**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ**

**«ХАРКІВСЬКИЙПОЛІТЕХНІЧНИЙ ІНСТИТУТ»**

Кафедра \_ соціології і публічногоуправління\_\_\_\_\_\_

(назва)

«**ЗАТВЕРДЖУЮ**»

Завідувач кафедри соціологіїі публічного управління

(назва кафедри )

\_\_\_\_\_\_\_\_\_\_\_\_ Мороз В.М. .

(підпис) (ініціали та прізвище)

«31» серпня 2022 року

**МЕТОДИ КОНТРОЛЮ ЗНАНЬ З НАВЧАЛЬНОЇ ДИСЦИПЛІНИ**

МЕТОДИ БАГАТОВИМІРНОГО АНАЛІЗУ ТА BIG DATA В СОЦІОЛОГІЇ

( названавчальноїдисципліни)

рівеньвищоїосвіти другий (магістерський)

перший (бакалаврський) / другий (магістерський)

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

(шифр і назва)

спеціальність 054 Соціологія

(((шифр і назва )

освітняпрограма Соціологічне забезпечення економічної діяльності

(назвиосвітніхпрограмспеціальностей )

вид дисципліни професійна підготовка (обов’язкова)

(загальнапідготовка (обов’язкова/вибіркова) / професійнапідготовка (обов’язкова/вибіркова))

форма навчання денна

(денна / заочна)

Харків – 2022 рік

**Контрольні роботи з курсу «**МЕТОДИ БАГАТОВИМІРНОГО АНАЛІЗУ ТА BIG DATA В СОЦІОЛОГІЇ»**.**

Контрольна робота полягає у самостійній роботі студента в поза аудиторний час над отриманими питаннями за темами курсу.

**Контрольна робота № 1**

### Варіант № 1

1. Проаналізуйте в чому полягає неодновимірність багатьох досліджуваних соціологом понять.
2. Визначте сутність переходу до простору довільної розмірності для успішного виконання завдання шкалірування.

### Варіант № 2

1. Обґрунтуйте в чому полягає особлива роль простору сприйняття респондентами запропонованих їм об'єктів.
2. Визначте сутність моделі ідеальної точки в багатовимірному випадку.

### Варіант № 3

1. Сформулюйте загальні принципи вивчення простору сприйняття як основного завдання БШ.
2. Обґрунтуйте в чому полягає особлива роль проблеми неметричного багатовимірного розгортання.

### Варіант № 4

1. Проаналізуйте ідеї Кумбса щодо урахування можливості упорядкування відстаней між об'єктами.
2. Визначте сутність виду вихідних даних при багатовимірному шкаліруванні.

### Варіант № 5

1. Обґрунтуйте в чому полягає проблема формального визначення близькості.
2. Проаналізуйте в чому полягає роль соціолога при отриманні даних, вихідних для багатовимірного шкалірування, та інтерпретації його результатів.

### Варіант № 6

1. Проаналізуйте процес створення функції відстані (аксіоматичне визначення).
2. Обґрунтуйте в чому полягають проблеми можливих способів одержання вихідних даних при багатовимірному шкаліруванні.

### Варіант № 7

1. Визначте особливості понять Евклідова відстань та Евклідовий простір.
2. Проаналізуйте особливості безпосереднього визначення близькості від респондентів, класифікація відповідних способів опитування; проблеми, що постають при такому способі збору даних.

### Варіант № 8

1. Сформулюйте загальні характеристики неявного порівняння відстаней між близькістю, яке закладене у формулі функції стресу для метричного шкалірування.
2. Визначте особливості прикладів розрахунку матриці близькості на основі аналізу достатньо надійних даних іншого роду при багатовимірному шкаліруванні.

### Варіант № 9

1. Обґрунтуйте в чому полягає особлива роль поняття монотонної регресії, що використовується при розрахунку функції стресу для неметричного шкалірування.
2. Проаналізуйте особливості використання формальних та неформальних методів при інтерпретації результатів багатовимірного шкалірування.

### Варіант № 10

1. Сформулюйте загальні характеристики важливості для соціології неметричного шкалірування.
2. Проаналізуйте значення змістовних концепцій дослідника при вирішенні проблем вибору розмірності евклідова простору і повороту його осей при багатовимірному шкаліруванні.

### Варіант № 11

1. Проведіть філософсько-соціологічний аналіз сутності формальних аспектів проблем розмірності евклідового простору і обертання, що визначають його осей координат.
2. Проаналізуйте в чому полягає особливості загального уявлення про моделювання частот таблиці спряженості при багатовимірному шкаліруванні.

