|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **МЕТОДИ БАГАТОВИМІРНОГО АНАЛІЗУ ТА BIGDATA В СОЦІОЛОГІЇ** СИЛАБУС | | | | | | | | | | | | |
| **Шифр і назва спеціальності** | | | **054 – Соціологія** | | | **Інститут / факультет** | | | **Інститут соціально-гуманітарних технологій** | | | |
| **Назва програми** | | | **Соціологічне забезпечення економічної діяльності** | | | **Кафедра** | | | **Соціології і публічного управління** | | | |
| **Тип програми** | | | **Освітньо-професійна** | | | **Мова навчання** | | | **Українська** | | | |
| **Викладач** | | | | | | | | | | | | |
| **Бірюкова Марина Василівна** | | | | | | | **Маryna.Biriukova@kpi.edu.ua** | | | | | |
| http://web.kpi.kharkov.ua/sp/wp-content/uploads/sites/95/2021/12/Biryukova-MV.png | | Доктор соціологічних наук, доцент, професор кафедри соціології і публічного управління. Досвід роботи – 34 роки  Лектор з дисциплін: «Математичні методи в соціології», «Практикум з аналізу соціологічних даних», «Комп’ютерні технології організації соціологічних дисциплін», «Технології соціального проектування», «Методи багатомірного аналізу соціологічних даних». | | | | | | | | | | |
| **Загальна інформація про курс** | | | | | | | | | | | | |
| **Анотація** | | Основними завданнями курсу є: вивчення методів наукових досліджень з теорії організації вибіркових спостережень, обробки та аналізу отриманої інформації, застосування багатовимірних методів та bigdata для соціального аналізу, ідентифікації та розпізнавання образів; моделювання і прогнозування соціальних процесів; використання інформаційних технологій для статистичного обгрунтування прийняття рішень при соціологічному забезпеченні економічної діяльності. | | | | | | | | | | |
| **Цілі курсу** | | освоєння методологічних і методичних основ використовування методів багатовимірного аналізу та bigdata для дослідження природи соціальних явищ, для побудови багатовимірних моделей існування та функціонування соціальних об'єктів. | | | | | | | | | | |
| **Формат** | | Лекції, практичні заняття, консультації. Підсумковий контроль – іспит. | | | | | | | | | | |
| **Семестр** | | 2 | | | | | | | | | | |
| **Обсяг (кредити) / Тип курсу** | 6 / Обов'язковий | | | **Лекції (години)** | 32 | | | **Практичні заняття (години)** | | 32 | **Самостійна робота (години)** | 116 |
| **Програмні компетентності** | | * ЗК05. Здатність оцінювати та забезпечувати якість виконуваних робіт * СК02. Здатність виявляти, діагностувати та інтерпретувати соціальні проблеми українського суспільства та світової спільноти. * CК03. Здатність проектувати і виконувати соціологічні дослідження, розробляти й обґрунтовувати їхню методологію. * СК04. Здатність збирати та аналізувати емпіричні дані з використанням сучасних методів соціологічних досліджень. * СК07. Здатність розробляти та оцінювати соціальні проекти і програми. * СК10. Здатність використовувати сучасні методи соціологічних досліджень в інтернет-просторі. | | | | | | | | | | | |

**Результати навчання:**

* ПР01. Аналізувати соціальні явища і процеси, використовуючи емпіричні дані та сучасні концепції і теорії соціології.
* ПР02. Здійснювати діагностику та інтерпретацію соціальних проблем українського суспільства та світової спільноти, причини їхнього виникнення та наслідки.
* ПР03. Розробляти і реалізовувати соціальні та міждисциплінарні проекти з урахуванням соціальних, економічних, правових, екологічних та інших аспектів суспільного життя.
* ПР04. Застосовувати наукові знання, соціологічні та статистичні методи, цифрові технології, спеціалізоване програмне забезпечення для розв’язування складних задач соціології та суміжних галузей знань.
* ПР05. Здійснювати пошук, аналізувати та оцінювати необхідну інформацію в науковій літературі, банках даних та інших джерелах.
* ПР09. Планувати і виконувати наукові дослідження у сфері соціології, аналізувати результати, обґрунтовувати висновки.
* ПР11. Застосовувати наукові знання, соціологічні та статистичні методи, цифрові технології, спеціалізоване програмне забезпечення для розв’язування складних задач соціології та суміжних галузей знань.

