МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
МІЖНАРОДНИЙ ГУМАНІТАРНИЙ УНІВЕРСИТЕТ
Кафедри комп’ютерних наук та інформаційних технологій

Григор’єва Т.І., Соловська І.М.

СУЧАСНІ ПРОБЛЕМИ ТЕОРІЇ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНОЇ СТАТИСТИКИ

Методичні рекомендації для самостійної роботи здобувачів

Одеса 2024

Затверджено Вченою Радою Міжнародного гуманітарного університету (протокол № 13 від 16.08.2024 р.)

Григор'єва Т.І., Соловська І.М.
Сучасні проблеми теорії ймовірностей та математичної статистики: методичні рекомендації для самостійної роботи здобувачів [Електронне видання]. / Григор'єва Т.І., Соловська І.М. Кафедри комп'ютерних наук та інформаційних технологій Міжнародного гуманітарного університету. Одеса, 2024. – 47 с.

Методичні рекомендації з курсу «Сучасні проблеми теорії ймовірностей та математичної статистики» розроблено відповідно до навчального плану, вони складаються з навчальної програми курсу, методичних рекомендацій з проведення практичних занять і завдань для самостійної роботи здобувачів, списку рекомендованої літератури. Матеріали призначено для студентів Міжнародного гуманітарного університету, які в магістратурі вивчають галузь знань - 11 Математика та статистика.

Вивчення дисципліни «Сучасні проблеми теорії ймовірностей та математичної статистики» сприятиме залученню здобувачів до науково-дослідницької діяльності та підготовки ними публікацій, кваліфікаційних й інших наукових робіт.

ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

<table>
<tr><th rowspan="2">Найменування показників</th><th rowspan="2">Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень</th><th>Характеристика навчальної дисципліни</th></tr>
<tr><th>денна форма навчання</th></tr>
<tr><td rowspan="7">Кількість кредитів – 4

Загальна кількість годин – 120</td><td>Галузь - 11 Математика та статистика</td><td>обов’язкова</td></tr>
<tr><td rowspan="6">Спеціальність – 112 Статистика</td><td>Рік підготовки:</td></tr>
<tr><td>1-й</td></tr>
<tr><td>Семестр</td></tr>
<tr><td>2-й</td></tr>
<tr><td>Лекції</td></tr>
<tr><td rowspan="9">Мова навчання – українська</td><td rowspan="9">Рівень вищої освіти – другий (магістерський) рівень</td><td>14 год.</td></tr>
<tr><td>Практичні, семінарські</td></tr>
<tr><td>26 год.</td></tr>
<tr><td>Лабораторні</td></tr>
<tr><td>-</td></tr>
<tr><td>Самостійна робота та індивідуальні завдання</td></tr>
<tr><td>80 год.</td></tr>
<tr><td>Вид контролю:</td></tr>
<tr><td>залік</td></tr>
</table>

Дисципліна "Сучасні проблеми теорії ймовірностей та математичної статистики" вивчає новітні теоретичні та практичні аспекти, які виникають у процесі розвитку цих галузей математики. Основна увага приділяється розробці й аналізу нових методів, що застосовуються для моделювання випадкових процесів, оцінки ймовірностей складних подій, а також вивченню статистичних методів, які використовуються для аналізу даних.

Метою викладення дисципліни "Сучасні проблеми теорії ймовірностей та математичної статистики" є формування у студентів глибокого розуміння сучасних концепцій і методів теорії ймовірностей та математичної статистики, а також розвиток навичок їх застосування для аналізу складних випадкових процесів та статистичних даних. Дисципліна спрямована на ознайомлення студентів з новітніми досягненнями в цих галузях та підготовку їх до проведення самостійних досліджень та розв'язання практичних задач у науці, техніці, економіці та інших сферах.

Передумови для вивчення дисципліни є знання і вміння, отримані студентом при вивченні навчальних дисциплін бакалаврської підготовки.

ЗАПЛАНОВАНІ РЕЗУЛЬТАТИ НАВЧАННЯ ЗА НАВЧАЛЬНОЮ ДИСЦИПЛІНОЮ

Знання:

1. Студенти мають знати основні сучасні теорії та моделі теорії ймовірностей та математичної статистики.
2. Знання основних методів і підходів, що використовуються для аналізу випадкових процесів і статистичних даних, включаючи стохастичні процеси, регресійний аналіз, методи кореляційного аналізу.
3. Знання про новітні дослідження та тренди в галузі теорії ймовірностей та математичної статистики, а також їх застосування в різних галузях.

Уміння:

1. Уміння застосовувати отримані теоретичні знання для аналізу випадкових процесів та статистичних даних.
2. Уміння будувати математичні моделі для прогнозування поведінки випадкових процесів, а також оцінювати ризики та ймовірності різних подій.
3. Уміння інтерпретувати результати статистичного та ймовірнісного аналізу, робити відповідні висновки та рекомендації.
4. Уміння розв'язувати практичні задачі, використовуючи методи теорії ймовірностей та математичної статистики.

Навички:

1. Розвиток навичок критичного мислення для аналізу наукових статей, даних і моделей, що дозволяє оцінювати їх валідність і надійність.
2. Навички проведення наукових досліджень, формулювання гіпотез, проведення аналізу результатів.
3. Навички ефективної комунікації отриманих результатів, включаючи підготовку наукових доповідей, статей, презентацій.
4. Навички інтеграції знань і методів теорії ймовірностей та математичної статистики в різні галузі, що дозволяє застосовувати їх у міждисциплінарних дослідженнях.

ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Тема 1. Випадкові ряди та граничні теореми

Закони великих чисел. Центральні граничні теореми. Розподіли сум випадкових рядів. Сингулярні розподіли та теорема Джессена-Вінтнера.

