

**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ ТОРГОВЕЛЬНО-ЕКОНОМІЧНИЙ
УНІВЕРСИТЕТ**
СИСТЕМА УПРАВЛІННЯ ЯКІСТЮ
Система забезпечення якості освітньої діяльності та якості вищої освіти
сертифікована на відповідність ДСТУ ISO 9001:2015 / ISO 9001:2015
Кафедра вищої та прикладної математики

СИЛАБУС

**МАТЕМАТИЧНІ МЕТОДИ ОБРОБКИ
СОЦІОЛОГІЧНИХ ДАНИХ /
MATHEMATICAL METHODS OF PROCESSING
SOCIOLOGICAL DATA**

SYLLABUS

освітній ступінь	бакалавр / bachelor
галузь знань	05 Соціальні та поведінкові науки / Social and behavioral sciences
спеціальність	054 Соціологія / Sociology
спеціалізація	Соціологія економічної діяльності (Sociology of Economic Activity)

Автор: Карташова С.С., професор кафедри вищої та прикладної математики

Силабус розглянуто і затверджено на засіданні кафедри
(протокол № 1 від 25 вересня 2022 р.).

Київ 2022

МАТЕМАТИЧНІ МЕТОДИ ОБРОБКИ СОЦІОЛОГІЧНИХ ДАНИХ

Метою викладання навчальної дисципліни «Математичні методи обробки соціологічних даних» є засвоєння та опанування основними математико-статистичними методами, які застосовуються для коректної математичної обробки та аналізу даних вибірових спостережень, зокрема, результатів соціологічного експерименту для отримання достовірних статистичних висновків.

Предметом вивчення дисципліни є кількісний аналіз, що ґрунтується на засадах теорії ймовірностей та математичної статистики, як базового етапу емпіричного соціологічного дослідження, який дає можливість в рамках визначеної кількісної моделі описувати стан соціальних об'єктів, пояснювати перебіг соціальних процесів, перевіряти формально висунуті гіпотези дослідження щодо генеральної сукупності за результатами вибірових спостережень.

В результаті вивчення дисципліни студент повинен:

знати:	вміти:
<ul style="list-style-type: none">- класичний, геометричний, статистичний підходи до обчислення ймовірностей випадкових подій;- сумісні та несумісні, залежні та незалежні випадкові події, формули повної ймовірності та Байєсса;- закони розподілу ймовірностей випадкових величин дискретного, неперервного та абсолютно неперервного типу;- найважливіші дискретні та абсолютно неперервні розподіли: Бернуллі, біноміальний, Пуассона, геометричний, рівномірний, експоненціальний, гамма, нормальний, Хі-квадрат, Стьюдента, Фішера;- числові характеристики випадкових величин та методи їх обчислення;- методологічні засади використання математико-	<ul style="list-style-type: none">- сприймати нову інформацію, аналізувати і узагальнювати, конкретизувати поставлену мету і вибирати шляхи її досягнення;- використовувати правила обробки баз даних, отриманих в результаті прикладних стохастичних експериментів;- формулювати прикладні завдання згідно з теоретико-ймовірнісним підходом та застосовувати адекватні методи для їх вирішення;- володіти інструментарієм для вирішення прикладних математико-статистичних завдань у своїй предметно-фаховій області;- проводити первинну статистичну обробку результатів соціологічного спостереження в тому числі їх зведення та групування;- проводити аналіз варіації,

<p>статистичних методів в обробці даних емпіричних соціологічних досліджень;</p> <ul style="list-style-type: none"> - основні поняття обробки даних із використанням математичних методів; - типологію та властивості шкал вимірювання; - форми представлення експериментальних даних; - основні методи побудови оцінок невідомих параметрів та перевірки статистичних гіпотез; - модель лінійної регресії та її застосування до аналізу даних емпіричних соціологічних досліджень; - можливості та обмеження застосування різних математичних методів до різних типів даних. 	<p>будувати ряди розподілу, та давати характеристику форми розподілу;</p> <ul style="list-style-type: none"> - розрахувати числові характеристики для заданої вибіркової сукупності; - використовуючи закони розподілу випадкової величини, обчислювати її числові характеристики; - вміти використовувати основні елементи теорії ймовірностей на практиці при вирішенні конкретних соціологічних задач; - проводити кореляційний аналіз даних, у тому числі кількісних, порядкових та номінальних даних, а також даних, що виміряні у різних шкалах; - інтерпретувати результати, отримані в процесі обробці даних, та робити обґрунтовані висновки, необхідні для емпіричного соціологічного дослідження. - самостійно опрацьовувати спеціальну літературу з теорії ймовірностей та статистики.
---	--

