**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ**

**«ХАРКІВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

Кафедра                соціології і публічного управління

(назва кафедри, яка забезпечує викладання дисципліни)

«**ЗАТВЕРДЖУЮ**»

Завідувач кафедри соціології і публічного управління

(назва кафедри )

\_\_\_\_\_\_\_\_\_\_\_\_ Мороз В.М .

(підпис) (ініціали та прізвище)

«31» серпня 2022 року

**ЗАВДАННЯ ДЛЯ САМОСТІЙНОЇ РОБОТИ З НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

Методи багатовимірного аналізу та bigdata в соціології

( назва навчальної дисципліни)

рівень вищої освіти\_\_\_\_\_другий (магістерський)

перший (бакалаврський) / другий (магістерський)

галузь знань\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_05 Соціальні та поведінкові науки

(шифр і назва)

спеціальність \_\_\_\_\_\_\_054 Соціологія

(шифр і назва )

освітня програма Соціологічне забезпечення економічної діяльності

(назви освітніх програм спеціальностей )

вид дисципліни професійна підготовка (обов’язкова)

(загальна підготовка / професійна підготовка; обов’язкова/вибіркова)

форма навчання денна

(денна / заочна/дистанційна)

Харків – 2022 рік

**Тема 1. Основні елементи формалізму**

1. Проблеми неодновимірності багатьох досліджуваних соціологом понять.
2. Особливості вивчення простору сприйняття соціологічних явищ та процесів – основне завдання БШ.
3. Ідеї Кумбса щодо урахування можливості упорядкування відстаней між об'єктами.
4. Векторна модель або модель ідеальної крапки як основа БШ.
5. Функція відстані (аксіоматичне визначення).
6. Відповідні функції стресу.

**Література:** 1, 3, 4, 5

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

1. Постановка завдання важливість врахування специфіки метрик окремих респондентів.
2. Вид вхідних і вихідних даних, функції стресу в індивідуальному БШ.
3. Одномірне розгортання.
4. Обґрунтування необхідності переходу до простору довільної розмірності для успішного виконання завдання шкалірування.
5. Неметричне багатовимірне розгортання.
6. Особливості інтерпретації результатів.
7. Розв’язання практичних завдань.

**Література:** 1, 3, 4, 5, 7

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

1. Роль соціолога при отриманні даних, вихідних для багатовимірного шкалірування та інтерпретації його результатів.
2. Класифікація відповідних способів опитування; проблеми, що постають при такому способі збору даних.
3. Приклади розрахунку матриці близькості на основі аналізу достатньо надійних даних іншого роду.
4. 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.
5. Розв’язання практичних завдань.

**Література:** 1, 3, 4, 5, 6

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

1. Загальне уявлення про моделювання частот таблиці спряженості.
2. Мультиплікативні та адитивні моделі частот.
3. Роль логарифмування мультиплікативної моделі.
4. Основне завдання канонічного аналізу. Принципи їх отримання на основі аналізу таблиці спряженості.
5. Моделі частот, що відповідають канонічному аналізу.
6. Зв'язок канонічних коефіцієнтів кореляції з критерієм «хі-квадрат".
7. Загальне уявлення про оцифрування значень номінальних ознак.
8. Канонічний аналіз як метод оцифровки і метод вимірювання зв'язку між двома номінальними ознаками зі "спільними альтернативами".
9. Розв’язання практичних завдань.

**Література:** 1, 3, 4, 5, 6

**Тема 5. Логлінейний аналіз**

1. Причини відмінності реального розподілу від рівномірного.
2. Моделі частот, що відповідають логлінейному аналізу.
3. Насичена модель.
4. Мета переходу до логарифмів частот.
5. Гіпотези про взаємозв'язок ознак. Їх роль при побудові моделей частот.
6. Розрахунок коефіцієнтів логлінейной моделі для двовимірного випадку. Відносини переважання. Інтерпретація коефіцієнтів через відносини переважання (для моделі довільної розмірності).
7. Порівняння логлінейного аналізу з номінальним регресійним і дисперсійним аналізом, а також з методом послідовних розбивок. Порівняння здійснюється на змістовному рівні.
8. 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.
9. Розв’язання практичних завдань.