Тема 2. Випадкові процеси

Закони розподілу випадкових процесів. Характеристики випадкових процесів. Математичне сподівання, дисперсія і кореляційна функція випадкового процесу. Нормована кореляційна функція випадкового процесу. Взаємна кореляційна функція двох стохастичних процесів. Перетворення випадкових процесів. Додавання, диференціювання та інтегрування стохастичних процесів.

Тема 3. Статистичні гіпотези

Емпірична функція розподілу. Теорема Глівенко. Критерій Коломогорова-Смирнова. Критерій Пірсона. Квантилі та процентні точки нормального розподілу.

Тема 4. Дисперсійний аналіз. Кореляційний та регресійний аналіз

Моделі: випадкова, детермінована, змішана. Однофакторний і двофакторний дисперсійний аналіз. Завдання кореляційного і регресійного аналізу. Парна лінійна регресійна модель. Рангова кореляція. Нелінійна парна кореляція. Множинна кореляція і регресія. Побудова регресійних моделей.

Тема 5. Статистичний аналіз даних

Класифікація методів статистичного аналізу. Описова статистика. Частотний розподіл. Відсоткові показники. Заходи центральної тенденції. Міри розкиду даних. Методи вторинної статистичної обробки результатів. Параметричні та непараметричні статистичні методи.

СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назви змістових модулів і тем	Кількість годин									
	Денна форма					Заочна форма				
	Усього	у тому числі				Усього	у тому числі			
		Лекц.	Прак.	Лаб.	Сам. роб.		Лекц.	Прак.	Лаб.	Сам. роб.
Тема 1. Випадкові ряди та граничні теореми	16	2	4		10	16				16
Тема 2. Випадкові процеси	26	4	4		10	26	2			24
Тема 3. Статистичні гіпотези	30	4	6		20	30	2			28
Тема 4. Дисперсійний аналіз. Кореляційний та регресійний аналіз	28	2	6		20	28		2		26
Тема 5. Статистичний аналіз даних	30	2	6		20	30		2		28
Усього годин	**120**	**14**	**26**	**-**	**80**	**120**	**4**	**4**	**-**	**112**
ПІДСУМКОВИЙ КОНТРОЛЬ – ЗАЛІК										

ПИТАННЯ ДО ПРАКТИЧНИХ ЗАНЯТЬ

Тема 1. Випадкові ряди та граничні теореми

Закони великих чисел. Центральні граничні теореми. Розподіли сум випадкових рядів. Сингулярні розподіли та теорема Джессена-Вінтнера.

Тема 2. Випадкові процеси

Елементарні випадкові функції. Криві реалізацій випадкової функції. Математичне сподівання випадкового процесу. Дисперсія випадкового процесу. Кореляційна функція випадкового процесу. Нормована кореляційна функція випадкового процесу. Взаємна кореляційна функція двох стохастичних процесів. Нормована взаємна кореляційна функція двох стохастичних процесів. Додавання стохастичних процесів. Диференціювання стохастичного процесу. Інтегрування стохастичного процесу.

Тема 3. Статистичні гіпотези

Емпірична функція розподілу. Поняття статистичної значимості (p-level). Теорема Глівенко. Критерій Коломогорова-Смирнова. Критерій Пірсона. Квантилі та процентні точки нормального розподілу.

Тема 4. Дисперсійний аналіз. Кореляційний та регресійний аналіз

Моделі: випадкова, детермінована, змішана. Однофакторний і двофакторний дисперсійний аналіз. Завдання кореляційного і регресійного аналізу. Вихідні передумови регресійного аналізу і властивості оцінок. Парна лінійна регресійна модель. Рангова кореляція. Нелінійна парна кореляція. Множинна кореляція і регресія. Завдання і проблеми кореляційного аналізу. Двовимірна кореляційний модель. Тривимірна кореляційний модель. Методи оцінки кореляційних моделей. Перевірка

значущості множинного рівняння регресії. Методика побудови математичних моделей часових рядів. Аналіз процесу. Приклади побудови моделей за статистичними даними.

Тема 5. Статистичний аналіз даних

Класифікація методів статистичного аналізу. Описова статистика. Частотний розподіл. Відсоткові показники. Заходи центральної тенденції. Міри розкиду даних. Методи вторинної статистичної обробки результатів. Параметричні та непараметричні статистичні методи. Методології аналізу та моделювання прикладної предметної області, виявлення інформаційних потреб і збору вихідних даних для статистичного дослідження процесів її функціювання. Програмні засоби статистичних досліджень. Використання результатів статистичної обробки.

САМОСТІЙНА РОБОТА

До самостійної роботи студентів включаються:

1. Знайомство з науковою та навчальною літературою відповідно зазначених у програмі тем.
2. Опрацювання лекційного матеріалу.
3. Підготовка до практичних занять.
4. Консультації з викладачем протягом семестру.
5. Самостійне опрацювання окремих питань навчальної дисципліни.
6. Підготовка до підсумкового контролю.

Тематика та питання до самостійної підготовки та індивідуальних завдань

Тема 1. Випадкові ряди та граничні теореми

Закони великих чисел. Центральні граничні теореми. Розподіли сум випадкових рядів. Сингулярні розподіли та теорема Джессена-Вінтнера. Закон повторного логарифма. Моделювання, оцінювання ризиків і прийняття рішень.

Тема 2. Випадкові процеси

Класифікація випадкових процесів. Характеристики випадкових процесів. Додавання стохастичних процесів. Диференціювання стохастичного процесу. Інтегрування стохастичного процесу. Стаціонарні та нестаціонарні випадкові процеси.

Тема 3. Статистичні гіпотези

Вибір тестової статистики. Обчислення тестової статистики та p-значення. Теорема Глівенко. Критерій Коломогорова-Смирнова. Критерій Пірсона. Квантилі та процентні точки нормального розподілу.

