



Uczelniana Oferta Studiów Zaawansowanych	
SYLABUS 2015/2016	
Nazwa przedmiotu	Analiza danych z pakietem R (grupa II – wersja zaawansowana) ADzR2
Liczba punktów ECTS	Proponowana liczba punktów: 3 ECTS

Osoby prowadzące	Tytuł naukowy	Imię i nazwisko	Katedra / Instytut / Centrum / Inne
	Dr hab. inż.	Anna Dembińska	Zakład Procesów Stochastycznych i Matematyki Finansowej, Wydział MiNI PW
Osoba odpowiedzialna za przedmiot	Dr hab. inż.	Anna Dembińska	Zakład Procesów Stochastycznych i Matematyki Finansowej, Wydział MiNI PW

Semestr studiów	Semestr letni 2015/2016
Typ przedmiotu (możliwości wyboru) obowiązkowy O fakultatywny F	F
Wymagania wstępne Zakres wiadomości / kompetencji / umiejętności, jakie powinien już posiadać student przed rozpoczęciem nauki przedmiotu, a także specyfikacja innych przedmiotów lub programów, które należy zaliczyć wcześniej.	Podstawy z rachunku prawdopodobieństwa: pojęcie prawdopodobieństwa i jego własności, niezależność zdarzeń losowych, podstawowe rozkłady prawdopodobieństwa ciągłe i dyskretne, podstawowe wiadomości dotyczące zmiennych losowych jedno- i wielowymiarowych, Centralne Twierdzenie Graniczne. Podstawy ze statystyki: znajomość podstawowych statystyk próbkowych i podstawowych metod graficznej prezentacji danych, znajomość pojęcia testu statystycznego.
Poziom przedmiotu Podstawowy P Średniozaawansowany Ś Zaawansowany Z	Ś

Charakter zajęć , liczba godzin w semestrze, liczba godzin w tygodniu. 1) podać rodzaj prowadzonych zajęć dla danego przedmiotu: wykłady (W); ćwiczenia (Ć); laboratorium (L); projekt (P) 2) podać liczbę godzin w tygodniu np. W - 2; Ć - 2; L - 3; P - 0 3) podać liczbę godzin w semestrze np. W - 30; Ć - 30; L - 45; P - 0	<i>W – 2 godz. w tygodniu, łącznie: 30 godzin</i>
Sugerowana liczba godzin pracy własnej	45 godzin obejmuje : 30 godzin przygotowywanie się słuchacza do wykładów, 15 – przygotowywanie się słuchacza do egzaminu.
Całkowita liczba godzin:	75 godzin
Aspekty międzynarodowe (jeśli są)	
Język wykładowy	Polski
Cel przedmiotu Opis zakładanych kompetencji i umiejętności, jakie student nabywa w wyniku zaliczenia przedmiotu.	Sprawność w rozwiązywaniu praktycznych problemów z wykorzystaniem metod statystycznych i pakietu R. Znajomość szerokiej gamy modeli statystycznych oraz umiejętność doboru stosownej procedury statystycznej i jej implementacji dla rzeczywistych danych.
Treść przedmiotu	
<ol style="list-style-type: none"> 1. Wstępna analiza danych z pakietem R: podstawowe miary liczbowe dla danych ilościowych, graficzna prezentacja danych ilościowych i jakościowych. 2. Rozkłady prawdopodobieństwa: wyznaczanie prawdopodobieństw i kwantyli, generowanie prób losowych, prezentacja graficzna. Badanie własności wybranych rozkładów dyskretnych i ciągłych przydatnych w modelowaniu statystycznym. 3. Badanie zgodności. Sprawdzanie normalności rozkładu: metody graficzne i testy normalności. Testowane zgodności z dowolnym rozkładem: test zgodności chi-kwadrat Pearsona i test Kołmogorowa-Smirnova. Szacowanie z próby nieznanymi parametrami rozkładu. 4. Regresja liniowa: dopasowywanie modelu do danych, sprawdzanie czy dopasowany model dobrze opisuje dane, obserwacje wpływowe i odstające, przekształcanie zmiennych, metoda ważonych najmniejszych kwadratów, dobór zmiennych objaśniających do modelu (częściowy test F, kryteria służące do wyboru najlepszego modelu, metoda dołączania, eliminacji i selekcji krokowej). Prognozowanie przy użyciu modelu regresji. 5. Jednoczynnikowa i dwuczynnikowa analiza wariancji: analiza wariancji jako szczególny przypadek regresji liniowej, wykresy średnich w grupach, sprawdzanie czy spełnione są założenia modelu, testy w modelu analizy wariancji, interakcje w dwuczynnikowej analizie wariancji. 6. Analiza kowariancji – model regresji liniowej w sytuacji, gdy wśród zmiennych objaśniających są zarówno zmienne typu ilościowego jak i jakościowego. 7. Uogólnione modele liniowe: regresja logistyczna, probitowa, Poissona i ujemna dwumianowa. Dopasowanie modelu do danych i diagnostyka jego dopasowania, odchylenie zerowe i odchylenie resztowe modelu, prognozowanie. 	
Spis zalecanych lektur	
LP.	Autor, Tytuł, Wydawnictwo,
1.	P. Dalgaard, „Introductory Statistics with R”, Springer, 2008
2.	J. Koronacki, J. Mielniczuk, „Statystyka dla studentów kierunków technicznych i przyrodniczych”, Wydawnictwa Naukowo-Techniczne, Warszawa, 2006
3.	P. Biecek, „Przewodnik po pakiecie R”, Oficyna Wydawnicza GIS, Wrocław, 2008