### Варіант № 12

1. Сформулюйте загальні характеристики постановки завдання важливості врахування специфіки метрик окремих респондентів.
2. Визначте в чому полягає проблема мультиплікативних та адитивних моделей частот при багатовимірному шкаліруванні.

### Варіант № 13

1. Обґрунтуйте в чому полягає особлива роль визначення способу обліку метрик в індивідуальному БШ.
2. Проаналізуйте роль логарифмування мультиплікативної моделі при багатовимірному шкаліруванні.

### Варіант № 14

1. Визначте сутність виду вхідних і вихідних даних, функції стресу в індивідуальному БШ.
2. Обґрунтуйте в чому полягає можливість різного розуміння як сенсу розглянутих вкладів, так і того "середнього" рівня, з яким порівнюються спостерігаються частоти в процесі їх моделювання при багатовимірному шкаліруванні.

### Варіант № 15

1. Сформулюйте основні закономірності одномірного розгортання.
2. Проаналізуйте особливості використання поняття зв'язку між двома групами ознак при багатовимірному шкаліруванні.

**Контрольна робота № 2**

### Варіант № 1

1. Визначте причини відхилення спостережуваних частот від їхніх середніх значень, тобто відмінності реального розподілу від рівномірного при багатовимірному шкаліруванні.
2. Сформулюйте загальні характеристики важливості для соціолога вивчення відповідних зв'язків при багатовимірному шкаліруванні.

### Варіант № 2

1. Проаналізуйте особливості вибору моделів частот, що відповідають логлінейному аналізу.
2. Визначте принципову різницю між статистичним та причинним зв'язком в контексті багатовимірного шкалірування.

### Варіант № 3

1. Обґрунтуйте в чому полягає особлива роль насиченої моделі при багатовимірному шкаліруванні.
2. Визначте схеми обчислення ковариаций (кореляцій) між будь-якими двома ознаками на основі графа зв'язків в контексті багатовимірного шкалірування.

### Варіант № 4

1. Сформулюйте основні вимоги до визначення мети переходу до логарифмів частот.
2. Сформулюйте загальні характеристики вивчення статистичних зв'язків на основі причинних схем в контексті багатовимірного шкалірування.

### Варіант № 5

1. Визначте особливу роль гіпотези про взаємозв'язок ознак при багатовимірному шкаліруванні.
2. Сформулюйте загальні характеристики обчислення структурних коефіцієнтів причинних схем в контексті багатовимірного шкалірування.

### Варіант № 6

1. Проаналізуйте роль критерію "хі-квадрат" при використанні логлінейного аналізу.
2. Сформулюйте загальні характеристики основної теореми причинного аналізу в контексті багатовимірного шкалірування.

### Варіант № 7

1. Проаналізуйте в чому полягає особливості розрахунку коефіцієнтів логлінейной моделі для двовимірного випадку.
2. Сформулюйте загальні характеристики поняття допоміжної теорії вимірювань Блейлока в контексті багатовимірного шкалірування.

### Варіант № 8

1. Проаналізуйте особливості порівняння логлінейного аналізу з номінальним регресійним і дисперсійним аналізом, а також з методом послідовних розбивок.
2. Визначте роль і місце причинного аналізу як концептуального підходу до вивчення соціальних явищ в контексті багатовимірного шкалірування.

### Варіант № 9

1. Сформулюйте загальні характеристики різного 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.
2. Проаналізуйте проблеми формалізації завдання вивчення причинно-наслідкових відносин в соціології в контексті багатовимірного шкалірування.

### Варіант № 10

1. Визначте принципові 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.
2. Сформулюйте загальні характеристики агломератівних та дівізімних алгоритмів в контексті багатовимірного шкалірування

### Варіант № 11

1. Визначте сутність поняття причини в соціології в контексті багатовимірного шкалірування.
2. Сформулюйте загальні характеристики функції відстані, які відмінні від евклідової: зважене евклидово, сіті-блок, Махаланобіса, Хеммінгово.

### Варіант № 12

1. Визначте роль принципової неможливості повністю формалізувати поняття причини в контексті багатовимірного шкалірування.
2. Сформулюйте в чому полягає специфіка вирішення соціологічних завдань побудови типології за допомогою методів автоматичної класифікації в контексті багатовимірного шкалірування.

### Варіант № 13

1. Проаналізуйте форми графу причинних зв'язків в контексті багатовимірного шкалірування.
2. Обґрунтуйте в чому полягає загальне уявлення про завдання розпізнавання образів в контексті багатовимірного шкалірування.

