

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
ІМЕНІ ТАРАСА ШЕВЧЕНКА
ФАКУЛЬТЕТ КОМП'ЮТЕРНИХ НАУК ТА КІБЕРНЕТИКИ
Кафедра прикладної статистики**



**РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
МАТЕМАТИЧНІ МЕТОДИ ПОБУДОВИ
DATASET/MATHEMATICAL METHODS FOR BUILDING A DATA
SET
для студентів / for students**

галузь знань	12 «Інформаційні технології»/ Information Technologies
спеціальність	122 «Комп'ютерні науки»/ Computer Science
освітній рівень	магістр / Master
освітня програма	«Штучний інтелект» / Artificial Intelligence
вид дисципліни	вибіркова / free choice

Форма навчання	денна
Навчальний рік	2020/2021
Семестр	4
Кількість кредитів ECTS	4
Мова викладання, навчання та оцінювання	англійська, українська/ Ukrainian, English
Форма заключного контролю	іспит / exam

Викладачі: **к.ф.-м.н, доц. Розора І.В.**

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

на 20__/20__ н.р. _____ (_____) «__» 20__р.

КИЇВ – 2020

Розробник: **Розора І.В.** канд. фіз.-мат. н., доцент, доцент кафедри Прикладної Статистики

ЗАТВЕРДЖЕНО

Зав. кафедри Прикладної Статистики

 (Лебедев Є.О.)

Протокол № 1 від «28» 08 2020 р.

Схвалено Гарантом освітньо-наукової програми «Штучний інтелект»

 (Крак Ю.В.)

«28» 08 2020 р.

Схвалено науково-методичною комісією факультету комп'ютерних наук та кібернетики

Протокол від «28» серпня 2020 року № 1

Голова науково-методичної комісії  (Омельчук Л.Л.)
(підпис) (прізвище та ініціали)

«28» серпня 2020 року

1. Мета дисципліни – ознайомлення студентів з сучасними математичними методами побудови Data Set на основі вибірових методів досліджень та практична підготовка студентів до проведення статистичних досліджень, їх аналізу та обробки.

Discipline aim. The purpose of the discipline "Mathematical methods for building a Data Set", acquaintance of students with modern mathematical methods of construction of the Data Set on the basis of sampling methods of researches and practical preparation of students for carrying out statistical researches, their analysis and processing..

2 Попередні вимоги до опанування або вибору навчальної дисципліни

Знати: основи дискретної математики, теорії ймовірностей та математичної статистики

Вміти: формалізувати умови задач та складати план розв'язку

Володіти елементарними навичками: розв'язувати типові задачі з теорії ймовірностей, математичної статистики та дискретної математики.

Preliminary demands to master or choice of the course discipline:

Know: basics of discrete mathematics, probability theory and mathematical statistics

Be able to: formalize the conditions of problems and make a solution plan

Have basic skills: solve typical problems in probability theory, mathematical statistics and discrete mathematics.

3 Анотація навчальної дисципліни

Дисципліна «Математичні методи побудови DataSet» є складовою частиною циклу професійної підготовки фахівців освітньо-кваліфікаційного рівня „магістр”; вона включає вивчення таких понять, як вибірові обстеження, вибірка для побудови Dataset, вибіровий дизайн; вивчення методів вибірових відборів. Особлива увага приділяється основним видам випадкових відборів та їх властивостям. Викладається у 4-му семестрі, обсяг 120 год. (4 кредити ECTS), з них лекції – 14 год., лабораторних – 10 год. консультації – 4 год., самостійна робота – 92 год. Передбачено 1 частину та іспит.

Synopsis of the course:

The discipline "Mathematical methods of building a DataSet" is an integral part of the cycle of professional training of specialists of educational and qualification level "master"; it includes the study of such concepts as sample surveys, sampling to build a dataset, sample design; study of sampling methods. Particular attention is paid to the main types of random selection and their properties. Taught in the 4th semester, volume 120 hours. (4 ECTS credits), of which lectures - 14 hours, laboratory - 10 hours. consultations - 4 hours, independent work - 92 hours. Part 1 and exam are provided.