**ТЕМИ ЩО РОЗГЛЯДАЮТЬСЯ**

**Тема 1. Основні елементи формалізму**

Аналіз соціологічної інформації, зібраної в ході емпіричних соціологічних досліджень, є не просто сукупністю технічних прийомів і методів. Неодновимірність багатьох досліджуваних соціологом понять. Непрямий її прояв – порушення транзитивності відношення порядку.

Метричне та неметричне БШ. Відповідні функції стресу. Неявне порівняння відстаней між близькістю, закладене у формулі функції стресу для метричного шкалірування. Поняття монотонної регресії, що використовується при розрахунку функції стресу для неметричного шкалірування.

Важливість для соціології неметричного шкалірування. Формальні аспекти проблем розмірності шуканого евклідового простору і обертання, що визначають його осей координат.

**Тема 2. Багатовимірне розгортання та індивідуальне багатовимірне шкалірування**

Постановка завдання; важливість врахування специфіки метрик окремих респондентів. Спосіб обліку таких метрик в індивідуальному БШ. Вид вхідних і вихідних даних, функції стресу в індивідуальному БШ.

Одномірне розгортання. Обґрунтування необхідності переходу до простору довільної розмірності для успішного виконання завдання шкалірування. Модель ідеальної точки в багатовимірному випадку. Неметричне багатовимірне розгортання. Вид вихідних даних. Функція стресу. Специфіка вихідних даних (наявність двох видів точок, що відповідають об'єктам і респондентам відповідно). Особливості інтерпретації результатів.

**Тема 3. Проблеми формування вихідних даних і інтерпретації результатів у багатовимірному шкалірування**

Роль соціолога при отриманні даних, вихідних для багатовимірного шкалірування, та інтерпретації його результатів. Можливі способи одержання вихідних даних. Безпосереднє отримання близькості від респондентів, класифікація відповідних способів опитування; проблеми, що постають при такому способі збору даних. Приклади розрахунку матриці близькості на основі аналізу достатньо надійних даних іншого роду.

Робота з БШ статистичними програмами - процедура БШ доступна в більшості статистичних програм. Існує вибір між метричним БШ (який дозволяє працювати з інтервалами чи даними про співвідношення рівня), і неметричним БШ (який працює з порядковими даними).

Використання формальних та неформальних методів при інтерпретації результатів багатовимірного шкалірування. Значення змістовних концепцій дослідника при вирішенні проблем вибору розмірності евклідова простору і повороту його осей.

**Тема 4. Канонічний аналіз. Загальне уявлення про методи, які засновані на моделях частот**

Загальне уявлення про моделювання частот таблиці спряженості. Змістовне розуміння таких моделей, їх роль для соціолога. Мультиплікативні та адитивні моделі частот. Роль логарифмування мультиплікативної моделі. Можливість різного розуміння як сенсу розглянутих вкладів, так і того "середнього" рівня, з яким порівнюються спостерігаються частоти в процесі їх моделювання.

Канонічний кореляційний аналіз – один із мето- дів багатовимірного аналізу даних.

Необхідність сполучення моделі, закладеної в конкретному методі оцифровки, з вмістом розглянутої задачі. Приклад моделі такого роду – модель, використовувана в методі шкалірування, званому методом послідовних розбивок. Канонічний аналіз як метод оцифровки і метод вимірювання зв'язку між двома номінальними ознаками зі "спільними альтернативами". Моделі частот, що відповідають канонічному аналізу. Побудова соціологічних індексів за допомогою техніки канонічного аналізу. Вирішення проблеми зважування складових індекс ознак.

**Тема 5. Логлінейний аналіз**

Логлінейний аналіз - метод багатовимірного статистичного аналізу для вивчення таблиць спряженості. Логлінейний аналіз дозволяє статистично перевіряти гіпотезу про систему одночасно мають місце парних і множинних взаємозв'язків в групі ознак, виміряних за номінальними шкалами. Багатовимірний статистичний аналіз.

Моделі частот, що відповідають логлінейному аналізу. Насичена модель. Мета переходу до логарифмів частот. Сенс вкладів різної розмірності.

