

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА**

МЕХАНІКО-МАТЕМАТИЧНИЙ ФАКУЛЬТЕТ

Кафедра теорії ймовірностей, статистики та актуарної математики



«ЗАТВЕРДЖУЮ»

Заступник декана
з навчальної роботи

Харитонов О.М.

2021 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Комп'ютерна статистика

для студентів

| | |
|------------------|---|
| галузь знань | 11 «Математика та статистика» |
| спеціальність | 112 «Статистика» |
| освітній рівень | другий (магістр) |
| освітня програма | «Прикладна та теоретична статистика» |
| вид дисципліни | обов'язкова |

| | |
|--|-------------------|
| Форма навчання | денна |
| Навчальний рік | 2021/2022 |
| Семестр | 1 |
| Кількість кредитів ECTS | 5 |
| Мова викладання, навчання та оцінювання | українська |
| Форма заключного контролю | залік |

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

Пролонговано: на 20 /20 н.р. () « » 20 р.
на 20 /20 н.р. () « » 20 р.

КИЇВ – 2021

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

ЗАТВЕДЖЕНО
Зав. кафедри
кафедри теорії ймовірностей,
статистики та актуарної математики
Мішура Ю.С.

Протокол № 1 від 31.08.2021 р.

Схвалено науково-методичною комісією механіко-математичного факультету

Протокол від "31" 08 2021 року № 1

Голова науково-методичної комісії Олійник А.С.
(підпис) професор, д.ф.-м.н. Олійник А.С.

1. Мета дисципліни – оволодіння сучасними методами, теоретичними положеннями та основними застосуваннями сучасної комп'ютерної та математичної статистики, зокрема основними методами регресійного аналізу, оцінювання, перевірки гіпотез, вибору та діагностики моделі, а також реалізації цих методів у програмній системі R.

2. Попередні вимоги до опанування або вибору навчальної дисципліни:

- 1) *Знати:* основи математичного аналізу, лінійної алгебри та теорії ймовірностей.
- 2) *Вміти:* розв'язувати задачі аналітичного диференціювання та інтегрування.
- 3) *Володіти елементарними навичками:* роботи на персональному комп'ютері.

3. Анотація навчальної дисципліни:

Навчальна дисципліна «Комп'ютерна статистика» є складовою освітньої програми підготовки фахівців за освітнім рівнем «магістр» галузі знань 11 Математика та статистика зі спеціальності 112 Статистика освітньої програми «Прикладна та теоретична статистика».

Дана освітня компонента є обов'язковою.

Дисципліна «Комп'ютерна статистика» включає в себе оцінювання коефіцієнтів регресії за методом найменших квадратів, методи перевірки гіпотез про коефіцієнти, аналіз впливу та діагностику моделей регресії, техніки регресії на головні компоненти, рідж-регресії, лассо-регресії, вибір оптимальної специфікації регресійної моделі, аналіз гетероскедастичних моделей. У курсі передбачено поглиблене вивчення комп'ютерної реалізації цих методів а також технологій програмування статистичних алгоритмів з використанням мови R.

Викладається у 1 семестрі в обсязі **150 год. (5 кредитів ECTS¹)** зокрема: *лекції – всього 38 год, практичні заняття – 14 год, самостійна робота – 102 год.* У курсі передбачено 2 змістових модулі та виконання 5 індивідуальних самостійних завдань. Завершується дисципліна заліком у першому семестрі.

4. Завдання (навчальні цілі):

Формування здатності розв'язувати складні спеціалізовані задачі статистичного аналізу та практичні проблеми в галузі дослідження процесів і систем, які мають стохастичну природу, зокрема в економіці, фінансах, медицині, соціології, комп'ютерних науках, інтелектуальному аналізі даних, управлінні та контролі якості, наукових дослідженнях з фізики, хімії, біології; набуття знань, умінь та навичок (компетентностей) на рівні новітніх досягнень у математиці та статистиці, відповідно до освітнього рівня «Магістр». Зокрема, професійне оволодіння компетентностями:

