

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ДВНЗ “ПРИКАРПАТСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ВАСИЛЯ СТЕФАНИКА”**

Факультет математики та інформатики

Кафедра математичного і функціонального аналізу

СИЛАБУС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

СТАТИСТИЧНИЙ АНАЛІЗ В R

Освітня програма “Прикладна та теоретична статистика”

Спеціальність 112 “Статистика”

Галузь знань 11 “Математика та статистика”

Затверджено на засіданні кафедри
Протокол № 1 від “27” серпня 2020 р.

ЗМІСТ

1. Загальна інформація
2. Анотація до курсу
3. Мета та цілі курсу
4. Результати навчання та компетентності
5. Організація навчання курсу
6. Система оцінювання курсу
7. Політика курсу
8. Рекомендована література

1. Загальна інформація					
Назва дисципліни	Статистичний аналіз в R				
Рівень вищої освіти	Другий (магістерський)				
Викладач (-і)	Дмитришин Роман Іванович				
Контактний телефон викладача					
Е-mail викладача	roman.dmytryshyn@pnu.edu.ua				
Формат дисципліни					
Обсяг дисципліни	6 кредитів ЄКТС				
Посилання на сайт дистанційного навчання	https://d-learn.pnu.edu.ua/				
Консультації	Очні консультації: згідно розкладу консультацій				
2. Анотація до курсу					
Курс “Статистичний аналіз в R” містить теоретичні відомості по роботі з інтерактивними компонентами R; опис мови і базових графічних можливостей середовища; опис поширених процедур обробки даних і побудови статистичних моделей; просторовий аналіз і зображення результатів на географічних картах і схемах.					
3. Мета та цілі курсу					
Мета та цілі курсу полягають у формуванні в студентів нових теоретичних знань і практичних навичок з аналізу та візуалізації даних з використанням середовища R.					
4. Результати навчання та компетентності					
Результати навчання:					
ПРН-9. Уміти використовувати в практичній діяльності та розробляти спеціалізоване статистичне програмне забезпечення.					
ПРН-13. Уміти інтегрувати знання з різних галузей для розв’язання теоретичних та практичних задач і проблем.					
Компетентності:					
ЗК-4. Здатність використовувати комп’ютерні та інформаційні технології в професійній діяльності.					
ФК-1. Знання на рівні новітніх досягнень, необхідні для дослідницької або практичної діяльності у сфері математики, статистики та їхніх практичних застосувань.					
ФК-7. Здатність аналізувати статистичні алгоритми, оцінювати їхню обґрунтованість та ефективність.					
ФК-8. Здатність використовувати спеціалізовані мови програмування та програмне забезпечення для розв’язання задач дослідницького або практичного характеру.					
5. Організація навчання курсу					
Обсяг курсу					
Вид заняття			Загальна кількість годин		
лекції			28		
семінарські заняття / практичні / лабораторні			0/16/16		
самостійна робота			120		
Ознаки курсу					
Семестр	Спеціальність	Курс (рік навчання)	Нормативний / вибірковий		
1	112 Статистика	I	Нормативний		
Тематика курсу					
Тема, план	Форма заняття	Література	Завдання, год	Вага оцінки	Термін виконання
Змістовий модуль 1. Програмне середовище R					

Тема 1. Основні компоненти статистичного середовища R. Історія появи та основні принципи організації середовища R. Робота з командною панеллю інтерфейсу R. Об'єкти, пакети, функції та пристрої.	Лекція, практичне заняття.	[1-10]	Опрацювання лекційного матеріалу, 2; розв'язування задач, 1.	0.2	За розкладом
Тема 2. Опис мови R. Типи даних мови R. Вектори і матриці. Фактори. Списки і таблиці. Імпортування даних в R. Зображення даних і часу: часові ряди. Організація обчислень: функції, розгалуження, цикли. Векторизовані обчислення в R із використанням apply-функцій.	Лекція, практичне заняття, лабораторне заняття.	[1-10]	Опрацювання лекційного матеріалу, 2; розв'язування задач, 3.	0.4	За розкладом
Тема 3. Графічні можливості R. Діаграми розсіювання plot() і параметри графічних функцій. Гістограми, функції ядерної густини і функція cdfplot(). Діаграми розмахів. Кругові і стовпчикові діаграми. Діаграми Клівленда та одновимірні діаграми розсіювання. Категоризовані графіки.	Лекція, практичне заняття, лабораторне заняття.	[1-10]	Опрацювання лекційного матеріалу, 2; розв'язування задач, 2.	0.4	За розкладом
Змістовий модуль 2. Статистичний аналіз					
Тема 4. Описова статистика і підбір розподілів. Оцінка вибірових параметрів із використанням спеціальних функцій. Використання функцій summary() і додаткових пакетів. Аналіз викидів. Заповнення пропущених значень у	Лекція, практичне заняття, лабораторне заняття.	[1-10]	Опрацювання лекційного матеріалу, 4; розв'язування задач, 3.	0.15	За розкладом