Викладання дисципліни передбачає проведення лекційних і практичних занять. Важливе місце відведено самостійній роботі студентів з вітчизняною та закордонною спеціальною літературою і виконанню студентами індивідуальних домашніх завдань, інтерактивним засобам навчання. Визначена загальна кількість годин/кредитів становить 180/6 з підсумковим контролем у формі екзамену.

Пререквізити: вивчення дисципліни базується на знаннях дисциплін «Алгебра», «Геометрія» базового рівня середньоосвітньої школи та дисципліни «Methodology and organization of sociological research».

Обсяг: 6 кредитів

Мова викладання: українська

Форма викладання: очна, заочна, дистанційна

Форми поточного контролю:

Модульні контрольні роботи (35+35= 70 балів) + індивідуальні домашні завдання + активна участь на заняттях(15 балів) + участь в науковій роботі (5 балів) = 100 балів

Форми підсумкового контролю: письмовий іспит – 100 балів (в т.ч. 5 задач по 20 балів)

Політика академічної доброчесності:

<https://knute.edu.ua/file/NjY4NQ==/bf27ad9293fa2bb6f9b2c3031d4b6e4a.pdf>

Список рекомендованих джерел

1. Найко Д.А. Шевчук О. Ф. Теорія ймовірностей та математична статистика: навч. посіб. Вінниця: ВНАУ, 2020. 382 с.
2. Паніотто В.І., Максименко В.С., Харченко Н.М. Статистичний аналіз соціологічних даних. К.: Видавничий дім “КМ Академія”, 2004. 270 с.
3. Телейко А. Б., Чорней Р. К. Математико-статистичні методи в соціології та психології: навч. посіб. К. : МАУП, 2007. 424 с.
4. Healey J.F. The Essentials of Statistics: A Tool for Social Research. Boston: Wadsworth Publishing, 2015. 452 p.
5. Levine D.M., Stephan D.F., Szabat K.A. Statistics for Managers Using Microsoft Excel. Pearson, 2021. 753 p.
6. Pollard J. H.. A Handbook of Numerical and Statistical Techniques: With Examples Mainly from the Life Sciences. Cambridge University Press. 1979. 368 p.
7. Upton J. G. Categorical data analysis by example. John Wiley & Sons, 2017. 216 p.

Викладач: Карташова Світлана Степанівна

доктор біологічних наук,

кандидат фізико-математичних наук, доцент, с.н.с.

професор кафедри вищої та прикладної математики

E-mail: s.kartashova@knute.edu.ua

Тел.: 050-351-89-03

Адреса: Київ, вул. Раєвського 36, ауд. Р-511

СТРУКТУРА ДИСЦИПЛІНИ ТА РОЗПОДІЛ ГОДИН ЗА ТЕМАМИ (ТЕМАТИЧНИЙ ПЛАН)

. Структура дисципліни та розподіл годин за темами.