### Варіант № 14

1. Проаналізуйте специфіку обчислення структурних коефіцієнтів в контексті багатовимірного шкалірування.
2. Визначте в чому полягає виділення завдань: пошук класів, опис класів, визначення найбільш ефективної системи ознак.

### Варіант № 15

1. Визначте принципи побудови часткових коефіцієнтів кореляції і регресії в контексті багатовимірного шкалірування.

**Тестові завдання з курсу «**МЕТОДИ БАГАТОВИМІРНОГО АНАЛІЗУ ТА BIG DATA В СОЦІОЛОГІЇ»

**1. Тестові завдання**

1. Оберіть правильне твердження:

а) нормальна випадкова величина ухиляється від свого середнього не більше, ніж на 2 кореня з дисперсії,

б) нормальна випадкова величина ухиляється від свого середнього не більше, ніж на 3 кореня з дисперсії; в) нормальна випадкова величина ухиляється від свого середнього не більше, ніж на 4 кореня з дисперсії.

1. Залежність виду Y = F (X) називається:

а) лінійна кореляція; б) лінійна регресія; в) часткова кореляція.

1. Скільки залежних змінних може бути в рівнянні регресії:

а) скільки завгодно; б) не більше 3; в) одна.

1. Не виконує завдання класифікації:

а) кластерний аналіз,

б) корреляційний аналіз;

в) дискримінантний аналіз.

1. Чи можливо обчислити коефіцієнт регресії Y на X, якщо через відомо коефіцієнт кореляції і середньоквадратичне відхилення:

а) не можливо, тому що необхідний показник дисперсії;

б) можливо;

в) залежить від виду аналізу.

1. До обмеження методу регресійного аналізу не належать:

а) нормальність розподілу ознак;

б) рівна кількість ознак змінних;

в) змінні відмінні від нуля.

1. До обмежень методу факторного аналізу не належить:

а) нормальність розподілу ознак;

б) рівна кількість ознак змінних;

в) рівність дисперсій.

1. До обмежень методу дисперсійного аналізу не належить:

а) нормальність розподілу ознак;

б) рівна кількість ознак змінних;

в) рівність дисперсій.

1. Задачу прогнозування не виконує:

а) дискримінантний аналіз;

б) факторний аналіз;

в) регресійний аналіз.

1. Для незалежних вибірок використовується:

а) дисперсійний аналіз з повторними змінами;

б) кореляційний аналіз;

в) однофакторний дисперсійний аналіз.

1. Таке значення змінної менше якої мають рівно половина цих змінних, називається ...

а) мода б) середнє

в) медіана г) проміжне

1. Чому дорівнює кореляція Спірмена і Кендалла для двох змінних:

|  |  |  |  |
| --- | --- | --- | --- |
| № | X | Y | а) 0  б) -1  в) 1 г) 0,5 |
| 1 | 32 | 2 |
| 2 | 16 | 6 |
| 3 | 20 | 4 |
| 4 | 8 | 10 |
| 5 | 11 | 8 |

1. Чи може одне і те ж чисельне значення кореляції для різних вибірок мати різну статистичну значущість?

