli dla danych zależnych i metod analizy takich danych.
- C7 Wyrobienie umiejętności analizy danych zależnych.
- C8 Przedstawienie miar zależności i miar zgodności oraz podstawowych pojęć analizy korespondencji.
- C9 Wyrobienie umiejętności wyznaczania miar zależności i miar zgodności oraz ich interpretacji oraz przeprowadzania analizy korespondencji.

C10 Przedstawienie modeli log-liniowych dla danych z tabel wielodzielczych.
 C11 Wyrobienie umiejętności wyboru modelu log-liniowego dla danych z tabel wielodzielczych i ich interpretacji.

PRZEDMIOTOWE EFEKTY KSZTAŁCENIA

Z zakresu wiedzy:

PEK_W01 zna rodzaje badań statystycznych i rodzaje danych ankietowych.
 PEK_W02 zna metody konstrukcji przedziałów ufności dla prawdopodobieństwa sukcesu.
 PEK_W03 zna testy stosowane w analizie danych ankietowych.
 PEK_W04 zna miary zależności i miary zgodności i podstawowe pojęcia analizy korespondencji.
 PEK_W05 zna metody analizy danych zależnych.
 PEK_W06 zna modele log-liniowe dla danych z tabel wielodzielczych i metody wyboru odpowiedniego modelu log-liniowego do danych.

Z zakresu umiejętności:

PEK_U01 potrafi wyznaczać przedziały ufności dla prawdopodobieństwa.
 PEK_U02 potrafi weryfikować hipotezy w analizie danych ankietowych.
 PEK_U03 potrafi wyznaczać miary zależności i miary zgodności i je interpretować oraz przeprowadzać analizę korespondencji.
 PEK_U04 potrafi analizować dane zależne.
 PEK_U05 potrafi dopasować odpowiedni model log-liniowy do danych z tabel wielodzielczych i podać jego interpretację.

Z zakresu kompetencji społecznych:

PEK_K01 potrafi korzystać z literatury zalecanej do kursu oraz samodzielnie wyszukiwać dodatkowe materiały w celu poszerzenia swojej wiedzy.
 PEK_K02 potrafi twórczo współdziałać w grupie studenckiej, budować pozytywne więzi emocjonalne z jej członkami.

| TREŚCI PROGRAMOWE | | |
|------------------------------|---|----------------------|
| Forma zajęć - wykłady | | Liczba godzin |
| Wy1 | Rodzaje badań i rodzaje danych ankietowych. Rozkład dwumianowy, rozkład hipergeometryczny, rozkład wielomianowy i produkt rozkładów wielomianowych. | 2 |
| Wy2 | Estymacja punktowa i przedziałowa prawdopodobieństwa sukcesu. | 4 |
| Wy3 | Testowanie hipotez dotyczących parametru rozkładu dwumianowego. Testowanie równości dwóch prawdopodobieństw sukcesu. | 2 |
| Wy4 | Dokładne i asymptotyczne testy niezależności w tabelach dwuwymiarowych. | 4 |
| Wy5 | Miary zależności i miary zgodności. | 2 |
| Wy6 | Analiza korespondencji. | 2 |
| Wy7 | Modele dla danych wielomianowych zależnych (powtarzanych). Testowanie symetrii, quasy symetrii i brzegowej jednorodności. | 4 |
| Wy8 | Paradoks Simpsona. Tabele wielodzielcze wyższych wymiarów. | 2 |
| Wy9 | Modele log-liniowe dla danych z tabeli wielodzielczych. Estymacja współczynników modelu log-liniowego. | 2 |
| Wy10 | Testowanie hipotez dotyczących parametrów modeli log-liniowych dla danych z tabeli wielodzielczych. | 2 |
| Wy11 | Wybór modelu log-liniowego. | 2 |
| Wy12 | Kolokwium zaliczeniowe. | 2 |
| Suma godzin | | 30 |

| Forma zajęć - laboratorium | | Liczba godzin |
|-----------------------------------|---|----------------------|
| La1 | Podstawowe informacje o pracy z wybranym pakietem statystycznym. Wprowadzanie danych, ich modyfikacja; sprawdzanie poprawności danych. Graficzna ilustracja danych ankietowych. | 2 |
| La2 | Symulacyjne porównania różnych przedziałów ufności dla parametru rozkładu dwumianowego w przypadku małej i dużej liczby danych. | 2 |
| La3 | Testowanie hipotez dotyczących parametru rozkładu dwumianowego. Testowanie równości dwóch prawdopodobieństw sukcesu. | 2 |
| La4 | Dokładne i asymptotyczne testy niezależności w tabelach dwuwymiarowych. | 2 |
| La5 | Obliczanie miar zależności i miar zgodności oraz ich interpretacja. Analiza korespondencji. | 2 |
| La6 | Modele dla danych wielomianowych zależnych (powtarzanych). Testowanie symetrii, quasi-symetrii i brzegowej jednorodności. | 2 |
| La7 | Paradoks Simpsona na przykładach danych rzeczywistych. Modele log-liniowe dla danych z tabel wielozmiennych – wybór i interpretacja. | 3 |
| Suma godzin | | 15 |

| STOSOWANE NARZĘDZIA DYDAKTYCZNE |
|---|
| <ol style="list-style-type: none"> 1. Wykład informacyjny, problemowy – metoda tradycyjna i prezentacja multimedialna. 2. Laboratorium. 3. Konsultacje. 4. Praca własna studenta – przygotowanie raportów z analizy danych. |

OCENA OSIĄGNIĘCIA EFEKTÓW KSZTAŁCENIA

| Oceny (F – formująca (w trakcie semestru), P – podsumowująca (na koniec semestru)) | Numer efektu kształcenia | Sposób oceny osiągnięcia efektu kształcenia |
|--|---|---|
| F1 | PEK_W01-PEK_W06, PEK_U01-PEK_U05, PEK_K01, PEK_K02. | Odpowiedzi ustne, raporty |
| F2 | PEK_W01-PEK_W06, PEK_K01. | Test |
| $P=0,7*F1+0,3*F2$ | | |

LITERATURA PODSTAWOWA I UZUPEŁNIAJĄCA

LITERATURA PODSTAWOWA:

- [1] Agresti A. *An Introduction to Categorical Data Analysis*. John Wiley & Sons, New York, 2007.
- [2] Fienberg, S. E. *The Analysis of Cross-Classified Categorical Data*. Springer, New York, 2007.
- [3] Rudas, T. *Lectures on Categorical Data Analysis*. Springer Science+Business Media, New York, 2018.
- [4] Santner T. J., Duffy D. E. *The Statistical Analysis of Discrete Data*. Springer-Verlag, New York, 1989.

LITERATURA UZUPEŁNIAJĄCA:

- [1] Collet D. *Modelling Binary Data*. Chapman & Hall, New York, 1991.
- [2] Jobson, J. D. *Applied Multivariate Data Analysis. Volume II: Categorical and Multivariate Methods*. Springer Science+Business Media, New York, 1992.
- [3] Magiera R. *Modele i metody statystyki matematycznej. Część II Wnioskowanie statystyczne*. GIS 2018.

OPIEKUN PRZEDMIOTU (IMIE, NAZWISKO, ADRES E-MAIL)

dr hab. Alicja Jokiel-Rokita, prof. uczelni (Alicja.Jokiel-Rokita@pwr.edu.pl)