4.	J.J. Faraway „Practical Regression and ANOVA Using R”, www.stat.lsa.umich.edu/~faraway/book
----	--

Metody oceny (ocena, egz. pisemny, egz. ustny, projekt)	Praca podczas zajęć – rozwiązywanie projektów.
--	--

Uwagi dodatkowe	Zajęcia odbędą się, jeżeli zapisze się co najmniej 20 osób.
------------------------	---

Tabela 1. Efekty kształcenia

Numer (symbol)	Efekty kształcenia słuchacz, który zaliczył przedmiot, potrafi	Sposób weryfikacji osiągnięcia efektu
WIEDZA		
ADzR2_W1	Zna i wie jak dopasować do danych następujące modele: model regresji liniowej jednokrotnej, regresji liniowej wielokrotnej oraz jednoczynnikowej i dwuczynnikowej analizy wariancji.	Projekt
ADzR2_W2	Zna uogólnienia modelu regresji liniowej na przypadek, gdy zmienna objaśniana lub jedna ze zmiennych objaśniających jest zmienną jakościową.	Projekt
ADzR2_W3	Zna metody graficzne i testy służące do analizy danych jakościowych.	Projekt
UMIEJĘTNOŚCI		
ADzR2_U1	Potrafi dobrać odpowiednią procedurę statystyczną do rozwiązania konkretnego, praktycznego problemu.	Projekt
ADzR2_U2	Mając rzeczywiste dane, potrafi zastosować do nich wybraną procedurę statystyczną a następnie zweryfikować adekwatność modelu skonstruowanego w oparciu o wybraną procedurę.	Projekt
ADzR2_U3	Przeprowadzenie formalnych testów podczas statystycznej analizy danych umie poprzedzić wstępną analizą wykorzystującą metody graficzne.	Projekt
ADzR2_U4	Umie sprawnie posługiwać się pakietem R podczas przeprowadzania analizy danych.	Projekt
KOMPETENCJE		
ADzR2_K1	Rozumie konieczność diagnozowania modelu, użytego do opisu rzeczywistych danych, i wprowadzania w nim stosownych modyfikacji.	Obserwacja na zajęciach
ADzR2_K2	Rozumie, że automatyczne użycie procedur statystycznych, bez wniknięcia w istotę i charakter danych, grozi wyciągnięciem błędnych wniosków.	Obserwacja na zajęciach