Тема 4. Дисперсійний аналіз. Кореляційний та регресійний аналіз

Вихідні передумови регресійного аналізу і властивості оцінок. Парна лінійна регресійна модель. Рангова кореляція. Нелінійна парна кореляція. Двовимірна кореляційний модель. Тривимірна кореляційний модель. Методи оцінки кореляційних моделей. Перевірка значущості множинного рівняння регресії. Кореляційний аналіз порядкових величин. Порядкові випадкові величини і ранги. Ранговий коефіцієнт кореляції Спірмена. Ранговий коефіцієнт кореляції Кендала.

Тема 5. Статистичний аналіз даних

t - критерій Стьюдента. Дисперсійний аналіз (ANOVA). U критерій Манна-Уїтні. Критерій серій Вальда-Вольфовиця. Двовибірковий критерій Колмогорова-Смірнова. Ранговий дисперсійний аналіз Фрідмана. Q критерій Кохрена. Коефіцієнт кореляції Браве-Пірсона. Коефіцієнт кореляції рангів Спірмена. Роль і значення графічного методу в статистиці.

ТЕМИ ДОПОВІДЕЙ

1. Закони великих чисел та їх сучасне застосування
2. Центральна гранична теорема та її застосування в машинному навчанні
3. Сингулярні розподіли: теорема Джессена-Вінтнера та нові напрямки досліджень

4. Порівняння та огляд критеріїв Колмогорова-Смирнова та Пірсона, їх застосування в сучасних дослідженнях, переваги та обмеження.
5. Сучасні проблеми оцінки ризиків за допомогою ймовірнісних методів
6. Методи оцінки ризиків у фінансових ринках, страхуванні та інших галузях.
7. Використання ймовірнісних підходів для моделювання ризикових подій.
8. Байєсівська статистика та її застосування в сучасній науці
9. Регресійний аналіз в кібербезпеці
10. Застосування регресійного аналізу для виявлення аномалій та прогнозування кіберзагроз.
11. Використання стохастичного програмування в управлінні складними системами
12. Теорія імовірностей в аналізі великих даних
13. Моделювання випадкових процесів у фінансовій сфері
14. Принципи функціонування та моделювання природничих, економічних та соціальних процесів.
15. Побудова математичних моделей систем і явищ з елементами випадковості
16. Імовірнісні розподіли, що застосовуються в прикладних сферах досліджень.
17. Приклади теоретичних та практичних задач з різних галузей, які розв’язуються статистичними методами
18. Приклади статистичної обробки даних економічного явища (процесу).
19. Програмні засоби статистичних досліджень

ВИДИ ТА МЕТОДИ КОНТРОЛЮ

Робоча програма навчальної дисципліни передбачає наступні види та методи контролю:

Види контролю	Складові оцінювання
поточний контроль, який здійснюється у ході: проведення практичних занять, виконання індивідуального завдання; проведення консультацій та відпрацювань.	**50%**
підсумковий контроль, який здійснюється у ході проведення заліку.	**50%**

Методи діагностики знань (контролю)	фронтальне опитування; наукова доповідь, усне повідомлення, індивідуальне опитування, робота у групах, розв`язання задач і практичних завдань, залік

Питання до заліку

1. Закони великих чисел. Центральні граничні теореми.
2. Розподіли сум випадкових рядів.
3. Сингулярні розподіли та теорема Джессена-Вінтнера.
4. Випадкові процеси. Закони розподілу випадкових процесів.
5. Математичне сподівання випадкового процесу. Дисперсія випадкового процесу.
6. Кореляційна функція випадкового процесу. Властивості.
7. Нормована кореляційна функція випадкового процесу.
8. Взаємна кореляційна функція двох стохастичних процесів. Нормована взаємна кореляційна функція двох стохастичних процесів.

9. Перетворення випадкових процесів. Додавання, диференціювання та інтегрування стохастичних процесів.

10. Теорема Глівенко.

11. Критерій Коломогорова-Смирнова. Критерій Пірсона.

12. Квантилі та процентні точки нормального розподілу.

13. Однофакторний і двофакторний дисперсійний аналіз.

14. Завдання кореляційного і регресійного аналізу.

15. Вихідні передумови регресійного аналізу і властивості оцінок.

16. Парна лінійна регресійна модель.

17. Рангова кореляція.

18. Нелінійна парна кореляція.

19. Множинна кореляція і регресія.

20. Завдання і проблеми кореляційного аналізу.

21. Двовимірна кореляційний модель.

22. Тривимірна кореляційний модель.

23. Методи оцінки кореляційних моделей.

24. Перевірка значущості множинного рівняння регресії.

25. Методика побудови математичних моделей часових рядів.

26. Приклади побудови моделей за статистичними даними.

27. Класифікація методів статистичного аналізу.

28. Описова статистика.

29. Частотний розподіл.

30. Відсоткові показники.

31. Заходи центральної тенденції. Міри розкиду даних.

32. Методи вторинної статистичної обробки результатів.

33. Параметричні та непараметричні статистичні методи.

КРИТЕРІЇ ПІДСУМКОВОЇ ОЦІНКИ ЗНАНЬ СТУДЕНТІВ

(для іспиту / заліку)

Рівень знань оцінюється:

- «відмінно» / «зараховано» А - від 90 до 100 балів. Студент виявляє особливі творчі здібності, вміє самостійно знаходити та опрацьовувати необхідну інформацію, демонструє знання матеріалу, проводить узагальнення і висновки. Був присутній на лекціях та семінарських заняттях, під час яких давав вичерпні, обґрунтовані, теоретично і практично правильні відповіді, має конспект з виконаними завданнями до самостійної роботи, проявляє активність і творчість у науково-дослідній роботі;