Різне розуміння залежної ознаки: кількісна ознака в дисперсійному аналізі, кількісна або номінальна – в номінальному регресійному і частота, що стоїть в клітці багатовимірної таблиці спряженості, – в логлінейном аналізі. Різні можливості пошуку поєднань значень предикторів: перевірка гіпотез про наявність багатовимірних зв'язків у логлінейном аналізі і можливість пошуку найбільш дієвих поєднань в методі послідовних розбивок і регресійному аналізі, заздалегідь заданий набір поєднань значень предикторів в дисперсійному аналізі.

**Тема 6. Причинний аналіз. Стратегія аналізу структури взаємозв'язків ознак**

Поняття причини в соціології. Принципова неможливість повністю його формалізувати. Роль статистичних методів при вивченні причинних відносин.

Граф причинних зв'язків. Структурні коефіцієнти. Вхідні (зовнішні, незалежні) і вихідні (внутрішні, залежні) змінні. Правила редукції причинних схем та формування рівнянь.

Повторення принципів побудови часткових коефіцієнтів кореляції і регресії. Важливість для соціолога вивчення відповідних зв'язків. Різниця між статистичним та причинним зв'язком. Поняття "помилкової" кореляції. Основні причинні схеми, що призводять до їх появи.

Проблема формалізації завдання вивчення причинно-наслідкових відносин в соціології. Поняття структури багатовимірної випадкової величини. Формування узагальнених показників на базі аналізу структури зв'язків ознак. Комплексне використання декількох методів вивчення зв'язків між ознаками для вирішення соціологічних задач (аналіз структури випадкової величини; факторний і дисперсійний аналіз; пошук детермінуючих поєднань значень предикторів).

**Тема 7. Завдання розпізнавання образів. Поняття автоматичної класифікації об'єктів**

Класифікація як один із фундаментальних процесів у науці. Ознаковий простір. Задача класифікації як пошук згущення точок – моделей об'єктів в ознаковому просторі.

Виділення задачі автоматичної класифікації об'єктів (синоніми: багатовимірна класифікація, розпізнавання образів без вчителя, кластерний аналіз, таксономія).

**Тема 8. Проблема "стикування" змісту і формалізму при використанні алгоритмів класифікації**

Специфіка рішення соціологічних завдань побудови типології за допомогою методів автоматичної класифікації. Сенс протиставлення термінів "класифікація" і "типологія". Підстава типології. Роль апріорних уявлень дослідника про шуканих типах у виборі і реалізації алгоритму, інтерпретації результатів його застосування. Виділення основних формальних елементів алгоритмів автоматичної класифікації, що вимагають стикування зі змістовними концепціями соціолога.

**Тема 9. Функції відстані між об'єктами**

Аксіоматичне визначення функції відстані і ролі цієї функції в соціології. Приклади непридатності евклідової відстані з точки зору апріорного змістовного розуміння типів об'єктів.

Можливість використання евклідової відстані в розглянутих прикладах за рахунок зміни ознакового простору.

Сучасний аналіз даних обумовлюється способами отримання величин, методами їх обробки й залежить від розвитку математичних методів і моделювання. Функції відстані, відмінні від евклідова: зважене евклидово, сіті-блок, Махаланобіса, Хеммінгово.

**Тема 10. Основні види процедур класифікації. Відстані між класами**

Актуальність дослідження сутності та методів багатовимірного аналізу соціологічної інформації обумовлена специфікою соціальної реальності, що завжди уявляється як складний, багатогранний та багатозначний феномен, який інтегрує багатовимірність суспільства з багатовимірністю внутрішнього світу окремої людини. Виділення ієрархічних і неієрархічні алгоритмів класифікації. Багатовимірний статистичний аналіз (у широкому значенні) - розділ математичної статистики, що поєднує методи вивчення даних, які характеризують багатовимірні об'єкти. Агломератівні та дівізімні алгоритми. Причини необхідності розгляду відстаней між класами в ієрархічних процедурах. Алгоритм найближчого сусіда як приклад способу класифікації, що використовує такі відстані.

**Тема 11. Гіпотези про розташування об'єктів у ознаковому просторі**

Роль гіпотез про характер розташування об'єктів у виборі алгоритму класифікації. Обумовленість цих гіпотез апріорними уявленнями дослідника про типи об'єктів. Основні види гіпотез: компактності, зв'язності (безперервності), унімодального розподілу.