**Література:** 1, 2, 3, 4, 5

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

1. Поняття причини в соціології. Принципова неможливість повністю його формалізувати.
2. Граф причинних зв'язків.
3. Повторення принципів побудови часткових коефіцієнтів кореляції і регресії. Важливість для соціолога вивчення відповідних зв'язків.
4. Поняття "помилкової" кореляції. Основні причинні схеми, що призводять до їх появи.
5. Координуючий шлях. Його ефект.
6. Обчислення ковариаций (кореляцій) між будь-якими двома ознаками на основі графа зв'язків.
7. Структурні рівняння.
8. Обчислення структурних коефіцієнтів. Їх зв'язок з частковими коефіцієнтами регресії.
9. Основна теорема причинного аналізу. Її роль у вивченні статистичних залежностей.
10. Поняття структури багатовимірної випадкової величини.
11. Формування узагальнених показників на базі аналізу структури зв'язків ознак.
12. Розв’язання практичних завдань.

**Література:** 1, 3, 4, 5, 6

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

1. Класифікація як один із фундаментальних процесів у науці.
2. Загальне уявлення про завдання розпізнавання образів (синоніми: образ, клас, кластер, таксон; неоднозначність трактування термінів в літературі).
3. Виділення завдань: пошук класів, опис класів, визначення найбільш ефективної системи ознак.
4. Виділення задачі автоматичної класифікації об'єктів (синоніми: багатовимірна класифікація, розпізнавання образів без вчителя, кластерний аналіз, таксономія).
5. Розв’язання практичних завдань.

**Література:** 1, 3, 4, 5, 6

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

1. Специфіка рішення соціологічних завдань побудови типології за допомогою методів автоматичної класифікації.
2. Сенс протиставлення термінів "класифікація" і "типологія".
3. Виділення основних формальних елементів алгоритмів автоматичної класифікації, що вимагають стикування зі змістовними концепціями соціолога.
4. Розв’язання практичних завдань.

**Література:** 1, 3, 4, 5, 6

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

1. Обумовленість гіпотез апріорними уявленнями дослідника про типи об'єктів.

2. Загальне уявлення про розмиті класифікації.

3. Доцільність комплексного використання декількох алгоритмів класифікації в соціологічних завданнях побудови типології.

**Література:** 1, 2, 3, 4, 5

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

* 1. Причини необхідності розгляду відстаней між класами в ієрархічних процедурах.
  2. Алгоритм найближчого сусіда як приклад способу класифікації, що використовує такі відстані.
  3. Приклади соціологічних задач, для яких змістовно адекватні різні способи вимірювання відстаней між класами.

**Література:** 1, 3, 4, 5, 7

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

1. Обумовленість гіпотез апріорними уявленнями дослідника про типи об'єктів.

2. Основні види гіпотез: компактності, зв'язності (безперервності), унімодального розподілу.

3. Приклади алгоритмів, що шукають закономірності розташування точок у ознаковому просторі, що відповідають кожній з гіпотез: алгоритм Форель (гіпотеза компактності), алгоритм найближчого сусіда (гіпотеза зв'язності), алгоритм, заснований на виділенні локальних максимумів функції приналежності (гіпотеза унімодального розподілу ).

4. Доцільність комплексного використання декількох алгоритмів класифікації в соціологічних завданнях побудови типології.

5. Коригування результатів класифікації з метою забезпечення відповідності класифікації і типології.

**Література:** 1, 3, 4

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

1. 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.

**Література:** 1, 3, 4, 5

**Тема 13.** **Дані. Метадані**

Підготувати відповіді на питання:

1. Що таке дані?

2. Які ДСТУ з визначеннями даних вам відомі?

3. Які визначення даються у ФЗ-149?

4. Що таке життєвий цикл даних?

5. Перерахуйте етапи життєвого циклу даних.

6. Для яких цілей потрібен етап “Синтез даних” (один із етапів життєвого циклу даних)?

7. З якою метою потрібен етап “Використання даних” (один із етапів життєвого циклу даних)?

8. З якою метою потрібен етап “Публікація даних” (один із етапів життєвого циклу даних)?

9. З якою метою потрібен етап “Архівація даних” (один із етапів життєвого циклу даних)?

**Література:** 4, 8, 9

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

Підготувати відповіді на питання:

1. Що таке Великі дані?

2. Які п'ять характеристик притаманні Великим даним?

3. Які є базові принципи обробки Великих даних?

4. Оцініть скільки необхідно для зберігання набору даних, що містить координати, швидкості та метаінформацію (тип молекули та час вимірювання по конкретній молекулі) для всіх молекул біля аеропорту.

5. Що таке стовпцеві бази даних?

6. Що таке сховища документів?

7. Що таке потокові дані?

8. Що таке сховища для ключів?

9. Що таке SQL на Hadoop?

10. Що таке новий SQL?

11. Що таке графові бази даних?

**Література:** 4, 8, 9

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

Підготувати відповіді на питання:

1. Які мови програмування використовуються для роботи з фреймворками даних?

2. Чи дозволяє ліцензія Apache 2.0, під якою випущено деякі фреймворки, вносити власні виправлення в програмний код забезпечення?

