

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА

МЕХАНІКО-МАТЕМАТИЧНИЙ ФАКУЛЬТЕТ
Кафедра теорії ймовірностей, статистики та актуарної математики

«ЗАТВЕРДЖУЮ»

Заступник декана
з навчальної роботи



Безущак Безущак О.О.

15 вересня 2018 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
СТАТИСТИКА БІОЛОГІЧНИХ ДАНИХ
для здобувачів освітньо-наукового рівня «доктор філософії»

галузь знань	11 «Математика та статистика»
спеціальність	112 «Статистика»
освітній рівень	третій (освітньо-науковий)
освітньо-наукова програма	«Статистика»
вид дисципліни	вибіркова

Форма навчання	денна
Навчальний рік	2018/2019
Рік навчання	2
Кількість кредитів ECTS	4
Мова викладання, навчання та оцінювання	англійська/українська
Форма заключного контролю	екзамен

Викладачі: : професор Майборода Ростислав Євгенович, д.ф.-м.н.

Пролонговано: на 20¹⁹/20²⁰ н.р. *Безущак* Безущак О.О. «18» вересня 2019 р.
на 20²⁰/20²¹ н.р. *Безущак* (Безущак О.О.) «14» вересня 2020 р.

КИЇВ – 2018

Розробник: професор кафедри теорії ймовірностей, статистики та актуарної математики
Майборода Ростислав Євгенович, д. ф.-м. н., проф.

ЗАТВЕРДЖЕНО


Завідувач кафедри теорії ймовірностей,
статистики та актуарної математики



(підпис) Мішура Ю.С.

Протокол №18 від «22» *червня* 2018 р.

Схвалено науково-методичною комісією механіко-математичного факультету

Протокол від «12» *09* 2018 року № *1*
Голова науково-методичної комісії 

(підпис) професор, д.ф.-м.н. Курченко О.О.

1. Мета дисципліни Розвиток навичок розв'язання комплексних проблем в галузі статистики, використання новітніх інформаційних і комунікаційних технологій, здатності до абстрактного мислення, здатності до пошуку, оброблення та аналізу інформації з різних джерел, вміння генерувати нові ідеї, навичок роботи в міжнародному науковому просторі, навичок формулювання дослідницьких задач з статистики, розробляти наукові проекти та керувати ними, спілкуватися на професійні теми з неспеціалістами у галузі статистики, застосовувати теоретичні знання до розв'язання прикладних задач математичного захисту інформації, забезпечувати належну якість власної наукової та прикладної роботи а також навичок статистичного аналізу даних медико-біологічних та генетичних досліджень.

2. Попередні вимоги до опанування або вибору навчальної дисципліни:

1. *Знати:* основні методи математичного аналізу, теорії міри, теорії ймовірностей та математичної статистики, дискретної математики та теорії алгоритмів.
2. *Вміти:* проводити критичний аналіз, оцінку і синтез нових ідей і підходів в галузі статистики, самостійно застосовувати методи теорії ймовірностей і математичної статистики при розв'язанні нових задач.

3. Анотація навчальної дисципліни:

Дисципліна «Статистика біологічних даних» належить до вибіркового компоненту освітньої програми, блоку дисциплін вільного вибору аспіранта. Вона забезпечує професійний розвиток, спрямована на формування концептуальних та методологічних знань у галузі математики і статистики, вміння критично аналізувати, оцінювати і синтезувати нові та комплексні ідеї, аналізувати наукові праці, формулювати методологічну базу власного наукового дослідження, здатність формулювати наукову проблему, робочі гіпотези досліджуваної проблеми. В рамках дисципліни вивчаються основні принципи та методи статистичного аналізу результатів медичних, біологічних та генетичних досліджень, сучасні програмні засоби такого аналізу.

4. Завдання (навчальні цілі): набуття знань, умінь та навичок (компетентностей) на рівні новітніх досягнень у математиці, відповідно науково-освітньої кваліфікації «Доктор філософії». Зокрема, розвивати: вміння аналізувати сучасні передові концептуальні та методологічні знання; здатність проводити критичний аналіз, оцінку і синтез нових та складних ідей; здатність застосовувати теоретичні та практичні підходи математики; вміння розробляти наукові і інформаційно-освітні ресурси для розв'язання професійних задач, пов'язаних з розвитком та використанням математики і статистики; зокрема, у аналізі біологічних даних. Набуття компетентностей:

1. Здатність до абстрактного мислення, аналізу та синтезу (ЗК-1).
2. Навички використання новітніх інформаційних і комунікаційних технологій (ЗК-2).
3. Здатність проведення самостійних досліджень на сучасному рівні (ЗК-3).
4. Здатність до пошуку, оброблення на аналізу інформації з різних джерел (ЗК-4).
5. Здатність генерувати нові ідеї (креативність) (ЗК-5).
6. Здатність працювати в міжнародному науковому просторі (ЗК-6).
7. Здатність розробляти та управляти науковими проектами (ЗК-7).
8. Здатність до застосування теоретичних знань до розв'язання прикладних задач (ЗК-9).
9. Здатність оцінювати і забезпечувати належну якість власної наукової та прикладної роботи (ЗК-11)
10. Здатність формулювати наукову проблему, робочі гіпотези досліджуваної проблеми, що передбачає глибоке переосмислення наявних та створення нових цілісних знань та/або професійної практики (ФК-1).
11. Здатність до самостійного опрацювання та аналізу літературних джерел у галузі математичної та прикладної статистики (ФК-2).

12. Здатність формулювати математико-статистичні моделі явищ реального світу (ФК-3).
13. Здатність до вибору існуючих та створення нових методів та методик аналізу реальних статистичних даних відповідно до поставленої задачі (ФК-4).
14. Здатність до оцінки ефективності методів статистичного аналізу даних за допомогою аналітичного дослідження та імітаційного моделювання (ФК-5).
15. Здатність представляти та обґрунтовувати результати теоретичних та прикладних статистичних досліджень у формі, яка відповідає можливостям сприйняття аудиторії (ФК-6).
16. Здатність до алгоритмічної реалізації нових методів статистичного аналізу даних (ФК-7).
17. Здатність до використання сучасного програмного забезпечення для прикладного статистичного аналізу та для дослідження якості нових статистичних алгоритмів (ФК-8).
18. Знання специфічних особливостей процесів навчання статистичним методам та формування статистичного мислення (ФК-9).

5. Результати навчання за дисципліною:

Результат навчання (РН) (1. знати; 2. вміти; 3. комунікація; 4. автономність та відповідальність)		Форми (та/або методи і технології) викладання і навчання	Методи оцінювання та пороговий критерій оцінювання (за необхідності)	Відсоток у підсумковій оцінці з дисципліни	
Код	Результат навчання				
РН 1.1	Знати означення баєсового, класифікатора. Знати методи побудови баєсових класифікаторів при заданих щільностях розподілу класів та апіорних ймовірностях..	<i>Лекція, самостійна робота</i>	<i>Контрольна робота 1 (60% правильних відповідей), екзамен, активна робота на лекції, усні відповіді</i>	5%	
РН 1.2	Вміти застосовувати емпірично-баєсової класифікації для аналізу даних біологічних досліджень.			5%	
РН 1.3	Вміти використовувати техніку лінійного дискримінантного аналізу і знати обмеження області її використання.			5%	
РН 1.4	Знати основні методи цілеспрямованого проектування, що застосовуються у аналізі біологічних та генетичних даних.		<i>Контрольна робота 2 (60% правильних відповідей), екзамен, активна робота на лекції, усні відповіді</i>	5%	
РН 1.5	Вміти застосовувати техніку багатовимірною шкалування для аналізу даних генетичних досліджень.			5%	
РН 1.6	Знати методи динамічного обчислення відстаней між ДНК-подібними структурами.			<i>Екзамен, активна робота на лекції, усні відповіді</i>	5%
РН 1.7	Знати поняття лагранжіана навантаженого графа та його зв'язок з технікою спектральної кластеризації.				5%
РН 2.1	Знати принципи побудови деревинних класифікаторів.	<i>Практичне заняття, самостійна робота</i>	<i>Контрольна робота 1 (60% правильних відповідей), екзамен, виконання завдань, винесених на самостійну роботу</i>	10%	
РН 2.2	Вміти застосовувати техніку CART для побудови класифікаторів біологічних об'єктів.			10%	
РН 2.3	Знати основні означення моделі скінченної суміші розподілів та вміти застосовувати їх для побудови ймовірнісних моделей, що описують розподіл даних біологічних досліджень.		<i>Контрольна робота 2 (60% правильних відповідей), екзамен, виконання</i>	10%	

			завдань, винесених на самостійну роботу	
PH 2.4	Знати принципи роботи EM-алгоритма.		Екзамен, виконання завдань, винесених на самостійну роботу	10%
PH 2.5	Вміти застосовувати техніки ієрархічних класифікацій та аналізувати дерево класифікації.			10%
PH3.1	Здатність працювати у міжнародному просторі, обґрунтовувати власний погляд на задачу та формулювати робочі гіпотези, спілкуватися з колегами з питань застосування методів та теорій математики, писати наукові роботи	Практичне заняття, самостійна робота	Виступ з доповіддю за темою наукового дослідження	5%
PH4.1	Демонстрація авторитетності, інноваційність, високий ступінь самостійності, академічна та професійна доброчесність, послідовна відданість розвитку нових ідей або процесів у передових контекстах професійної та наукової діяльності.			5%
PH4.2	Відповідально ставитися до виконуваних робіт, нести відповідальність за їх якість			5%