Тема	Кількість годин з них				Форми контролю
	Усього годин/ кредитів	Лекції	Практичні заняття/модульні й контроль	Самостійна робота	
Тема 1. Вступ: загальна характеристика результатів спостережень в соціологічних дослідженнях та методів їх обробки. Пакети прикладних статистичних програм: «Аналіз даних» (MS EXCEL), EPI INFO, SPSS.	2,5	2	0,5		О
Тема 2. Типи соціологічних даних. Поняття про шкали вимірювання. Візуалізація даних спостережень.	9,5	2	1,5	6	О
Тема 3. Аналітична модель соціологічного дослідження: ймовірно-статистичний підхід	16	6	6	4	О, ПСР
Тема 4. Випадкова величина як ймовірнісна характеристика даних соціологічних спостережень. Системи випадкових величин	24	8	8	8	О, ПСР, АКР
Тема 5. Основні поняття та описові методи математико - статистичної обробки соціологічних даних.	10	2	2	6	О, ПСР
Тема 6. Точкові оцінки параметрів розподілів соціологічних даних	8	2	2	4	О, ПСР
Тема 7. Інтервальні оцінки. Визначення мінімально достатнього обсягу вибіркового соціологічного дослідження.	14	4	4	6	О, ПСР
Тема 8. Загальні принципи та методи перевірки статистичних гіпотез.	7	2	2	3	О, ПСР
Тема 9. Параметричні критерії перевірки гіпотез узгодженості, центру розподілу та варіації в соціологічних даних.	17	6	6	5	О, ПСР

Тема 10. Непараметричні критерії перевірки гіпотез про відмінності рівнів, наявності змін та значущих тенденцій за фактором впливу	12	4	4	4	О, ПСП
Тема 11. Елементи дисперсійного аналізу	12	4	4	4	О, ПСП
Тема 12. Параметричний та непараметричний кореляційний аналіз категорійних і числових соціологічних даних.	28	10	10	8	О, ПСП
Тема 13. Стандартний та стратифікований аналіз таблиць сполучених ознак.	8	2	2	4	О, ПСП
Тема 14. Лінійні моделі регресійного аналізу	12	2	4	6	О, ПСП, АКР
Разом	180/6	56	56	68	
Підсумковий контроль - екзамен					

Умовні позначення: О – опитування; ПСП – перевірка самостійної роботи; АКР – аудиторна контрольна робота

Зміст тем дисципліни «Математичні методи обробки соціологічних даних»

Тема 1. Загальна характеристика результатів соціологічних спостережень та кількісних методи їх обробки

Мета і завдання дисципліни, її місце в системі підготовки фахівців зі спеціальності «соціологія». Об'єкти досліджень та основні групи задач обробки соціологічних даних, класифікація напрямків та методів розв'язання. Шляхи та умови застосування математичних методів в соціологічних дослідженнях. Презентація пакетів прикладних статистичних програм обробки та аналізу соціологічних даних: «Аналіз даних» (MS EXCEL), EPI INFO, SPSS.

Тема 2. Типи соціологічних даних. Поняття про шкали вимірювання. Візуалізація даних спостережень.

Вимірювання як передумова застосування математичних методів у соціології. Структура та чотири основні способи класифікації наборів соціологічних даних. Дані за типом вимірювання: категорійні (якісні) та кількісні (числові) дані. Номінальна, рангова, інтервальна, відносні, абсолютна шкала вимірювання. Основні характеристики. Візуалізація даних: стандартні графіки та діаграми.

Тема 3. Аналітична модель соціологічного дослідження: ймовірнісно-статистичний підхід. Математична модель стохастичного експерименту: роль та місце при дослідженні соціологічних даних. Класичний, геометричний та статистичний підходи до визначення ймовірностей. Основне правило

комбінаторики та теорема суми. Обчислення класичних ймовірностей з використанням функцій COMBIN та FACT в середовищі MS Excel. Умовна ймовірність та поняття про незалежність подій. Формули повної ймовірності та Байєса. Модель повторних незалежних випробувань. Формула Бернуллі. Застосування теореми про найімовірніше число «успіхів» в соціологічних дослідженнях.

Тема 4. Випадкова величина як ймовірнісна характеристика даних соціологічних спостережень. Системи випадкових величин Означення випадкових величин та їх класифікація. Функція розподілу та її властивості. Числові характеристики центру розподілу та розсіяння: початкові та центральні моменти, мода, медіана, асиметрія, ексцес.