3. Перерахуйте кілька фреймворків, які забезпечують обробку даних у реальному часі.

4. Перерахуйте кілька фреймворків, які забезпечують аналітичну обробку даних.

5. Перерахуйте кілька фреймворків, які забезпечують зберігання даних.

6. Перерахуйте кілька фреймворків, які забезпечують керування потоками даних.

**Література:** 4, 8, 9

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

Підготувати відповіді на питання:

1. Яке обладнання потрібне для обробки великих даних?

2. Центр обробки даних якого рівня забезпечує максимальну надійність?

3. Центр обробки даних якого рівня забезпечує резервування?

4. Центр обробки даних якого рівня дозволяє проводити обслуговування обладнання одночасно з обробкою даних?

5. Як багато часу потрібно для створення центру обробки даних "під ключ"?

**Література:** 4, 8, 9

**ІНДИВІДУАЛЬНІ ЗАВДАННЯ**

**Проектна робота**

(вид індивідуального завдання)

|  |  |  |
| --- | --- | --- |
| **№**  **з/п** |  | **Терміни виконання  (на якому тижні)** |
| **1** | **Розробка індивідуального проекту щодо багатомірного шкалірування об’єктів за темою дипломної роботи** | 7-9 тижні |

**Реферат**

|  |  |  |
| --- | --- | --- |
| **№**  **з/п** | **Теми рефератів** | **Терміни виконання  (на якому тижні)** |
| 1 | **Теми рефератів**   1. Когнітивний аналіз опитувального інструментарію. 2. Огляд пакетів статистичного аналізу. 3. Основи статистики в комп'ютерних пакетах. 4. Так-система. Сильні і слабкі сторони логічних пакетів. 5. Програмування в SPSS. 6. Математичні методи в маркетингу. 7. Математичні методи вивчення конфліктних ситуацій. 8. Контент-аналіз і його реалізація на прикладі. 9. Дослідження Інтернету. Контент-аналіз, статистика та інтерпретація. 10. Факторний аналіз у політичних дослідженнях. 11. Кластеризація ринків. 12. Кластеризація національної економіки. 13. Кластеризація ціннісних орієнтацій. 14. Документообіг в світі безпаперових технологій. 15. Вітчизняний ринок програм обробки соціологічних даних. 16. Використання інформаційних систем для соціології: можливості і проблеми. 17. Консалтингова підтримка діяльності підприємств. 18. Бази знань та експертні системи. 19. Проблеми захисту інформації в автоматизованих системах. 20. Інноваційні напрями розвитку інформаційних технологій. 21. Використання кореляційно-регресійного аналізу для обробки соціологічних даних. 22. Застосування первинних і вторинних угруповань в аналізі соціологічних даних. 23. Багатовимірні угруповання в статистиці 24. Роль графіків в узагальненні та аналізі соціологічних даних. 25. Графічний метод у вивченні соціальної реальності. 26. Графічне зображення в узагальненні та аналізі статистичних даних. 27. Середні величини в статистиці, їх значення, види. 28. Застосування структурних середніх величин для аналізу соціальних явищ. 29. Роль показників варіації в оцінці достовірності даних проведених досліджень. 30. Використання різних методик розрахунку показників варіації 31. Вибіркове спостереження, як основний метод проведення статистичного дослідження: його етапи, властивості, переваги та недоліки. 32. Предмет аналізу даних і витоки формування його методології. 33. Модель вивчення властивості об'єкту 34. Типи емпіричних даних 35. Шкалювання та кодування в процесі вимірювання. 36. Індекси при зборі та аналізі даних. 37. Специфічні прийоми вимірювання соціальної установки 38. Аналіз взаємозв'язку ознак 39. Заходи зв'язку, засновані на поняттях «статистична залежність» і «детермінації». 40. Заходи зв'язку засновані на моделі прогнозу | 2-3  4-5  6-7  8-9 |

**РЕКОМЕНДОВАНА 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 аналіз соціологічних 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.. |

**ІНФОРМАЦІЙНІ РЕСУРСИ В ІНТЕРНЕТІ**

1. ukrstat.gov.ua (СтатистичніматеріалиДержкомстатуУкраїни)
2. http://i-soc.com.ua/institute/el\_library.php (Електронна бібліотека Інституту соціології НАН України)
3. <http://sau.in.ua/theme/simya/> (Соціологічна асоціація України, дослідження, сім’я)
4. <https://genderindetail.org.ua/> (Гендер в деталях)
5. <https://idss.org.ua/public> (Інститут демографії та соціальних досліджень імені М.В.Птухи Національної академії наук України, публікації)