4 Завдання (навчальні цілі)

Набуття знань, умінь та навичок (компетентностей) на основі використання системного підходу та його принципів для створення інтелектуальних бізнес систем, на основі сучасних інформаційних та комп'ютерних технологій, моделей і методів штучного інтелекту. Зокрема, розвивати:

- Здатність спілкуватися іноземною мовою.
- Здатність аналізувати математичні методи побудови DataSet.

Objectives of study:

Acquisition of knowledge, skills and abilities (competencies) based on the use of a systems approach and its principles for creating intelligent business systems, based on modern information and computer technologies, models and methods of artificial intelligence. In particular, to develop:

- Ability to communicate in a foreign language.
- Ability to analyze mathematical methods for constructing DataSet.

5 Результати навчання за дисципліною

Результат навчання (РН) (1 – знати; 2 – вміти; 3. комунікація; 4. автономність та відповідальність)		Форми викладання та навчання	Методи оцінювання	Відсоток у підсумковій оцінці з дисципліни
Код	Результат навчання			
РН1.1	Знати і розуміти основні методи побудови DataSet з використанням різних методів відборів / Know and understand the basic methods of building a DataSet using different sampling methods	Лекції, лабораторні заняття, самостійна робота / Lecture, Lab.work, Individual work	Поточне оцінювання (ПО) контрольна робота, іспит / Current assessment (PO) test, exam	40%
РН2.1	Вміти будувати DataSet з використанням різних методів відборів. / Be able to build a DataSet using different sampling methods.			
РН3.1	Демонструвати навички взаємодії з іншими людьми, уміння працювати в командах / Demonstrate skills of interaction with other people, ability to work in teams	Лекції, лабораторні заняття, самостійна робота / Lecture, Lab.work, Individual work	Поточне оцінювання (ПО) контрольна робота / Current assessment (PO) test	
РН.4.1	Виявляти здатність до самонавчання та продовження професійного розвитку / Demonstrate the ability to self-study and continue professional development	Лекції, лабораторні заняття, самостійна робота / Lecture, Lab.work, Individual work	Поточне оцінювання (ПО) контрольна робота, іспит / Current assessment (PO) test, exam	50%
РН.4.2	Уміти організувати власну діяльність та одержувати результат у рамках обмеженого часу / Be able to organize their own activities and get results within a limited time	Лекції, лабораторні заняття, самостійна робота / Lecture, Lab.work, Individual work	Поточне оцінювання (ПО) контрольна робота / Current assessment (PO) test	10%

6 Співвідношення результатів навчання дисципліни з програмними результатами навчання

Програмні результати навчання	Результати навчання дисципліни				
	РН1.1	РН2.1	РН3.1	РН4.1	РН4.2
(з опису освітньої програми)					

ПРН20.2 Володіти математичними методами побудови DataSet.	+	+	+	+	+
---	---	---	---	---	---

7 Схеми формування оцінки/ Evaluation scheme

7.1 Форми оцінювання студентів: / Forms of evaluation:

- семестрове оцінювання: / semester evaluation:

1. Контрольна робота / Test (PH1.1, PH2.1, PH3.1, PH4.2): 30 балів (points) / 18 балів (points).
2. Поточне оцінювання / Current assessment (PH1.1, PH2.1, PH3.1, PH4.2): 30 балів (points) / 18 балів (points).

Підсумкове оцінювання / final evaluation: іспит /exam:

- максимальна кількість балів які можуть бути отримані студентом: 40;
- результати навчання, які оцінюються: PH1.1, PH2.1, PH.4.1;
- форма проведення: письмова робота
- види завдань: два теоретичні питання (40%), дві задачі (60%).