- «добре» / «зараховано» В - від 82 до 89 балів. Студент володіє знаннями матеріалу, але допускає незначні помилки у формуванні термінів, категорій, проте за допомогою викладача швидко орієнтується і знаходить правильні відповіді. Був присутній на лекціях та семінарських заняттях, має конспект з виконаними завданнями до самостійної роботи, проявляє активність і творчість у науково-дослідній роботі;

- «добре» / «зараховано» С - від 74 до 81 балів. Студент відтворює значну частину теоретичного матеріалу, виявляє знання і розуміння основних положень, з допомогою викладача може аналізувати навчальний матеріал, але дає недостатньо обґрунтовані, невичерпні відповіді, допускає помилки. При цьому враховується наявність конспекту з виконаними завданнями до самостійної роботи та активність у науково-дослідній роботі;

- «задовільно» / «зараховано» D - від 64 до 73 балів. Студент був присутній не на всіх лекціях та семінарських заняттях, володіє навчальним матеріалом на середньому рівні, допускає помилки, серед яких є значна кількість суттєвих. При цьому враховується наявність конспекту з виконаними завданнями до самостійної роботи;

- «задовільно» / «зараховано» Е - від 60 до 63 балів. Студент був присутній не на всіх лекціях та семінарських заняттях, володіє навчальним матеріалом на рівні, вищому за початковий, значну частину його відтворює на репродуктивному рівні, на всі запитання дає необґрунтовані, невичерпні відповіді, допускає помилки, має

неповний конспект з завданнями до самостійної роботи.

- «незадовільно з можливістю повторного складання» / «не зараховано» FX – від 35 до 59 балів. Студент володіє матеріалом на рівні окремих фрагментів, що становлять незначну частину навчального матеріалу.

- «незадовільно з обов'язковим повторним вивченням дисципліни» / «не зараховано» F – від 1 до 34 балів. Студент не володіє навчальним матеріалом.

Таблиця відповідності результатів контролю знань за різними шкалами

100-бальною шкалою	**Шкала за ECTS**	**За національною шкалою**	
		екзамен	**залік**
90-100 (10-12)	A	Відмінно	зараховано
82-89 (8-9)	B	Добре	
74-81(6-7)	C		
64-73 (5)	D	Задовільно	
60-63 (4)	E		
35-59 (3)	Fx	незадовільно	не зараховано
1-34 (2)	F		

ПЛАН – КОНСПЕКТ ЛЕКЦІЙ ДИСЦИПЛІНИ

СУЧАСНІ ПРОБЛЕМИ ТЕОРІЇ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНОЇ СТАТИСТИКИ

Тема 1. Випадкові ряди та граничні теореми

Випадкові ряди та граничні теореми є ключовими темами в теорії ймовірностей і математичній статистиці, які використовуються для аналізу поведінки суми незалежних випадкових величин та їхніх властивостей у великому масштабі.

Випадкові ряди

Випадковий ряд — це сума послідовності випадкових величин. Якщо $X_1, X_2,\ldots, X_n$ — це послідовність незалежних і однаково розподілених випадкових величин, то випадковий ряд має вигляд: $S_n = X_1 + X_2 +\ldots+X_n$.

Ці ряди використовуються для моделювання різних випадкових процесів, таких як доходи від інвестицій, погрішності вимірювань, або сума виграшів у серії ігор.

Основні поняття:

1. **Випадкові змінні**. Випадковий ряд складається з випадкових змінних $X_1, X_2,\ldots, X_n$ кожна з яких може приймати різні значення відповідно до певного розподілу ймовірностей.
2. **Незалежні випадкові змінні.** Випадкові змінні в ряді можуть бути незалежними, тобто ймовірність настання подій, пов'язаних з однією змінною, не залежить від подій, пов'язаних з іншими змінними.
3. **Математичне сподівання та дисперсія**. Для кожної випадкової змінної можна визначити математичне сподівання (середнє значення) і дисперсію (міра розсіювання значень навколо середнього).

Властивості випадкових рядів:

1. **Збіжність**. Випадковий ряд може збігатися до певного значення за ймовірністю, майже напевно або в середньому квадраті. Це поняття важливе для аналізу довгострокової поведінки ряду.
2. **Закон великих чисел.** Цей закон стверджує, що середнє арифметичне великої кількості незалежних і однаково розподілених випадкових змінних буде наближатися до їхнього математичного сподівання.
3. **Центральна гранична теорема (ЦГТ)**. Згідно з ЦГТ, сума великої кількості незалежних випадкових змінних із однаковим розподілом наближається до нормального розподілу незалежно від початкового розподілу цих змінних.

Випадкові ряди використовуються для моделювання та аналізу різних стохастичних процесів.

Сфери використання:

1. **Фінансові ринки**. Випадкові ряди моделюють коливання цін на фінансових ринках. Наприклад, ціни акцій часто описуються випадковими рядами.
2. **Фізичні процеси**. Випадкові ряди використовуються для моделювання природних процесів, таких як коливання температури або радіоактивний розпад.
3. **Економіка**. Аналіз часових рядів, наприклад, ВВП, рівня безробіття, споживчих витрат, часто базується на вивченні випадкових рядів.

Граничні теореми

Граничні теореми розглядають поведінку суми випадкових величин, коли кількість доданків прямує до нескінченності. Основні граничні теореми включають **Закон великих чисел** та **Центральну граничну теорему**.

Закон великих чисел стверджує, що середнє арифметичне великої кількості незалежних і однаково розподілених випадкових величин наближається до

математичного сподівання цих величин при збільшенні кількості доданків. **Закон великих чисел** — це фундаментальна теорема теорії ймовірностей, яка описує поведінку середнього значення випадкових величин при збільшенні кількості спостережень. Цей закон стверджує, що в середньому, при достатньо великій кількості незалежних випробувань, середнє значення випадкової величини збігається з її математичним очікуванням.

Різні формулювання закону великих чисел

Слабкий закон великих чисел

Слабкий закон великих чисел стверджує, що для великої кількості спостережень середнє значення вибірки збігається за ймовірністю з математичним очікуванням.