Закон розподілу дискретної випадкової величини. Таблиця та полігон розподілу, їх представлення в середовищі MS Excel. Основні закони дискретних розподілів (вироджений, гіпергеометричний, Бернуллі, звичайний та від'ємний біноміальні, Пуассона, геометричний) формальне визначення, числові характеристики, прикладне застосування та графічне представлення в MS Excel.

Щільність та числові характеристики абсолютно неперервних випадкових величин, що використовуються при соціологічних дослідженнях. Рівномірний закон розподілу та його застосування в теорії помилок. Експоненціальний закон розподілу: властивість відсутності післядії, найпростіший потік подій. Нормальний закон розподілу ймовірностей, його стандартне представлення та правило трьох σ . Логарифмічно-нормальний розподіл. Розподіли хі-квадрат, Стюдента та Фішера та їх застосування при побудові критеріальних статистик.

Закони розподілу та числові характеристики системи випадкових величин: безумовні, умовні, маргінальні. Коваріація та коефіцієнт кореляції як міра залежності між випадковими величинами: визначення, властивості, обчислення. Випадкові вектори та послідовності. Двомірний нормальний розподіл, його числові характеристики та властивості. Закони великих чисел, умови їх виконання та значення при побудові точкових статистичних оцінок. Центральна гранична теорема як теоретична база для побудови інтервальних статистичних оцінок.

Тема 5. Основні поняття та описові методи математико - статистичної обробки соціологічних даних. Стандартні підходи до означення генеральної сукупності. Основні положення вибіркового дослідження та методи формування випадкової вибірки. Об'єм вибірки та її репрезентативність. Емпірична функція та гістограма. Вибіркові моменти та їх властивості.

Описові статистики центру розподілу та варіації даних (середнє, стандартна помилка, медіана, мода, стандартне відхилення, дисперсія вибірки, ексцес, асиметричність), їх обчислення засобами пакету «Аналіз даних» в середовищі MS Office/ Excel.

Тема 6. Точкові оцінки параметрів розподілів соціологічних даних.

Визначення незміщеної, консистентної, ефективної статистичної оцінки та встановлення її асимптотичної поведінки. Методи знаходження оцінок: метод моментів (ММ) та метод максимальної вірогідності (ММВ). Приклади побудови за результатами проведених спостережень. Порівняння точкових оцінок.

Тема 7. Інтервальні оцінки. Визначення мінімально достатнього обсягу вибіркового соціологічного дослідження. Інтервальні оцінки (довірчі інтервали, ДІ) за рівнем значущості: означення та загальний алгоритм їх побудови. ДІ для генеральних значень середнього (при відомому/невідомому стандартному відхиленні) та дисперсії (при відомому/невідомому математичного очікуванні). Мінімально достатній обсяг вибіркового соціологічного дослідження для оцінки генеральних характеристик досліджуваних ознак.

Тема 8. Загальні принципи та методи перевірки статистичних гіпотез

Статистичні гіпотези в соціологічних дослідженнях. Нульова та альтернативна гіпотези. Поняття про рівень значущості статистичного висновку. Тестова (критеріальна) статистика та її критичне значення. Області відхилення та прийняття гіпотез. Типи помилок та потужність критерію. Ризики, що виникають при перевірці гіпотез, пов'язані з помилками першого та другого роду. Потенційні проблеми та етичні питання, що пов'язані з перевіркою гіпотез.

Тема 9. Параметричні критерії перевірки гіпотез узгодженості, центру розподілу та варіації в соціологічних даних.

Критерії порівнянь емпіричного (вибіркового) розподілу з теоретичним. Порівняння декількох емпіричних розподілів. Критерій узгодженості «хі-квадрат» для розподілу Пуассона, біноміального (поліноміального) та нормального розподілу. Критерій узгодженості Колмогорова - Смірнова для декількох емпіричних розподілів.

Критерії перевірки значень частки, середнього та варіації досліджуваної ознак (випадок однієї вибірки): $N_{(0,1)}$ – критерій, t - критерій Стьюдента, χ^2 – критерій.