Факторний аналіз найбільш яскраво відображує риси багатомірного аналізу в частині дослідження зв'язку між ознаками. Кластерний аналіз ці риси відображує з боку класифікації об'єктів.

Загальне уявлення про розмиті класифікації. Роль функції належності у відповідних алгоритмах. Доцільність комплексного використання декількох алгоритмів класифікації в соціологічних завданнях побудови типології.

Змістовні уявлення соціолога про типи та умови вибору кроку розбиття при інтерпретації результатів. Коригування результатів класифікації з метою забезпечення відповідності класифікації і типології.

**Тема 12. Поняття інтерпретації вихідних даних і основні методологічні принципи використання методів аналізу даних в соціології**

Інтерпретація вихідних даних як одне з основних ланок "стикування" соціології і математики. Основні фактори, що визначають інтерпретацію вихідних даних: апріорні уявлення дослідника про спосіб породження цих даних (у тому числі – про моделі сприйняття респондентами пропонованих ним питань, об'єктів, про ймовірнісну природу даних і т. д.); мета дослідження; концептуальні уявлення соціолога про досліджуване явище; характер моделі явища, "закладеної" в математичному методі, використання якого планується; розгляд спостережуваних змінних як непрямих показників латентних факторів, насправді цікавлять дослідника і т. п.

Виділення методологічних принципів, дотримання яких є необхідним для того, щоб аналіз соціологічних даних був ефективний, не відводив соціолога в сторону від реальності: забезпечення певної однорідності вихідних даних; облік моделі, "закладеної" в кожному методі аналізу даних, при виборі алгоритму аналізу, два основні принципи інтерпретації результатів аналізу: необхідність її узгодження з інтерпретацією вихідних даних і заповнення при її здійсненні тих втрат, які мали місце при переході до формалізму; необхідність комплексного використання декількох методів для вирішення одного завдання і т. д.

**Тема 13. Дані. Метадані**

Згідно з ГОСТ, дані – подання інформації у формалізованому вигляді, придатному для передачі, інтерпретації та обробки.

Вихідне поняття даних - філософське, воно виникає в епістемології під час розгляду основою проблеми гносеології – пізнаваності світу, пошуку та осмислення істини. Процедури верифікації чи фальсифікації даних створюють інформацію, осмислення істини створює знання.

Життєвий цикл даних – це послідовність етапів, яку конкретна порція даних проходить від початкового етапу створення чи отримання до моменту архівації чи видалення.

При зборі даних виникають метадані, що містять будь-яку інформацію про зібрані дані.

Огляд основних аналітичних інструментів роботи з Big data соціальних науках (Python, R, SAS, та ін). Читання та запис даних, формати файлів. Завантаження даних із різних джерел. Взаємодія з базами даних. Читання даних із Excel. Робота з CSV файлами та даними у форматі JSON. Парсинг простих даних XML. Читання даних із таблиць HTNL. Читання даних із файлу SAS. Взаємодія з HTML та Web API.

**Тема 14. Великі дані. Системи керування великими даними**

Великі дані можуть бути різних типів. Інформацію, отриману в результаті обліку або вимірювання будь-яких об'єктів або параметрів, називають майстер-даними (Master Data). Наприклад, облік кількості, виміри координат і швидкостей конкретних молекул - це майстер-дані.

Транзакційні дані (в англомовній літературі застосовуються терміни Transactional Data, Application Specific Data, Operational Data) – це дані, що відображають результат виконання будь-яких операцій. Транзакційні дані описують взаємодія об'єктів один з одним або з навколишнім світом, які можна отримати за допомогою обробки майстер-даних.

Ретроспективні дані (Historical data) – це дані, забезпечені позначки часу.

Посилальні дані (довідники, НСІ, нормативно-посилальна інформація,

Reference Data, Lookup Data, Dictionaries) – це базові незмінні дані, заздалегідь відомі із зовнішніх джерел, такі як нормативи, скорочення, акроніми, словники, стандарти.

Формат даних. Структуровані дані мають заздалегідь визначений формат. Напівструктуровані або слабоструктуровані дані - це дані, які часто зібрані з різних джерел.

**Тема 15. Програмні платформи та системи для Великих даних**

В даний час використовується значна кількість платформ та систем Великих даних. Системи обробки великих даних є фреймворками, тобто каркасами, для використання яких необхідно з'єднати їх з іншими фреймворками, прикладним програмним забезпеченням користувача та системою зберігання даних.