- 1) здатність учитися, здобувати нові знання, уміння, у тому числі в галузях, відмінних від статистики (ЗК-1);
- 2) здатність використовувати у професійній діяльності знання з галузей математичних, природничих, соціально-гуманітарних та економічних наук (ЗК-2);
- 3) здатність вирішувати проблеми у професійній діяльності на основі абстрактного мислення, аналізу, синтезу та прогнозу (ЗК-3);
- 4) здатність до пошуку, оброблення й аналізу інформації з різних джерел, необхідної для розв'язування наукових і професійних завдань (ЗК-4);
- 5) здатність генерувати нові ідеї (ЗК-5);
- 6) здатність спілкуватися державною мовою і усно, і письмово (ЗК-8);
- 7) здатність спілкуватися іноземною мовою (ЗК-9);
- 8) здатність грамотно будувати комунікацію, виходячи з мети і ситуації спілкування (ЗК-10);

¹ кредитів ECTS – кредит кратний 30 годинам.

- 9) здатність критично оцінювати та переосмислювати власний і чужий досвід, аналізувати свою професійну й соціальну діяльність (ЗК-11).
- 10) Знання на рівні новітніх досягнень, необхідні для дослідницької та/або інноваційної діяльності у сфері статистики та її практичних застосувань (ФК-1);
- 11) Спроможність розуміти проблеми та виділяти їхні суттєві риси (ФК-4);
- 12) Спроможність розробляти математичну модель ситуації з реального світу та переносити математичні знання в нематематичні контексти (ФК-5);
- 13) Здатність доводити знання та власні висновки до фахівців та нефхівців (ФК-6);
- 14) Здатність до розвитку нових та удосконалення існуючих статистичних методів аналізу, моделювання, прогнозування, розв'язування нових проблем у нових галузях знань (ФК-8).

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 (Формуються розробником) | | | |
| РН 1.1 | Знати основні методи оцінювання параметрів на основі методу найменших квадратів, методи аналізу впливу та діагностики моделі, та їх реалізацію в R. | <i>Лекція, практичне заняття</i> | <i>Залік, активна робота на лекції, усні відповіді, результати виконання самостійних індивідуальних завдань</i> | 5% |
| РН 1.2 | Знати поняття головних компонент вибірки як напрямів оптимальної проєкції, їхнє застосування для підгонки регресійних моделей і реалізацію в R. | | | 5% |
| РН 1.3 | Знати теорію і техніку рідж- та лассо-оцінювання, вміти реалізовувати ці оцінки в R. | | | 10% |
| РН 1.4 | Знати спектральний та сингулярний розклади матриць, їх використання для побудови псевдооберненої матриці Мура-Пенроуза. | | | 10% |
| РН 1.5 | Знати методи аналізу гетероскедастичних регресійних моделей на основі техніки навантажених найменших квадратів. | | | 10% |
| РН 2.1 | Вміти вибирати оптимальний набір регресорів з використанням критерію Мелуза з використанням R. | <i>Практичне заняття, самостійна робота</i> | <i>Розв'язання задач на практичних заняттях Залік, виконання завдань, винесених на самостійну роботу</i> | 15% |
| РН 2.2 | Вміти використовувати базові функції R для проведення попереднього збору та обробки даних. | | | 15% |
| РН 2.3 | Вміти використовувати техніку крос-валідації для вибору оптимальних параметрів налаштування алгоритмів. | | | <i>Виконання завдань, винесених на самостійну роботу</i> |
| РН 3.1 | Здатність грамотно будувати комунікацію, виходячи з мети і ситуації спілкування | <i>Практичне заняття</i> | <i>активна робота практичних заняттях, усні</i> | 2.5% |

| | | | | |
|--------|--|--------------------------|---|------|
| | | | <i>відповіді</i> | |
| РН 3.2 | Вироблення навиків командної роботи | <i>Практичне заняття</i> | <i>активна робота практичних заняттях, усні відповіді</i> | 2.5% |
| РН 4.1 | Демонстрація авторитетності, інноваційність, високий ступінь самостійності, академічна та професійна доброчесність, послідовна відданість розвитку нових ідей або процесів у передових контекстах професійної та наукової діяльності | <i>Практичне заняття</i> | <i>активна робота практичних заняттях, усні відповіді</i> | 2.5% |
| РН 4.2 | Відповідально ставитися до виконуваних робіт, нести відповідальність за їх якість. | <i>Практичне заняття</i> | <i>активна робота практичних заняттях, усні відповіді</i> | 2.5% |