6. Співвідношення результатів навчання дисципліни із програмними результатами навчання

Результати навчання дисципліни	PH 1.1	PH 1.2	PH 1.3	PH 1.4	PH 1.5	PH 1.6	PH 1.7	PH 2.1	PH 2.2	PH 2.3	PH 2.4	PH 2.5	PH 3.1	PH 4.1	PH 4.2
	Програмні результати навчання														
<i>(з опису освітньої програми)</i>															
ПРН-3-1. Знати праці провідних зарубіжних вчених, наукові школи та фундаментальні праці у галузі дослідження, формулювати мету власного наукового дослідження як складову загально цивілізаційного процесу;	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ПРН-3-2. Знати принципи фінансування науково-дослідної роботи та структуру кошторисів на її виконання, вміння підготувати запит на отримання фінансування, звітну документацію.														+	+
ПРН-3-3. Моніторинг наукових джерел інформації відносно досліджуваної проблеми.	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ПРН-3-4. Визначати методологічні принципи та методи наукового дослідження галузі інформаційних технологій в залежності від об'єкту і предмету, використовуючи міждисциплінарні підходи.	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

ПРН-У-10. Формулювати, експериментально підтверджувати, обґрунтовувати і застосовувати на практиці нові конкурентоздатні ідеї, методи, технології розв'язку професійних, науково-технічних задач, в тому числі нестандартних;	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+				
ПРН-У-11. Розробляти наукові і інформаційно-освітні ресурси для розв'язання професійних задач, пов'язаних з розвитком та використанням статистики;	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+				
ПРН-У-12. Розуміти сутність інформації, проводити критичну оцінку кількості і змісту інформації	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+				
ПРН-У-14. Прогнозувати розвиток статистики															+	+	+			
ПРН-У-15. Розуміти, аналізувати, цілеспрямовано шукати і вибирати необхідні для рішення професійних наукових задач інформаційно-довідникові та науково-технічні ресурси і джерела знань з урахуванням сучасних досягнень науки і техніки.	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			
ПРН-У-18. Формулювати наукову проблему з огляду на ціннісні орієнтири сучасного суспільства та стан її наукової розробки.	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			
ПРН-У-19. Здійснювати процедуру встановлення інформаційної цінності джерел шляхом порівняльного аналізу з іншими джерелами	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+			
ПРН-У-20. Ініціювання наукових проєктів в галузі комп'ютерних наук та інформаційних систем, лідерство та повна автономність під час їх реалізації																+	+	+		
ПРН-У-23. Оформляти результати досліджень у вигляді статей і доповідей на наукових конференціях.																	+	+	+	
ПРН-У-29. Працювати зі студентською аудиторією в галузі математики, вміти організувати їх навчальний процес																		+	+	+

7. Схема формування оцінки.

7.1. Форми оцінювання здобувачів освітньо-наукового ступеня:

- оцінювання впродовж навчального періоду:

1. Активна робота на лекції, усні відповіді: РН1.1, РН1.2, РН1.3, РН1.4 – 5 балів/3 бали;
2. Виконання завдань, винесених на самостійну роботу: РН2.1, РН2.2 – 5 балів/3 бали;
3. Контрольна робота 1: РН1.1, РН1.2, РН1.3, РН1.4 – 15 балів/9 балів;

4. *Контрольна робота 2*: РН2.1, РН2.2 – 15 балів/9 балів;

6. *Виступ з доповіддю за темою наукового дослідження*: РН2.3, РН3.1, РН4.1, РН4.2, – 20 балів/12 балів;

- **підсумкове оцінювання**: екзамен.

- *максимальна кількість балів які можуть бути отримані*: 40 балів;

- *результати навчання які будуть оцінюватись*: РН1.1, РН1.2, РН1.3, РН1.4, РН2.1, РН2.2;

- *форма проведення і види завдань*: письмова робота.

7.2. Організація оцінювання:

Обов'язковим є виконання завдань, винесених на самостійну роботу, та модульних контрольних робіт за графіком робочої програми.

У частину 1 входять теми 1 - 3, у частину 2 – теми 4 – 6 у частину 3 – теми 7 – 9. Обов'язковим для екзамену є виконання усіх контрольних робіт та доповідь за темою наукового дослідження до вказаної викладачем дати, перед початком екзаменаційної сесії, згідно навчального плану. Переписування чи перескладання тем не практикується..