таблицях даних. Відтворюваність результатів при використанні генератора випадкових чисел. Закони розподілу ймовірностей, реалізовані в R. Підбір закону і розподілу в R. Перевірка на нормальність розподілу.					
Тема 5. Класичні методи і критерії статистики. Гіпотеза про рівність середніх двох генеральних сукупностей. Використання рангових критеріїв Уїлкоксона – Манна – Уїтні. Рандомізація, бутстреп та оцінка статистичної потужності (на прикладі двохвибіркового <i>t</i> -критерія). Гіпотеза про однорідність дисперсій. Вступ в дисперсійний аналіз. Оцінка кореляції двох випадкових величин. Критерій хі-квадрат. Точний тест Фішера, критерії Мак-Немара і Кохрана – Мантеля – Хензеля. Оцінка статистичної потужності при порівнянні часток.	Лекція, практичне заняття, лабораторне заняття.	[1-10]	Опрацювання лекційного матеріалу, 4; розв'язування задач, 3.	0.15	За розкладом
Тема 6. Лінійні моделі в дисперсійному аналізі. Протокол розвідувального аналізу даних. Лінійні моделі дисперсійного аналізу. Структура модельних об'єктів дисперсійного аналізу. Оцінка адекватності моделі дисперсійного аналізу.	Лекція, практичне заняття, лабораторне заняття.	[1-10]	Опрацювання лекційного матеріалу, 4; розв'язування задач, 4.	0.2	За розкладом

Дисперсійний аналіз за Краскелом – Воллісом. Моделі дво- і багатofакторного дисперсійного аналізу. Особливості в лінійних моделях, що містять категоріальні предиктори. Проблема численних перевірок статистичних гіпотез. Методи порівняння групових середніх в дисперсійному аналізі.					
Тема 7. Регресійні моделі залежності між кількісними змінними. Статистична модель. Проста лінійна регресія: який вік Всесвіту? Моделі регресії при різних видах функції втрат. Критерії вибору моделей оптимальної складності. Поліноміальні і нелінійні моделі регресії. Модель множинної регресії і вибір її специфікації. Процедури діагностики моделей множинної регресії. Регресія гребеня і ласо: використання головних компонент. Порівняння ефективності різних моделей при прогнозуванні.	Лекція, практичне заняття, лабораторне заняття.	[1-10]	Опрацювання лекційного матеріалу, 4; розв'язування задач, 4.	0.2	За розкладом
Тема 8. Узагальнені, структурні та інші моделі регресії. Моделі згладжування. Узагальнені моделі регресії. Моделі logit- та probit-регресії. Узагальнені моделі для оцінки показників екологічної толерантності. Коваріаційний аналіз.	Лекція, практичне заняття, лабораторне заняття.	[1-10]	Опрацювання лекційного матеріалу, 4; розв'язування задач, 3.	0.15	За розкладом

Моделі зі змішаними ефектами для ієрархічно організованих даних. Індуктивні моделі (метод групового обліку аргументів). Моделювання структурними рівняннями.					
Тема 9. Просторовий аналіз і створення криптограм. Проста карта: використання растрового малюнка та обчислення відстаней. Аналіз просторового розміщення точок. Використання сервісів картографічної системи Google Maps. Створення картограм за допомогою R.	Лекція, практичне заняття, лабораторне заняття.	[1-10]	Опрацювання лекційного матеріалу, 2; розв'язування задач, 3.	0.15	За розкладом
6. Система оцінювання курсу					
Загальна система оцінювання курсу	<p>100 бальна: 14 балів за змістовний модуль 1 (6 балів – модульна контрольна робота, 4 бали – лабораторні роботи, 4 бали – опитування); 36 балів за змістовний модуль 2 (16 балів – модульна контрольна робота, 10 балів – лабораторні роботи, 10 балів – опитування). 50 балів за екзамен (20 балів за теоретичні питання, 30 балів за практичні завдання).</p> <p>Критерії оцінювання знань, умінь і навичок студентів: 90 – 100 (відмінно) – студент демонструє повні і глибокі знання навчального матеріалу, достовірний рівень розвитку умінь та навичок, правильне й обґрунтоване формулювання практичних висновків, наводить повний обґрунтований розв'язок прикладів та задач, аналізує причинно-наслідкові зв'язки; вільно володіє науковими термінами; 70 – 89 (добре) – студент демонструє повні знання навчального матеріалу, але допускає незначні пропуски фактичного матеріалу, вміє застосувати його до розв'язання конкретних прикладів та задач, у деяких випадках нечітко формулює загалом правильні відповіді, допускає окремі несуттєві помилки та неточності розв'язках; 50 – 69 (задовільно) – студент володіє більшою частиною фактичного матеріалу, але викладає його не досить послідовно і логічно, допускає істотні пропуски у відповіді, не завжди вміє правильно застосувати набуті знання до розв'язання конкретних прикладів та задач, нечітко, а інколи й невірно формулює основні твердження та причинно-наслідкові зв'язки;</p>				