В аналітичному звіті Big Data Analytics Market Study, 2017 Edition наводиться така діаграма інфраструктур Великих даних, впроваджених на підприємствах, представлена у розрізі розмірів підприємств

Розподілена обробка даних тісно пов’язана з паралельною обробкою даних. Однак така обробка завжди виконується за допомогою окремих машин у кластері, підключеному до мережі. Розподілена обробка даних - це метод виконання прикладних програм групою систем. Користувач може працювати з мережевими службами та прикладними процесами, розташованими в кількох взаємопов’язаних абонентських системах. Розподілена обробка даних підвищує ефективність інформаційних потреб користувачів і забезпечує ефективність та результативність рішень.

**Тема 16. Машинне навчання за допомогою бібліотеки Scikit-learn.**

Види машинного навчання. Основні бібліотеки машинного навчання Python (Scikit-learn, Keras, TensorFlow). Створення тренувальних наборів - передобробка даних. Точність та достовірність моделі. Вибір найкращої моделі.

Кроки типового практичного сценарію машинного навчання. Завантаження набору даних. Дослідження даних за допомогою Pandas. Візуалізація ознак за допомогою Matplotlib. Розбиття даних для навчання та тестування. Створення моделі. Вивчення моделі. Тестування моделі.

Налаштування параметрів моделі та оцінка її точності. Формування прогнозів на на підставі «живих» даних, які ще невідомі моделі.

Функціонал бібліотеки Scikit-Learn. Класифікація за допомогою К-сусідів.

Лінійні моделі для регресії та класифікації (модель лінійної регресії, логістична регресія, та ін). Наївні байєсівські класифікатори. Дерева рішень та випадковий ліс. Спосіб опорних векторів. Основи нейронних мереж.

Метод основних компонентів. Алгоритми кластеризації (кластеризація методом К-середніх, ієрархічна кластеризація, та ін).

**Форма та методи навчання**

Під час проведення лекційних занять з навчальної дисципліни передбачено застосування таких методів навчання: пояснювально-ілюстративний метод (демонстрація на екрані слайдів презентацій, візуалізації навчального матеріалу);метод проблемного викладення**;**частково-пошуковий, абоевристичний метод (лекції за окремими темами викладаються в проблемний формі)

Під час проведення семінарських занять застосовується: репродуктивний метод (засвоєння базових понять курсу); частково-пошуковий, або евристичний метод(під час підготовки індивідуальних проєктів); дослідницький метод (студенти самостійно вивчають літературу, джерела, ведуть дослідження, виміри та виконують інші пошукові дії для створення презентації на задану тему).Семінарські заняття можуть бути побудовані у формі дискусії.

**Лекції –** викладення теоретичного матеріалу лектором згідно навчальної програми і розподілу годин поміж темами.

**Практичні заняття** – проводяться у формі семінарських занять. Для семінарських занять студенти опрацьовують лекційний матеріал, готують виступи з використанням навчальної і наукової літератури, виступають з презентаціями. Лектор оцінює активність студентів впродовж семінару за прийнятою шкалою оцінок в балах. Під час семінарського заняття обов’язково за кожною темою оцінюються рівень знань студентів за допомогою тестових завдань та письмової самостійної роботи на знання основних понять за темою.

**Індивідуальне завдання** – вид самостійної роботи поза аудиторними годинами, коли студент, використовуючи лекційний матеріал та додаткові джерела знань, розробляє особисту тему.

**Підготовка презентації –** вид самостійної роботи, що виконується студентом (або 2-3 студентами) поза аудиторними годинами. Студент вільно обирає тематику з числа тем, які пропонуються планами семінарських занять, або узгоджує з викладачем ініціативну тематику. Наступним кроком студент здійснює бібліографічний пошук, використовуючи бібліотечні фонди або Інтернет-ресурси. Також складає план презентації або ставить питання, на які треба отримати аргументовану відповідь. Опанувавши джерела за темою, студент розкриває зміст питань та представляє виконану роботу на семінарі. Обсяг презентації – 16-25 слайдів, текст доповіді – 4-6 стандартних сторінок, набраних на комп’ютері. Основний зміст презентації доповідається у вільній формі на семінарському занятті, і студент отримує оцінку від викладача.