Терміни проведення форм оцінювання:

1. *Контрольна робота*: до 5 тижня навчального періоду.

2. *Контрольна робота*: до 13 тижня навчального періоду.

3. *Доповідь за темою наукового дослідження*: до 10 тижня навчального періоду.

У випадку відсутності з поважних причин відпрацювання та перездачі контрольні роботи здійснюються у відповідності до „Положення про організацію освітнього процесу”.

7.3. Шкала відповідності оцінок

Відмінно / Excellent	90-100
Добре / Good	75-89
Задовільно / Satisfactory	60-74
Незадовільно / Fail	0-59

При визначенні оцінки визначальною є робота в семестрі. Після завершення розгляду тем проводяться письмові контрольні роботи та теоретичне опитування.

8. СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ. ТЕМАТИЧНИЙ ПЛАН ЛЕКЦІЙ І ПРАКТИЧНИХ ЗАНЯТЬ

№	Назва лекції	Кількість годин		
		Лекції	Практичні	Самостійна робота
Частина 1. „Методи класифікації біологічних даних”				
1	Тема 1. Вступ, основні означення та поняття. Задача статистичної класифікації. Характеризація якості класифікатора. Поняття про баєсів класифікатор. <i>Самостійна робота:</i> Побудова теоретичних баєсових класифікаторів при заданих щільностях розподілу класів та апіорних ймовірностях.	2		8
2	Тема 2. Параметричні і непараметричні оцінки щільності розподілу. Емпірично-баєсова класифікація біологічних даних. <i>Самостійна робота:</i> Застосування емпірично-баєсової класифікації для аналізу даних біологічних досліджень. Лінійний дискримінантний аналіз.	2		12
3	Тема 3. Древа класифікації. Сучасні техніки побудови деревинних класифікаторів. <i>Самостійна робота:</i> Використання алгоритмів CART для побудови класифікаторів біологічних об’єктів..	1	2	8
<i>Контрольна робота 1</i>		1		
Частина 2. „Методи зниження вимірності багатовимірних біологічних даних”				
4	Тема 4. Метод головних компонент та техніки цілеспрямованого проектування у аналізі біологічних даних. <i>Самостійна робота:</i> Застосування методів цілеспрямованого проектування до аналізу даних прикладних біологічних досліджень.	2		12
5	Тема 5. Відстані між біологічними об’єктами. Генетичні відстані і їх обчислення алгоритмами динамічного програмування. <i>Самостійна робота:</i> Реалізація алгоритмів динамічного розрахунку відстаней у R.	2	2	8
6	Тема 6. Метричні і не метричні методи багатовимірного шкалування. <i>Самостійна робота:</i> Застосування багатовимірного шкалування до аналізу даних генетичних досліджень.	2		12
Частина 3. „Класифікація без вчителя для біологічних даних”				
7	Тема 7. Техніка спектральної кластеризації. Лагранжіан та його властивості. Зв’язок між	2		12

	власними векторами лагранжіана і оптимальною кластеризацією. <i>Самостійна робота:</i> Стандартні техніки кластеризації у біостатистиці.			
8	Тема 8. Кластеризація на основі моделі скінченної суміші розподілів. EM-алгоритм та його застосування у задачах кластеризації. <i>Самостійна робота:</i> Реалізація EM-алгоритма у R..	2		12
9	Тема 9. Техніки ієрархічної класифікації злиттям та розбиттям. Їх застосування при побудові філогенетичних дерев. <i>Самостійна робота:</i> Особливості застосування різних методів ієрархічної класифікації для дослідження біологічних об'єктів.	1		12
<i>Контрольна робота 2</i>		1		
ВСЬОГО		18	4	96

Загальний обсяг 120 годин, в тому числі:

Лекцій – **18 годин**,

Практичних занять – **4 години**

Консультації - **2 години**.

Самостійна робота – **96 години**.

9. Рекомендовані джерела

Основні:

1. Gerry P. Quinn, Michael J. Keough - Experimental Design and Data Analysis for Biologists- Cambridge University Press, 2002.- 558p.
2. Gareth James , Daniela Witten , Trevor Hastie , Robert Tibshirani Introduction to Statistical Learning with Applications in R . Springer, 2013 – 764p.
3. Alboukadel Kassambara , Practical Guide To Cluster Analysis in R , Unsupervised Machine Learning , Published by STHDA,2017 – 180p.

Додаткові:

4. Wickham H. Advanced R.CRC Press 2015 – 240 p.
5. Dalgaard P. Introductory Statistics With R. Springer 2008 – 344p.