Формулювання: Нехай $X_1, X_2, \ldots, X_n$ — послідовність незалежних однаково розподілених випадкових величин з однаковим математичним очікуванням E[Xi]=μ і дисперсією Var(Xi)= σ^2. Тоді, для будь-якого є>0,

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| \geq \epsilon\right) = 0$$

Це означає, що середнє арифметичне послідовності випадкових величин збігається за ймовірністю до математичного очікування μ.

Тобто для послідовності $X_1, X_2, \ldots, X_n$ незалежних і однаково розподілених випадкових величин із математичним сподіванням E[X] та дисперсією Var(X), середнє арифметичне

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

збігається до математичного сподівання з імовірністю 1 при n→∞.

Сильний закон великих чисел

Сильний закон великих чисел говорить про те, що середнє значення вибірки збігається до математичного очікування майже напевно (зі ймовірністю 1).

Формулювання: Нехай $X_1, X_2, \ldots, X_n$ — послідовність незалежних однаково розподілених випадкових величин з однаковим математичним очікуванням E[Xi]=μ. Тоді,

$$P\left(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \mu\right) = 1$$

Це означає, що майже напевно середнє значення випадкових величин збігається з математичним очікуванням.

Закон великих чисел Бернуллі

Це класичне формулювання закону великих чисел, яке застосовується до послідовності незалежних випробувань з двома можливими результатами (успіх і неуспіх), де ймовірність успіху залишається постійною.

Формулювання: Нехай $X_1, X_2, \ldots, X_n$ — послідовність незалежних випробувань Бернуллі з ймовірністю успіху ppp. Тоді частота успіхів у цих випробуваннях збігається за ймовірністю до ймовірності успіху p:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = p$$

Це означає, що при великій кількості випробувань частота успіху наближається до істинної ймовірності успіху p.

Таким чином, закон великих чисел підтверджує, що при великій кількості спостережень середні значення випадкових величин на практиці будуть близькі до їх математичного очікування. Різні формулювання закону великих чисел розглядають

це зближення з різних точок зору, залежно від типу збіжності (за ймовірністю або майже напевно).

Закон повторного логарифма

Закон повторного логарифма уточнює характер збіжності для суми випадкових величин у Законі великих чисел і встановлює точні межі коливань цієї суми.

Формулювання:

Для незалежних і однаково розподілених випадкових величин із нульовим математичним сподіванням і кінцевою дисперсією:

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = \sigma$$

з ймовірністю 1,

де $S_n = X_1 + X_2 + \ldots + X_n$— сума перших n доданків, а σ^2— дисперсія випадкових величин.

Центральна гранична теорема

Центральна гранична теорема стверджує, що незалежно від початкового розподілу випадкових величин, сума великої кількості незалежних і однаково розподілених випадкових величин наближається до нормального розподілу, якщо кількість доданків зростає.

Формулювання:

Якщо $X_1, X_2, \ldots, X_n$ — незалежні і однаково розподілені випадкові величини з математичним сподіванням μ і дисперсією σ^2, то сума $S_n = X_1 + X_2 + \ldots + X_n$ нормально розподілена з параметрами nμ та nσ^2 при n→∞.

Ця теорема є основою для багатьох статистичних методів, зокрема для оцінки параметрів і перевірки гіпотез.

Розподіли сум випадкових рядів

Розподіли сум випадкових рядів досліджують характеристики сум випадкових величин, які можуть приймати різні форми залежно від властивостей складових випадкових величин.

Випадкові ряди та граничні теореми є фундаментальними інструментами в теорії ймовірностей. Вони дають змогу аналізувати поведінку складних систем із великою кількістю випадкових компонентів, що важливо для моделювання, оцінювання ризиків і прийняття рішень у різних галузях науки та інженерії.

Сингулярні розподіли та теорема Джессена-Вінтнера

Розглянемо випадки, коли розподіл суми випадкових величин є сингулярним, тобто не може бути описаний стандартними (дискретними або абсолютно неперервними) розподілами.

Сингулярні розподіли та теорема Джессена-Вінтнера є важливими поняттями в теорії ймовірностей та випадкових процесів.

Сингулярні розподіли — це особливий клас ймовірнісних розподілів, які не є ні абсолютно неперервними, ні дискретними. Вони мають властивості, характерні як для неперервних, так і для дискретних розподілів, але водночас не належать жодному з цих класів. Сингулярні розподіли можуть виникати в різних контекстах, наприклад, при розгляді ймовірнісних мір на множинах з фрактальною структурою.

Властивості сингулярних розподілів:

1. Сингулярний розподіл має нульову міру на кожному одиничному інтервалі, але повністю покриває ймовірнісний простір.

2. Функція розподілу сингулярного розподілу є неперервною, але має похідну, що дорівнює нулю майже всюди.
3. Багато сингулярних розподілів мають фрактальні властивості, наприклад, знаменита крива Коха або крива Кантора.

Приклад сингулярного розподілу

Класичний приклад сингулярного розподілу — це **розподіл Кантора**. Його функція розподілу є прикладом неперервної функції, яка не має похідної майже всюди і зростає з нульової швидкістю на проміжках.

Теорема Джессена-Вінтнера

Теорема Джессена-Вінтнера є важливим результатом у теорії ймовірностей і описує умови, за яких характер функції розподілу ймовірності (абсолютно неперервний, сингулярний або дискретний) може бути визначений з урахуванням характеристик випадкової величини.

Суть теореми Джессена-Вінтнера

Теорема стверджує, що для послідовності незалежних і однаково розподілених випадкових величин існують три можливі типи асимптотичного поведінки:

1. **Абсолютно неперервний розподіл**. Якщо функція розподілу є абсолютно неперервною, вона має похідну, яка є щільністю розподілу ймовірності.
2. **Дискретний розподіл**. Якщо функція розподілу є дискретною, її можна виразити у вигляді кусково-постійної функції.
3. **Сингулярний розподіл**. В інших випадках розподіл може бути сингулярним, тобто не належати ні до абсолютно неперервного, ні до дискретного типу.