Методи контролю

**1. Підсумковий (семестровий) контроль проводиться у формі екзамену або шляхом накопичення балів за поточним контролем по змістовним модулям.**

**Екзамен –** письмова або усна відповідь на питання, що містяться в екзаменаційному білеті, або відповідь на підсумковий тест по курсу (студент обирає форму). Питання білетів доводяться до студентів заздалегідь.

**Контрольні питання з курсу до екзамену.**

1. Багатовимірне шкалірування: коло вирішуваних завдань.

2. Багатовимірне шкалірування: основні елементи формалізму (близькості, відстані, функція стресу).

3. Індивідуальне багатовимірне шкалірування: основні ідеї, мета використання в соціології, функція відстані.

4. Багатовимірне розгортання: основні ідеї, сенс вирішуються за його допомогою соціологічних завдань.

5. Багатовимірне шкалірування: проблеми формування вихідних даних і інтерпретації результатів.

6. Багатовимірне шкалірування: проблеми інтерпретації результатів.

7. Поняття багатовимірної зв'язку. Відносини переважання.

8. Логлінейний аналіз: мета використання в соціології, моделі частот.

9. Розрахунок параметрів логлінейной моделі для чотириклетинної таблиці спряженості. Зв'язок одержуваних величин з відносинами переважання.

10. Логлінейний аналіз: проблема формування гіпотез.

11. Порівняння можливостей логлінейного і номінального регресійного аналізу.

12. Порівняння можливостей логлінейного і дисперсійного аналізу.

13. Порівняння можливостей логлінейного аналізу та алгоритмів послідовних розбивок.

14. Канонічний аналіз: постановка завдання, канонічні кореляції.

15. Поняття канонічної кореляції як узагальнення множинного коефіцієнта кореляції.

16. Канонічний аналіз: моделі частот, використання при аналізі таблиць спряженості.

17. Канонічний аналіз: використання для побудови соціологічних індексів.

18. Оцифровка значень номінальних і порядкових ознак. Цілі використання відповідних методів. Подання про модель, що стоїть за кожним методом.

19. Канонічний аналіз як метод оцифровки.

20. Причинний аналіз: граф причинних зв'язків, структурні коефіцієнти, що координує шлях, його ефективність.

21. Причинний аналіз: обчислення коваріації (кореляції) будь-яких двох ознак на основі графа причинних зв'язків, шляхові коефіцієнти.

22. Зв'язок структурних коефіцієнтів з регресійним. Структурні рівняння.

23. Основна теорема колійного аналізу.

24. Роль латентних факторів у причинному аналізі.

25. Комплексне використання різних методів при аналізі структури взаємозв'язків ознак.

26. Загальне уявлення про завдання розпізнавання образів. Ознаковий простір. Поняття автоматичної класифікації об'єктів.

27. Сенс термінів "класифікація" і "типологія". Їх роль при вирішенні соціологічних завдань побудови типології об'єктів.

28. Роль функції відстані між об'єктами в процесі класифікації. Проблема її адекватності змістовному розуміння типу об'єктів.

29. Евклідова відстань. Виважена евклидова відстань. Відстань Хемінга.

30. Загальне уявлення про ієрархічних і неієрархічні алгоритмах класифікації.

31. Роль функції відстані між класами при реалізації алгоритмів класифікації. Її види.

32. Вибір форми шуканих класів при використанні методів класифікації. Гіпотези про розташування об'єктів у признаковом просторі.

33. Гіпотеза компактності. Алгоритм ФОРЕЛЬ

34. Гіпотеза зв'язності. Алгоритм найближчого сусіда.

35. Гіпотеза унімодального розподілу. Алгоритм, заснований на виділенні локальних максимумів функції приналежності.

36. Забезпечення відповідності класифікації і типології в процесі інтерпретації результатів класифікації.

37. Поняття інтерпретації даних. ЇЇ роль в соціології.

38. Принципи сполучення формалізму і змісту, зв'язку всіх етапів дослідження один з одним як основні методологічні принципи застосування методів аналізу даних в соціології. Приклади їх реалізації.

39. Забезпечення однорідності досліджуваної сукупності об'єктів як один з основних методологічних принципів застосування методів аналізу даних в соціології. Приклади його реалізації.

40. Методологічні принципи інтерпретації результатів застосування математичного методу для вирішення соціологічної завдання. Приклади їх реалізації.