Студент допускається до іспиту, якщо в семестрі набрав не менше ніж 36 балів. Для отримання загальної позитивної оцінки з дисципліни оцінка за екзамен має бути не менше 24 балів.

- the maximum number of points that can be obtained by a student: 40 points;
- learning outcomes that are evaluated: PH1.1, PH2.1, PH.4.1;
- form of conducting and types of tasks: written work;
- types of tasks: two theoretical questions (40%), two tasks (60%).

A student is admitted to the exam if he scored at least 36 points in the semester. To receive an overall positive grade in the discipline, the grade for the exam must be at least 24 points.

Контрольні запитання до частини I

1. Основні поняття: генеральна сукупність, вибірка, схеми відбору
2. Вибірковий дизайн для побудови DataSet, приклади.
3. Ймовірності включення елемента у вибірку (DataSet).
4. Індикатори включення елемента у вибірку(DataSet), їх властивості.
5. Розмір DataSet, як випадкова величина. Його середнє та дисперсія.
6. Оцінка Горвіца-Томпсона та її властивості.
7. Форма Єйтса-Гранді-Сена.
8. Простий випадковий відбір без повернення (ПВВБП)
9. Оцінка Горвіца-Томпсона при ПВВБП
10. Оцінка параметрів підсукупностей при ПВВБП
11. Побудова довірчих інтервалів при ПВВБП
12. Визначення розміру DataSet
13. Відбір Бернуллі: Оцінка Г-Т.
14. Дизайн-ефект
15. Систематичний відбір
16. Розмір вибірки при систематичному відборі.
17. Ефективність систематичного відбору
18. Міри однордності
19. Оцінювання дисперсії при систематичному відборі.
20. Відбір з поверненням: основні відмінності
21. Оцінка Хансена-Гурвіца при відборі з поверненням

Test questions for part I

1. Basic concepts: population, sample, sampling schemes
2. Sampling design for building a DataSet, examples.
3. Probabilities of inclusion of an element in the sample (DataSet).
4. Indicators of inclusion of an element in the sample (DataSet), their properties.
5. DataSet size as a random variable. Its mean and variance.
6. Horwitz-Thompson Estimator and its properties.
7. The Sen-Yates-Grundy theorem.
8. Simple random sampling without replacement (SRS)
9. Horwitz-Thompson Estimator at SRS
10. Estimation of domain parameters at SRS
11. Construction of confidence intervals in SRS
12. Determining the size of the DataSet
13. Bernoulli sampling: Horwitz-Thompson Estimator.
14. Design effect
15. Systematic sampling
16. Sample size for systematic sampling.
17. The effectiveness of systematic sampling
18. Measures of homogeneity in systematic sampling
19. Estimation of variance in systematic sampling.
20. Simple random sampling with replacement (SRSwR): the main differences
21. Hansen-Hurwitz estimator in SRSwR

7.2 Організація оцінювання

Терміни проведення оцінювання

Контрольна № 1 – до 7 тижня.

Поточне оцінювання – протягом семестру.

За відсутності студента з поважних причин передача екзамену здійснюється відповідно до «Положення про порядок оцінювання знань студентів при кредитно-модульній системі організації навчального процесу» від 1 жовтня 2010 року.

7.2 Assessment

Terms of assessment

Test № 1 - up to 7 weeks.

Formative assessment - during the semester.

In the absence of a student for valid reasons, the reexamination is carried out in accordance with the "Regulations on the procedure for assessing students' knowledge in the credit-module system of the educational process" from October 1, 2010.