Приклад застосування теореми

Розглянемо випадкову величину, яка є результатом лінійної комбінації незалежних випадкових величин з різними розподілами. Залежно від властивостей

цих випадкових величин, згідно з теоремою Джессена-Вінтнера, результат може бути абсолютно неперервним, дискретним або сингулярним розподілом.

Сингулярні розподіли і теорема Джессена-Вінтнера відіграють важливу роль у розумінні різноманітності можливих ймовірнісних розподілів, які можуть виникати в складних системах. Вони дають уявлення про те, як властивості елементів системи можуть впливати на її загальний розподіл, і надають інструменти для аналізу ситуацій, коли класичні підходи до моделювання ймовірностей виявляються недостатніми.

Тема 2. Випадкові процеси

Випадкові процеси – це математичні моделі, що описують еволюцію систем, які зазнають впливу випадкових факторів з часом. Вони знаходять застосування у різних галузях, таких як фінанси, фізика, інженерія, економіка та багато інших.

Закони розподілу випадкових процесів

Закон розподілу випадкового процесу визначає ймовірності різних станів, які цей процес може прийняти в різні моменти часу. Якщо розглядається кілька випадкових величин, то їхній спільний розподіл називається **багатовимірним розподілом**.

Характеристики випадкових процесів

1. **Математичне сподівання** випадкового процесу X(t) в момент часу t визначається як:

$$\mathbb{E}[X(t)] = \int_{-\infty}^{\infty} x \cdot f_{X(t)}(x)\, dx$$

де $f_{X(t)}(x)$ – функція щільності ймовірності випадкової величини X(t).

2. **Дисперсія** випадкового процесу визначається як міра розсіювання його значень навколо математичного сподівання:

$$\mathrm{Var}[X(t)] = \mathbb{E}[(X(t) - \mathbb{E}[X(t)])^2]$$

3. **Кореляційна функція** визначає ступінь зв'язку між значеннями процесу в різні моменти часу:

$$R_X(t_1, t_2) = \mathbb{E}[X(t_1) \cdot X(t_2)]$$

   Кореляційна функція є важливим інструментом у аналізі залежностей у часі для випадкових процесів.

4. **Нормована кореляційна функція**:

$$\rho_X(t_1, t_2) = \frac{R_X(t_1, t_2)}{\sqrt{R_X(t_1, t_1) \cdot R_X(t_2, t_2)}}$$

   Вона має значення в діапазоні від -1 до 1 і показує, наскільки сильно значення процесу в моменти часу t_1 і t_2 пов'язані між собою.

5. **Взаємна кореляційна функція** двох стохастичних процесів X(t) і Y(t):

$$R_{XY}(t_1, t_2) = \mathbb{E}[X(t_1) \cdot Y(t_2)]$$

   Ця функція використовується для оцінки зв'язку між двома процесами.

Перетворення випадкових процесів

1. **Додавання**. Якщо X(t) і Y(t) – два випадкові процеси, то їх сума Z(t)=X(t)+Y(t) також є випадковим процесом. Характеристики нового процесу можна визначити через характеристики вихідних процесів.
2. **Диференціювання**. Процес

$$Z(t) = \frac{dX(t)}{dt}$$

є похідною випадкового процесу. Його характеристики можна визначити за допомогою властивостей вихідного процесу X(t).

3. **Інтегрування**. Інтеграл від випадкового процесу $Z(t)=\int X(t)\,dt$ також є випадковим процесом. Він застосовується в різних задачах, де потрібно акумулювати випадкові впливи з часом.

Таким чином, випадкові процеси та їх характеристики є ключовими інструментами в аналізі стохастичних систем, дозволяючи моделювати і прогнозувати поведінку систем, що підпадають під вплив випадкових факторів.

Тема 3. Статистичні гіпотези

Статистичні гіпотези — це припущення або твердження про розподіл випадкової величини або параметри цього розподілу, які перевіряються на основі наявних даних. Перевірка статистичних гіпотез є центральним поняттям у статистиці і широко використовується в наукових дослідженнях для прийняття обґрунтованих рішень.

Основні поняття

Нульова гіпотеза (H_0) – це гіпотеза, яку перевіряють за замовчуванням. Вона передбачає відсутність ефекту або зв'язку між досліджуваними явищами.

Альтернативна гіпотеза (H_1 або H_a) – це гіпотеза, яка приймається, якщо нульова гіпотеза відхиляється. Вона передбачає наявність ефекту або зв'язку.

Статистична значимість (p-level) — це показник, що використовується у статистиці для визначення ймовірності того, що спостережуваний результат у дослідженні є

випадковим, а не викликаним якимось фактором. Рівень статистичної значимості позначається як p-значення (рівень значущості).

Основні аспекти поняття статистичної значимості

1. **p-значення**

p-значення — це ймовірність отримання результатів, які є такими ж або більш екстремальними, ніж ті, що спостерігаються, за умови, що нульова гіпотеза (гіпотеза відсутності ефекту) є правдивою.

Наприклад, якщо p=0.05, це означає, що є 5% ймовірності того, що результат дослідження є випадковим.

2. **Нульова гіпотеза (H_0)**

Нульова гіпотеза припускає, що між змінними немає жодного ефекту або зв'язку.

Статистична значимість використовується для тестування нульової гіпотези, щоб визначити, чи є достатньо підстав для її відхилення.

3. **Рівень значимості (α)**

Рівень значимості (α) — це встановлений поріг для прийняття або відхилення нульової гіпотези. Зазвичай використовується значення α=0.05, що відповідає 5% ймовірності. Якщо p-значення менше або дорівнює α, нульова гіпотеза відхиляється, і результат вважається статистично значущим.