а) Так б) Ні

**2. Допишіть речення**

1. Дві вибірки є залежними, якщо …
2. Основний спосіб забезпечення репрезентативності вибірки відносно генеральної сукупності це …
3. Основна властивість вибірки, що визначає її якість – це …
4. Медіана як міра центральної тенденції не придатна для змінних в наступних шкалах …
5. Середнє як міра центральної тенденції не придатна для змінних в наступних шкалах …
6. Як співвідносяться дисперсії двох рядів чисел: 1) 5, 8, 10, 12, 11 і 2) 1, 4, 6, 8, 7 …
7. Гомогенність (рівність) дисперсій перевіряється перед …
8. Якщо при перевірці статистичної достовірності кореляції (при α = 0,05) р> 0.1, то коректний висновок, що …
9. Для перевірки достовірності відмінності двох незалежних груп, члени яких ранжовані за ступенем вираженості «тривожності», використовують критерій
10. Для перевірки достовірності відмінності двох повторних вимірювань, члени яких ранжовані за ступенем вираженості «тривожності», використовують критерій …
11. Для перевірки достовірності відмінностей студентів 1 і 5 курсів за змінної «сімейний стан» (неодружений - ні) слід застосувати критерій …
12. Для перевірки достовірності відмінностей 2-х вибірок по перемінної "стать" (чоловік - дружин) слід застосувати критерій …
13. Для порівняння викладачів та студентів з «домінантності» (метрична шкала), слід застосувати критерій …
14. Якщо необхідно порівняти два повторних вимірювання кількісної змінної, що має помітні викиди, то застосовують критерій …
15. Для перевірки достовірності відмінності двох залежних вибірок по змінній, яка вимірюється в ранговій шкалі, застосовують критерій …
16. Для перевірки достовірності відмінності старших (1-я вибірка) і їхніх молодших (2-я вибірка) братів за рівнем домінантності, виміряної в метричній шкалі, застосовують критерій
17. Гіпотезу про взаємозв'язок читача (2 градації) і метричної змінних доцільно перевіряти за допомогою …
18. Статистична значимість поліпшення стану (рангова шкала) до і після терапії визначається за критерієм …
19. Гіпотезу про взаємозв'язок рангової і номинативной змінної, що має дві градації (напр., стать), доцільно вивчати за допомогою критерію …
20. Для перевірки гіпотези про відмінність 2 груп за ступенем індивідуальної мінливості (дисперсії) застосовують критерій …
21. Гіпотезу про взаємозв'язок метричної та номинативной змінної, що має дві градації (напр., стать), доцільно вивчати за допомогою критерію …
22. Гіпотезу про взаємозв'язок метричної та номинативной змінної, що має 5 градацій (наприклад, хобі), доцільно перевіряти за допомогою критерію …
23. Гіпотезу про взаємозв'язок порядкової і номинативной змінної, що має 4 градації (напр., посада), доцільно вивчати за допомогою критерію …
24. Гіпотезу про взаємозв'язок метричної і порядкової змінної, що має 15 градацій, доцільно вивчати за допомогою
25. Гіпотезу про взаємозв'язок 2-х кількісних змінних, що мають помітні викиди (асиметрії) доцільно перевіряти за допомогою
26. Для перевірки гіпотези про взаємозв'язок однієї метричної змінної та двох номінативних змінних доцільно застосовувати:
27. Для множинного порівняння середніх в рамках дисперсійного аналізу застосовують …
28. Метод множинного порівняння середніх який вимагає попереднього отримання статистично значимого результату дисперсійного аналізу це …
29. Багатовимірний дисперсійний аналіз призначений для вивчення впливу …
30. Модель багатовимірного дисперсійного аналізу включає …
31. Багатовимірний етап багатовимірного дисперсійного аналізу припускає-лага перевірку гіпотез …
32. Одновимірна етап багатовимірного дисперсійного аналізу проводиться для …
33. Дисперсійний аналіз з повторними вимірами дозволяє вивчать вплив на залежні змінні …
34. Різниця двох методів класифікації полягає в тому, що в першому задано число класів і приналежність деяких об'єктів до цих класів, а в другому – не задано ні те ні інше …
35. Подібність двох багатовимірних методів полягає в тому, що аналізуються кореляції між ознаками …
36. Частина дисперсії «залежною» змінної, обумовлена впливом «незалежних» змінних – це ...
37. Якщо незалежна змінна х в множині регресійному аналізі корелює з іншими незалежними змінними, то її внесок в дисперсію залежної змінної ...
38. Якщо незалежна змінна х в множині регресійному аналізі не корелює з іншими незалежними змінними, то її внесок в оцінку залежної змінної ...
39. Якщо в багатовимірному регресійному аналізі (y - залежна змінна, x1, x2 - незалежні змінні) r12 = 0,4; r1y = 0,8; r2y = - 0,5; β1 = 0,5; β2 = - 0, 2, то коефіцієнт множинної детермінації R2 дорівнює …
40. Метод «повної зв'язку» («далекого сусіда») в кластерному аналізі, в порівнянні з методом «одиночній зв'язку» («найближчого сусіда») дає ...
41. Метод «середньої зв'язку» (СЗ) в порівнянні з методами «далекого сусіда» (ДС) і «ближнього сусіда» (БС) зазвичай дозволяє отримати число кластерів …
42. Ієрархічний кластерний аналіз за (N - 1) кроків кластеризації (N - число об'єктів кластеризації) дає об'єднання …
43. Статистична значимість внеску кожної змінної на відміну класів визначається ...
44. Показником належності об'єкта до класу є ...
45. Основним заходом якості рішення в багатовимірному шкалировании є ...
46. Для попередньої оцінки числа факторів у факторному аналізі використовують ...
47. Величина факторної навантаження даної змінної з даного фактору свідчить ...

**Питання, що виносяться на екзамену з курсу «**МЕТОДИ БАГАТОВИМІРНОГО АНАЛІЗУ ТА BIG DATA В СОЦІОЛОГІЇ»

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.