Тема 10. Непараметричні критерії перевірки гіпотез про відмінності рівнів, наявності змін та значущих тенденцій за фактором впливу

Критерії для оцінки різниці між вибірками за рівнем кількісно вимірної ознаки:

- Q – критерій Розенбаума;
- U – критерій Манна - Уїтні;
- H – критерій Крускала–Уолліса;
- S – критерій тенденцій Джонкіра.

Критерії розпізнавання зсуву та напрямку змін:

- G – критерій знаків;
- T – критерій Уїлкоксона;

- χ_r^2 – критерій Фрідмана;
- L – критерій тенденцій Пейджа.

Тема 11. Елементи дисперсійного аналізу. Однофакторний дисперсійний аналіз. Мета та необхідні умови проведення дисперсійного аналізу за статистичними даними вибіркових соціологічних спостережень. Множинні порівняння при дії одного фактору впливу. Використання F–критерію Фішера для оцінки різниці між декількома математичними очікуваннями (генеральними середніми). Критерій Левине для перевірки однорідності дисперсій.

Двохфакторний дисперсійний аналіз: оцінка факторів впливу та ефектів їх взаємодії. Інтерпретація ефектів взаємодії. Множинні порівняння при дії декількох факторів впливу.

Тема 12. Параметричний та непараметричний кореляційний аналіз категорійних і числових соціологічних даних. Кореляційний зв'язок та його статистичне вивчення. Завдання кореляційного аналізу. Статистична значущість та теоретичні засади. Основні характеристики: існування, силу, структуру та напрям. Лінійний коефіцієнт кореляції Пірсона та його дослідження.

Статистична залежність між змінним, вимірними за номінальною шкалою. Міри зв'язку на основі критерію хі-квадрат. Пропорційне зменшення помилки. Міра пропорційного зменшення помилки для номінальних змінних лямбда (λ).

Зв'язок між змінним, вимірними за порядковою шкалою. Пропорційне зменшення помилки (PRE). Обчислення коефіцієнта гамма (γ). Визначення напрямку зв'язку. Інтерпретація статистичного зв'язку за допомогою таблиць спряженості на прикладі участі громадян у суспільному житті. Коефіцієнти рангової кореляції Спірмена (ρ_{sp}) та Кендела (τ). Критерії перевірки на значущість статистичного зв'язку між двома ознаками, порядкове вимірними, за коефіцієнтами γ та ρ_{sp} .

Взаємозв'язок між змінним, вимірними за інтервальною чи відносною шкалою. Діаграма розсіювання. Коефіцієнт кореляції Пірсона (r_p) та практичний приклад обчислення за допомогою спеціальних функцій в середовищі MS Office/Excel. Критерій перевірки коефіцієнт кореляції Пірсона (r_p) на значущість.

Тема 13. Стандартний та стратифікований аналіз таблиць сполучених ознак. Принципи побудови стратифікованих таблиць сполучених ознак IxJ. Таблиці 2x2: точний критерій Фішера, точкові та інтервальні оцінки відношення переваг (OR) та відносного ризику (RR). Стратифікований аналіз таблиць сполучених ознак 2x2: точкові та інтервальні оцінки сумарних показників OR та RR. Застосування пакету статистичних програм EPI INFO при розв'язанні типових задач в соціологічних дослідженнях.

Тема 14. Лінійні моделі регресійного аналізу. Статистична модель парної лінійної регресії. Метод найменших квадратів: точкові оцінки

параметрів моделі. Інтервальні оцінки параметрів моделі та перевірка гіпотез про їх значення. Критерії значущості лінії регресії. Коефіцієнт детермінації. Таблиця дисперсійного аналізу парної лінійної регресії. Перевірка регресійної моделі на адекватність даним спостережень. Проблема автокореляції та критерій Дарбіна-Уотсона. Застосування пакету «Аналіз даних» (MS EXCEL) для проведення парного регресійного аналізу даних соціологічних досліджень. Поняття про множинну лінійну регресійну модель. Мультиколінійність. Гетероскедастичність. Регресія та прогноз. Небезпеки, застереження та рекомендації, пов'язані з побудовою прогнозу.