4. **Практичне значення**

Важливо розуміти, що статистична значимість не обов'язково означає практичну значимість. Результат може бути статистично значущим, але не мати великого практичного впливу.

Приклад: Припустимо, ви проводите експеримент, щоб перевірити, чи новий метод навчання покращує успішність студентів порівняно з традиційним методом. Нульова гіпотеза полягає в тому, що між двома методами немає різниці (H0: новий метод = традиційний метод). Після проведення експерименту ви отримали p-значення 0.03. Оскільки p=0.03 менше встановленого рівня значимості α=0.05, ви відхиляєте нульову гіпотезу та робите висновок, що новий метод має статистично значущий вплив на успішність студентів. Однак, важливо також оцінити практичну значимість цього впливу, щоб зрозуміти, чи варто впроваджувати новий метод на практиці.

Процедура перевірки гіпотез

1. **Формулювання гіпотез**. Спочатку визначають нульову та альтернативну гіпотези.
2. **Вибір рівня значущості (α)**. Рівень значущості (часто встановлюється на 0,05) є ймовірністю того, що нульова гіпотеза буде відхилена, коли вона насправді істинна.
3. **Вибір тестової статистики**. Залежно від типу даних і дослідження, вибирається відповідний статистичний критерій.
4. **Обчислення тестової статистики та p-значення**. p-значення показує ймовірність отримання результатів, які є такими ж екстремальними або більш екстремальними, ніж ті, що були отримані, за умови, що нульова гіпотеза є істинною.
5. **Прийняття рішення.** Якщо p-значення менше за рівень значущості, нульову гіпотезу відхиляють.

Емпірична функція розподілу. Теорема Глівенко

Емпірична функція розподілу — це функція, яка описує розподіл вибіркових даних. Вона оцінює імовірність того, що випадкова величина не перевищує певного значення.

Визначення: Для вибірки $X_1, X_2, \ldots, X_n$ із загальної сукупності, емпірична функція розподілу $F_n(x)$ визначається як:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x),$$

де *I(x)* — індикаторна функція.

Теорема Глівенко-Кантеллі стверджує, що емпірична функція розподілу сходиться до справжньої функції розподілу з імовірністю 1, коли розмір вибірки n прямує до нескінченності:

$$\sup_x |F_n(x) - F(x)| \to 0$$

майже напевно, при n→∞,

де *F(x)* — справжня функція розподілу.

Критерій Колмогорова-Смирнова використовується для перевірки відповідності емпіричної функції розподілу теоретичній. Він ґрунтується на максимальному відхиленні між емпіричною і теоретичною функціями розподілу.

Критерій Колмогорова-Смирнова (К-С тест) використовується для перевірки відповідності емпіричної функції розподілу даних певному теоретичному розподілу.

Визначення: Тест Колмогорова-Смирнова оцінює максимальне відхилення між емпіричною функцією розподілу $F_n(x)$ і теоретичною функцією розподілу *F(x):*

$$D_n = \sup_x |F_n(x) - F(x)|.$$

Гіпотези:

H0: Вибірка відповідає теоретичному розподілу *F(x)*.

H1: Вибірка не відповідає теоретичному розподілу *F(x)*.

Критичне значення: Якщо D_n більше за певне критичне значення, то нульова гіпотеза відхиляється.

Критерій Пірсона (χ^2-тест) — це статистичний тест, що використовується для перевірки гіпотези про відповідність частот розподілу спостережень певному теоретичному розподілу.

Визначення: Тест Пірсона розраховується за формулою:

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i},$$

де O_i — спостережена частота, E_i— очікувана частота, k — кількість категорій.

Гіпотези:

H0: Немає значущої різниці між спостереженими і очікуваними частотами.

H1: Є значуща різниця між спостереженими і очікуваними частотами.

Критичне значення. Визначається за допомогою таблиць розподілу χ^2, залежно від рівня значущості та кількості ступенів свободи.

Критерій Пірсона (хи-квадрат) використовується для перевірки відповідності емпіричного розподілу теоретичному. Цей критерій порівнює частоти спостережень у категоріях із частотами, очікуваними за теоретичним розподілом.

Квантилі та процентні точки нормального розподілу

Квантилі — це точки, що ділять розподіл на певні частини. Наприклад, медіана — це 0,5-квантиль.

Квантиль порядку *p* — це таке значення x_p, що імовірність того, що випадкова величина приймає значення, менше за x_p, дорівнює *p*: $P(X \leq x_p)=p$.

Процентні точки — це квантилі, що ділять розподіл на 100 частин. Наприклад, 95-та процентна точка нормального розподілу відповідає такому значенню, що 95% значень лежать нижче за нього.

Зв'язок з нормальним розподілом: Для стандартного нормального розподілу (із середнім 0 і дисперсією 1), квантилі знаходяться за допомогою таблиць нормального розподілу або числових методів.

Цей розділ зосереджений на перевірці статистичних гіпотез і оцінці відповідності емпіричних даних певним теоретичним моделям, що є ключовим аспектом в аналізі даних.

Статистичні гіпотези і методи їх перевірки є основними інструментами в аналізі даних, що дозволяють робити обґрунтовані висновки і приймати рішення на основі наявних даних. Вибір відповідного критерію перевірки залежить від характеру даних та гіпотези, що перевіряється.

Тема 4. Дисперсійний аналіз. Кореляційний та регресійний аналіз

Дисперсійний аналіз

Дисперсійний аналіз (ANOVA, Analysis of Variance) — це статистичний метод, що використовується для дослідження впливу одного або декількох факторів на залежну змінну та для порівняння середніх значень у кількох групах. Він дозволяє визначити, чи існують статистично значущі відмінності між середніми значеннями кількох груп, що дозволяє робити висновки про фактори, які можуть впливати на результати.