7.3 Шкала відповідності оцінок / Rating scale

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

8. Структура навчальної дисципліни. Тематичний план лекцій і лабораторних занять / The structure of the discipline. Thematic plan of lectures and laboratory classes

№	Назва лекції	Кількість годин		
		Лекції	Лабора-Торні	Самост. Робота
Змістовий модуль 1.				
1	<p>Тема 1. Вступ. Основні поняття: генеральна сукупність, вибірка- DataSet, схеми відбору, вибірковий дизайн/ Introduction. Basic conseps: population, sample-Dataset, sampling schemes, sampling design</p> <p><i>Самостійна робота/ individual work:</i> Вступ. Схеми відбору, вибірковий дизайн/ sampling schemes, sampling design</p>	2		15
2	<p>Тема 2. Індикатори включення елемента у DataSet, їх властивості./ Indicators of inclusion of an element in the sample (DataSet), their properties.</p> <p><i>Самостійна робота/ individual work:</i> Ймовірності включення елемента у вибірку (DataSet)./ Probabilities of inclusion of an element in the sample (DataSet).</p>	2	2	15
3	<p>Тема 3. Розмір DataSet, як випадкова величина. Його середнє та дисперсія./ DataSet size as a random variable. Its mean and variance.</p> <p><i>Самостійна робота/ individual work:</i> Основні співвідношення для випадкового відбору зі сталим розміром вибірки/ The main relationships in random sampling with fixed sample size</p>	2		10
4	<p>Тема 4. Оцінка Горвіца-Томпсона та її властивості. Форма Єйтса-Гранді-Сена./ Horwitz-Thompson Estimator and its properties. The Sen-Yates-Grundy theorem.</p> <p><i>Самостійна робота individual work:</i> Незміщеність та конзистентність оцінки Горвіца-Томпсона./ Unbiasedness and consistency of Horwitz-Thompson Estima</p>	2	2	15
5	<p>Тема 5. Простий випадковий відбір без повернення (ПВВБП). Оцінка параметрів підсукупностей при ПВВБП / Simple random sampling(SRS). Estimation of domain parameters in SRS</p> <p><i>Самостійна робота/ individual work:</i> Оцінка параметрів підсукупностей при ПВВБП. Побудова довірчих інтервалів / Estimation of</p>	2	2	15

	domain parameters in SRS. Construction of confidence intervals in SRS.			
6	Тема 6. Відбір Бернуллі(ББ): Оцінка Г-Т. Дизайн ефект./ Bernoulli sampling(BS): Horwitz-Thompson Esimator. Design effect. <i>Самостійна робота / individual work:</i> Визначення розміру DataSet при ББ. Простий випадковий відбір з повторенням. Оцінка Хансена-Гурвіца./ Determining the size of the DataSet in BS. Simple random sampling with replacement (SRSwR). Hansen-Hurwitz estimator in SRSwR	2	2	11
7	Тема 7. Систематичний відбір/ Systematic Sampling <i>Самостійна робота/ individual work:</i> Розмір DataSet при систематичному відборі. Ефективність при систематичному відборі. Міри однорідності/ Sample size for systematic sampling. The effectiveness of systematic sampling. Measures of homogeneity in systematic sampling	2	1	11
	Контрольна робота		1	
Іспит				
ВСЬОГО		14	10	92

Загальний обсяг 120 год.¹, в тому числі:

Лекцій – 14 год.

Лабораторні – 10 год.

Самостійна робота - 92 год.

Консультації – 4 год.

9. Рекомендовані джерела / Literature

1. О. І. Василик, Т. О. Яковенко, Лекції з теорії і методів вибіркового обстеження : навчальний посібник. – К. : Видавничо-поліграфічний центр "Київський університет", 2010. – 208 с.
2. T. Connolly, C. Begg, Database Solutions: A step by step guide to building databases (2nd Edition) 2nd Edition, Addison-Wesley;2003, 552 p.
3. Ardilly, P. Tille Y. Sampling Methods. Exercises and Solutions, Springer Science+Business Media Inc., 2006.
4. Kish, L. Survey Sampling, Wiley, 1995.
5. Lohr, S. Sampling: design and analysis, New York : Duxbury Press, 1999.
6. Sarndal, C.-E. ,B. Swensson, J. Wretman, Model Assisted Survey Sampling, New York : Springer-Verlag, 1992.

¹ Загальна кількість годин, відведених на дану дисципліну згідно навчального плану.