41. Визначення даних. Філософський, юридичний підходи й життєвий цикл даних.

42.Поняття метаданих. Життєвий цикл метаданих

43. Специфікація системних вимог. Система метаданих

44. Розподілені файлові системи

45. Бенчмаркінг

46. Системи керування потоками даних

47. Системи зберігання Великих даних

48. Аналітичні платформи Види машинного навчання.

49. Основні бібліотеки машинного навчання Python (Scikit-learn, Keras, TensorFlow).

50. Функціонал бібліотеки Scikit-Learn.

**2.Поточний контроль проводиться за результатами роботи студентів на семінарських заняттях, методом оцінювання контрольних робіт, оцінювання тестів, самостійних робот, індивідуальних завдань, командних проектів, презентацій.**

**Контроль на семінарських заняттях –** оцінювання виступів студентів, відповідей на питання поставлені викладачем, оцінці виконання тестових завдань, оцінок під час самостійних робот, оцінювання внеску окремих студентів у групову роботу при підготовки командного проекту, активність в діловій грі.

**Контрольна робота** – вид поточного контролю знань студентів, який має на меті виявити рівень знань студентів, що отримані за пройденим матеріалом. Дата проведення контрольної роботи доводиться до студентів і призначається по завершенню вивчення змістовного модулю. Питання та тестові завдання готує викладач, що веде практичні заняття, вони узгоджуються з лекційними питаннями і тематикою семінарських занять. Контрольна робота виконується у письмовій формі в присутності викладача, оцінюється за прийнятою шкалою і оцінка може використовувати викладачем для підрахунку кумулятивного балу за підсумками вивчення дисципліни.

**Індивідуальні завдання** – оцінюються викладачем або за результатами доповіді на практичному занятті або окремо за наданим текстом.

|  |  |  |  |
| --- | --- | --- | --- |
|  | |  |  | | --- | --- | |  | Виконання проекту передбачає командну(2-3 студента) або індивідуальну дослідницьку роботу за вибраною темою, підготовку письмового звіту та проведення презентації за допомогою мультимедійного обладнання в присутності викладачів кафедри. Командний проект – це пізнавально-аналітична робота. | |

*Ціль проекту* полягає в перевірці успішності засвоєння студентами методів багатомірного аналізу даних та уміння використовувати соціологічну уяву для аналізу явищ і процесів, що відбуваються у суспільстві.

Розподіл балів, які отримують студенти

Таблиця 1. – Розподіл балів для оцінювання успішності студента для іспиту

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Робота на практичному занятті | Контрольні роботи | Індивідуальні завдання (проекти) | Реферат | Екзамен | Сума |
| 48 (3\*16) | 10 (5\*2) | 12 | 10 | 20 | 100 |

Таблиця 2. – Шкала оцінюваннязнань та умінь: національна та ЕСТS

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Рейтингова**  **Оцінка, бали** | **Оцінка ЕСТS та її визначення** | **Національна оцінка** | **Критерії оцінювання** | |
| **позитивні** | **негативні** |
| **1** | **2** | **3** | **4** | **5** |
| 90-100 | А | Відмінно | - **Глибоке знання** навчального матеріалу модуля, що містяться в **основних і додаткових літературних джерелах;**  - **вміння аналізувати** явища, які вивчаються, в їхньому взаємозв’язку і розвитку;  - **вміння** проводити **теоретичні розрахунки**;  - **відповіді** на запитання **чіткі**, **лаконічні, логічно послідовні;**  **- вміння вирішувати складні практичні задачі.** | Відповіді на запитання можуть містити **незначні неточності** |
| 82-89 | В | Добре | - **Глибокий рівень знань** в обсязі **обов’язкового матеріалу**, що передбачений модулем;  - вміння давати **аргументовані відповіді** на запитання і проводити **теоретичні розрахунки**;  - вміння вирішувати **складні практичні задачі.** | Відповіді на запитання містять **певні неточності;** |
| 75-81 | С | Добре | - **Міцні знання** матеріалу, що вивчається, та його **практичного застосування;**  **-** вміння давати **аргументовані відповіді** на запитання і проводити **теоретичні розрахунки**;  - вміння вирішувати **практичні задачі.** | **-** невміння використовувати теоретичні знання для вирішення **складних практичних задач.** |
| 64-74 | D  Д | Задовільно | - Знання **основних фундаментальних положень** матеріалу, що вивчається, та їх **практичного застосування**;  - вміння вирішувати прості **практичні задачі**. | Невміння давати **аргументовані відповіді** на запитання;  - невміння **аналізувати** викладений матеріал і **виконувати розрахунки;**  - невміння вирішувати **складні практичні задачі.** |
| 660-63 | Е | Задовільно | - Знання **основних фундаментальних положень** матеріалу модуля,  - вміння вирішувати найпростіші **практичні задачі**. | Незнання **окремих (непринципових) питань** з матеріалу модуля;  - невміння **послідовно і аргументовано** висловлювати думку;  - невміння застосовувати теоретичні положення при розвязанні **практичних задач** |
| 35-59 | FХ  (потрібне додаткове вивчення) | Незадовільно | **Додаткове вивчення** матеріалу модуля може бути виконане **в терміни, що передбачені навчальним планом**. | Незнання **основних фундаментальних положень** навчального матеріалу модуля;  - **істотні помилки** у відповідях на запитання;  - невміння розв’язувати **прості практичні задачі.** |
| 1-34 | F  (потрібне повторне вивчення) | Незадовільно | - | - Повна **відсутність знань** значної частини навчального матеріалу модуля;  - **істотні помилки** у відповідях на запитання;  -незнання основних фундаментальних положень;  - невміння орієнтуватися під час розв’язання **простих практичних задач** |