6. Співвідношення результатів навчання дисципліни із програмними результатами навчання

| Результати навчання дисципліни | Програмні результати навчання | | | | | | | | | | | |
|--|-------------------------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
| | РН 1.1 | РН 1.2 | РН 1.3 | РН 1.4 | РН 1.5 | РН 2.1 | РН 2.2 | РН 2.3 | РН 3.1 | РН 3.2 | РН 4.1 | РН 4.2 |
| ПРН-3-1 - Знати та розуміти фундаментальні і прикладні аспекти наук у сфері статистики | + | + | + | + | + | | | | | | | |
| ПРН-3-2 - Відтворювати знання фундаментальних розділів статистики в обсязі, необхідному для володіння математичним та економічним апаратами відповідної галузі знань і використання статистичних методів у обраній професії | + | + | + | + | + | + | + | + | | | | |
| ПРН-3-3 - Володіти основами математичних дисциплін і економічних теорій, зокрема які вивчають моделі природничих і соціальних процесів | + | + | + | + | + | + | + | + | | | | |
| ПРН-У-1 - Уміти використовувати фундаментальні закономірності статистики у професійній діяльності | | | | | | + | + | + | + | + | + | + |
| ПРН-У-2 - Читати і розуміти фундаментальні розділи математичної та економічної літератури та демонструвати майстерність їх відтворення в аргументованій усній та/або письмовій доповіді | + | + | + | + | | | | | + | | + | + |
| ПРН-У-3 - Доносити професійні знання, | + | + | + | + | + | | | | + | + | + | + |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|--|---|---|
| власні обґрунтування і висновки до фахівців і широкого загалу | | | | | | | | | | | | | |
| ПРН-У-8 - Бути наполегливим у досягненні мети під час вирішення математичної проблеми | + | + | + | + | + | + | + | + | + | + | | + | + |
| ПРН-У-10 - Усно й письмово спілкуватися рідною та англійською мовами в науковій, виробничій та соціально-суспільній сферах діяльності із професійних питань; читати спеціальну літературу; знаходити, аналізувати та використовувати інформацію з різних джерел | + | + | + | + | + | + | + | + | + | + | | | |
| ПРН-У-11 - Використовувати раціональні способи пошуку та використання науково-технічної інформації в галузі статистики, включаючи засоби електронних інформаційних мереж; застосовувати інформаційні ресурси, у тому числі електронні, для пошуку відповідних математичних моделей | + | + | + | + | + | + | + | + | | | | | |

7. Схема формування оцінки.

7.1. Форми оцінювання студентів:

- оцінювання впродовж навчального періоду:

1. Активна робота на лекції, усні відповіді: РН1.1,РН1.2,РН1.3, РН1.4, РН1.5, РН2.1, РН2.2, РН2.3, РН3.1, РН3.2, РН4.1, РН4.2 – 6 балів/5 балів;
 2. Виконання індивідуальних самостійних завдань 1-2: РН1.1, РН1.2, РН3.1 – 22 балів/12 балів;
 3. Виконання індивідуального самостійного завдання 3: РН2.3, РН3.1, РН3.2– 10 балів/5 балів;
 4. Виконання індивідуального самостійного завдання 4-5: РН1.3, РН1.4 – 22 балів/13 балів;
- Разом має бути 60/35

- підсумкове оцінювання: залік.

- максимальна кількість балів, які можуть бути отримані: 40 балів;
- результати навчання, які будуть оцінюватись: РН1.1, РН1.2, РН1.3, РН1.4, РН1.5, РН2.1, РН2.2, РН2.3
- форма проведення і види завдань: письмова робота з усним обговоренням.