	0 – 49 (незадовільно) – студент не володіє достатнім рівнем необхідних знань, умінь, навичок, науковими термінами.
Вимоги до письмової роботи	Відповідно до робочого навчального плану, студент виконує дві контрольні роботи. Головна мета їх – перевірка самостійної роботи студентів у процесі навчання, виявлення ступеня засвоєння ними теоретичних положень курсу. При розв'язанні задач студент має детально вказувати, яким саме був хід його роздумів, якими формулами і твердженнями він користувався.
Семінарські заняття	Практичні заняття проводиться з метою формування у студентів умінь і навичок з курсу, вирішення сформульованих завдань, їх перевірка та оцінювання. За метою і структурою практичні заняття є ланцюжком, який пов'язує теоретичне навчання і навчальну практику з курсу, а також передбачає попередній контроль знань студентів. Оцінки за практичні заняття враховуються при виставленні підсумкової оцінки з курсу.
Умови допуску до підсумкового контролю	Оцінка за опитування і модульні контрольні роботи має бути не менша, ніж 25 балів .
7. Політика курсу	
<p>Самостійне виконання навчальних завдань, завдань поточного та підсумкового контролю результатів навчання (для осіб з особливими освітніми потребами ця вимога застосовується з урахуванням їхніх індивідуальних потреб і можливостей). Посилання на джерела інформації у разі використання ідей, розробок, тверджень, відомостей. Надання достовірної інформації про результати власної навчальної (наукової, творчої) діяльності, використанні методики досліджень і джерела інформації. Засвоєння пропущеної теми лекції з поважної причини перевіряється під час складання підсумкового контролю. Пропуск лекції з неповажної причини відпрацьовується студентом відповідно вимог кафедри, що встановлені на засіданні кафедри (співбесіда, реферат тощо). Пропущені практичні та/або лабораторні заняття, незалежно від причини пропуску, студент відпрацьовує згідно з графіком консультацій. Поточні незадовільні оцінки, отримані студентом під час засвоєння відповідної теми на практичному та/або лабораторному занятті, перескладаються викладачеві, який веде заняття до складання підсумкового контролю з обов'язковою відміткою у журналі обліку роботи академічних груп.</p>	
8. Рекомендована література	
<ol style="list-style-type: none"> 1. Буховец А.Г. и др. Статистический анализ данных в системе R. Воронеж: ВГАУ, 2010. 2. Волкова П.А., Шипунов А.Б. и др. Статистическая обработка данных в учебно-использовательских работах. Москва: Форум, 2012. 3. Baayen R.H. Analyzing Linguistic Data: A Practical Introduction to Statistics using R. Cambridge University Press, 2008. 4. Bivand R.S., Pebesma E., Gómez-Rubio V. Applied Spatial Data Analysis with R. Springer, 2013. 5. Bloomfield V.A. Using R for Numerical Analysis in Science and Engineering. Chapman, Hall/CRC, 2014. 6. Carmona R. Statistical Analysis of Financial Data in R. 2nd ed. Springer, 2014. 7. Hollander M., Wolfe D.A., Chicken E. Nonparametric Statistical Methods. Wiley, 2014. 8. Shumway R.H., Stoffer D.S. Time Series Analysis and Its Applications: With R Examples. Springer, 2011. 9. Krijnen W. Applied Statistics for Bioinformatics Using R. Hanze University, 2009. 10. Zhao Y., Cen Y. Data Mining Applications with R. Academic Press, 2013. 	

Викладач _____