Основна література:

**Базова література**

|  |  |
| --- | --- |
| 1 | Горбачик А.П., Сальнікова С.А. Аналіз даних соціологічних дослідженьзасобами SPSS: Навч. посіб.- Луцьк, 2008. – 164 с. |
| 2 | IBM SPSS 20 інструкція користувача// https://www.xn--80aaexjatkpdggghih8b1a2yhv.com.ua/ibm/spss-20/%D1%96%D0%BD%D1%81%D1%82%D1%80%D1%83%D0%BA%D1%86%D1%96%D1%8F-%D0%BA%D0%BE%D1%80%D0%B8%D1%81%D1%82%D1%83%D0%B2%D0%B0%D1%87%D0%B0 |
| 3 | Паніотто В.І., Максименко В. С., Харченко Н.М. Статистичний аналізсоціологічних даних. - Київ, 2004. – 270 с. |
| 4 | Литвин В.В. Аналіз даних та знань: підручник/ В.В. Литвин, В.В. Пасічник, Ю.В. Нікольський.- Л.: Магнолія, 2020.- 276с. (базовий підручник). |

**Допоміжна література**

|  |  |
| --- | --- |
| 5 | Лупан I.В., Авраменко О.В., Акбаш К.С. Комп’ютерні статистичні пакети: навчально-методичний посібник. - 2-е вид. - Кіровоград: ’КОД”. 2015. - 230 с. - http://dspace.cuspu.edu.ua/jspui/bitstream/123456789 |
| 6 | Making Sense of Multivariate Data Analysis//https://us.sagepub.com/en-us/nam/book/making-sense-multivariate-data-analysis |
| 7 | Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів В.Є. Бахрушин. - Запоріжжя : КПУ, 2011. - 26В с. - http://web.kpi.kharkov.ua/auts/wp-content/uploads/sites/67/2017/02/DAMAP\_Ivashko\_posobie2.pdf |
| 8 | Інтелектуальний аналіз даних: практикум/ М.Т. Фісун, І.О. Кравець, П.П. Казмірчук.- Л.: Новий Світ-2000, 2020.- 162с. |
| 9 | Гладун А.Я., Рогушина Ю. В. Data Mining: пошук знань в даних. Київ. ТОВ «ВД «АДЕФ- Україна», 2016. — 452 с.. |

**Структурно-логічна схема вивчення навчальної дисципліни**

Таблиця 4. – Перелік дисциплін

|  |  |
| --- | --- |
| Вивчення цієї дисципліни безпосередньо спирається на: | На результати вивчення цієї дисципліни безпосередньо спираються: |
| Математичні методи в соціології | Переддипломна практика |
| Практикум з комп’ютерної обробки соціологічних даних | Атестація |
| Соціологічний супровід економічної діяльності |  |
| Інтернет-дослідження економічної діяльності |  |

**Провідний лектор:** проф. Бірюкова М.В.  **\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

(посада, звання, ПІБ) (підпис)