7.2. Організація оцінювання:

Критично-розрахунковий мінімум балів за навчання впродовж семестру становить **20** балів, рекомендований мінімум, розрахований з урахуванням специфіки дисципліни становить **35** балів. Студенти, які протягом семестру набрали сумарно меншу кількість балів ніж рекомендований мінімум **35** балів для підвищення балів отримують можливість написати додаткову контрольну роботу та доскласти домашні завдання. Мінімальна кількість балів, які додаються до семестрових – 24 бали, тобто, якщо оцінка студента на іспиті є нижчою від мінімального порогового рівня (24 бали), то бали за залік не додаються до семестрової оцінки (вважаються рівними нулю), а підсумкова оцінка із дисципліни є незадовільною;

У випадку відсутності студента з поважних причин відпрацювання та перездачі форм контролю здійснюються у відповідності до „Положення про організацію освітнього процесу в Київському національному університеті імені Тараса Шевченка” (2018), <http://www.univ.kiev.ua/pdfs/official/Organization-of-the-educational-process.pdf>.

Форма заліку – письмово-усна. Екзаменаційний білет заліку складається із 4 завдань, перші два з яких є теоретичними, одне – розрахункове і одне – по обговоренню результатів виконання індивідуальних самостійних завдань. Кожне завдання оцінюється від 0 до 10 балів. Додатково від 0 до 5 балів студент отримує за додаткові запитання на усному опитуванні. Всього за залік можна отримати від 0 до 40 балів.

Терміни проведення форм оцінювання:

1. Виконання індивідуальних самостійних завдань 1-2: на 1-му - 6-му тижні навчального періоду.

2. Виконання індивідуальних самостійних завдань 3-5: на 7-му - 12-му тижні навчального періоду

7.3. Шкала відповідності оцінок

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

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

| теми | Назва теми | Кількість годин | | | | |
|---|---|-----------------|-------------------|----------------|------------------------|-----------------------|
| | | Лекції | Практичні заняття | Самост. робота | Контр. модульна робота | Інші форми контролю |
| Змістовий модуль 1 „Метод найменших квадратів” | | | | | | |
| 1 | Метод найменших квадратів у лінійній регресії. Застосування до аналізу медико-біологічних даних. | 4 | 2 | 14 | | 1 самостійне завдання |
| 2 | Діагностика моделі та аналіз впливу | 6 | 4 | 18 | | 1 самостійне завдання |
| 3 | Регресія на головні компоненти. застосування у геноміці | 4 | 4 | 14 | | 2 самостійне завдання |
| Змістовий модуль 2 „Оцінювання параметрів і перевірка гіпотез” | | | | | | |
| 4 | Рідж- та лассо-регресія | 4 | 2 | 14 | | 3 самостійне завдання |
| 5 | Техніка крос-валідації | 2 | 2 | 14 | | 3 самостійне завдання |
| 6 | Вибір оптимального набору регресорів | 4 | 2 | 14 | | 4 самостійне завдання |

| | | | | | | |
|---------------------------|--|----|----|-----|--|-----------------------|
| 7 | Навантажений метод найменших квадратів для гетероскедастичних моделей. Гетероскедастичні моделі біологічних явищ. | 4 | 1 | 14 | | 5 самостійне завдання |
| Всього годин за I семестр | | 28 | 14 | 102 | | |

**Загальний обсяг 150 годин, у тому числі:
лекції – 28 годин,
практичні заняття – 14 годин,
консультації – 6 годин,
самостійна робота – 102 години.**

9. Рекомендовані джерела

Основні:

1. Майборода Р.Є. Регресія: Лінійні моделі. К.:ВПЦ «Київський університет», 2007. – 296 с.
2. Майборода Р.Є. Комп'ютерна статистика. К.:ВПЦ «Київський університет», 2019. – 589с.
3. Майборода Р.Є. Самостійні роботи по курсу: Комп'ютерна статистика (для студентів магістратури за спеціальністю статистика). К. 2017. – 16с. Режим доступу <http://probability.univ.kiev.ua/userfiles/mre/tasksCSmag4.pdf>
4. Карташов М.В. Імовірність, процеси, статистика. К.:ВПЦ «Київський університет», 2008. – 504 с.
5. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning & Data Mining, Inference, and Prediction/Springer, 2013. – 764p.

Додаткові:

1. Shao J. Mathematical statistics./ Springer-Verlag, New York, 1998.
2. Себер Дж. Линейный регрессионный анализ. М.: Мир, 1980. – 456с.