Основні поняття дисперсійного аналізу:

1. **Фактор**. Незалежна змінна, яка може впливати на залежну змінну (результат). Фактори можуть мати кілька рівнів (значень).

2. **Групи**. Рівні фактора, між якими порівнюються середні значення залежної змінної.
3. **Дисперсія**. Величина, що характеризує розсіювання даних навколо середнього значення.
4. **Загальна дисперсія**: Загальне розсіювання залежної змінної по всіх групах.
5. **Внутрішньогрупова дисперсія**. Дисперсія в межах кожної групи, яка виникає через випадкові варіації.
6. **Міжгрупова дисперсія**. Дисперсія між групами, яка виникає через різницю між середніми значеннями груп.

Мета дисперсійного аналізу. Основна мета дисперсійного аналізу полягає в перевірці нульової гіпотези про те, що середні значення залежної змінної у різних групах рівні. Іншими словами, ми перевіряємо, чи різниця між середніми значеннями груп є настільки великою, що її можна пояснити не випадковими відмінностями, а впливом фактора.

Формули в дисперсійному аналізі

1. **Загальна дисперсія**:

$$SST = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

де SST (Sum of Squares Total) — загальна сума квадратів відхилень спостережень від загального середнього $\bar{Y}$.

2. **Міжгрупова дисперсія**:

$$SSB = \sum_{j=1}^{k} n_j (\bar{Y}_j - \bar{Y})^2$$

де SSB (Sum of Squares Between) — сума квадратів відхилень середніх групових значень від загального середнього, n_j — кількість спостережень у групі j, $\bar{Y}_j$ — середнє значення в групі j.

3. **Внутрішньогрупова дисперсія**:

$$SSW = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (Y_{ij} - \bar{Y}_j)^2$$

де SSW (Sum of Squares Within) — сума квадратів відхилень спостережень від середнього значення у своїй групі.

4. **F-статистика**:

$$F = \frac{MSB}{MSW}$$

$MSB = \frac{SSB}{k-1}$ — середньоквадратична відхилення між групами,

$MSW = \frac{SSW}{n-k}$ — середньоквадратична відхилення всередині груп.

Типи дисперсійного аналізу:

1. **Однофакторний дисперсійний аналіз (One-Way ANOVA)**

Використовується для порівняння середніх значень між кількома групами, які визначені одним фактором.

2. **Багатофакторний дисперсійний аналіз (Multi-Way ANOVA)**

Включає два або більше факторів і дозволяє оцінювати вплив кожного фактора, а також їх взаємодію.

3. **Повторні вимірювання (Repeated Measures ANOVA)**

Використовується, коли одні й ті ж суб'єкти вимірюються кілька разів за різних умов або в різний час.

Приклад використання дисперсійного аналізу

Припустимо, що ви хочете перевірити, чи є різниця в середніх оцінках студентів, які використовували три різні методики навчання. У цьому випадку:

Фактор — методика навчання.

Рівні — три різні методики.

Залежна змінна — оцінки студентів.

За допомогою однофакторного дисперсійного аналізу можна визначити, чи існують статистично значущі відмінності в середніх оцінках між групами студентів, які використовували різні методики навчання.

Таким чином, дисперсійний аналіз є потужним інструментом для порівняння середніх значень між кількома групами та виявлення впливу одного або кількох факторів на результати. Його широко використовують у різних галузях науки, включаючи медицину, психологію, економіку та інші.

Кореляційний та регресійний аналіз

Кореляційний і регресійний аналіз є важливими інструментами статистичного аналізу, які використовуються для виявлення зв'язків між змінними та побудови моделей прогнозування. Розглянемо основні аспекти, моделі та методи цих аналізів.

Кореляційний аналіз вивчає силу та напрямок зв'язку між двома або більше змінними. Основною метою є визначення, чи є між змінними зв'язок, і наскільки сильним він є.

Завдання кореляційного аналізу:

1. Визначення наявності зв'язку між змінними.
2. Оцінка сили зв'язку.
3. Визначення напрямку зв'язку (позитивний або негативний).
4. Вивчення залежності між ранговими даними (рангова кореляція).

Основні типи кореляції:

- **Парна лінійна кореляція**. Вимірює лінійний зв'язок між двома змінними за допомогою коефіцієнта кореляції Пірсона.
- **Рангова кореляція**. Використовується для аналізу зв'язків між ранговими даними. Найбільш відомий коефіцієнт — коефіцієнт кореляції Спірмена.
- **Нелінійна парна кореляція**. Вивчає зв'язок між змінними, що не підкоряються лінійним залежностям.
- **Множинна кореляція.** Аналізує зв'язок між однією залежною змінною та декількома незалежними змінними.

Регресійний аналіз

Регресійний аналіз використовується для моделювання відносин між залежною змінною (результатом) та однією або кількома незалежними змінними (факторами). Метою є побудова математичної моделі, яка описує залежність між змінними та дозволяє робити прогнози.

Завдання регресійного аналізу:

1. Побудова математичної моделі, яка описує зв'язок між змінними.
2. Оцінка впливу незалежних змінних на залежну змінну.
3. Прогнозування значень залежної змінної на основі значень незалежних змінних.
4. Визначення значущих факторів, що впливають на залежну змінну.

Основні типи регресійних моделей:

Парна лінійна регресійна модель. Описує лінійний зв'язок між однією незалежною змінною та однією залежною змінною.

Множинна регресія. Описує зв'язок між кількома незалежними змінними та залежною змінною.

Нелінійна регресія. Використовується, коли зв'язок між змінними не є лінійним.

Змішана модель. Поєднує як випадкові, так і детерміновані елементи у моделюванні залежної змінної.

Моделі

1. **Випадкова модель**. Включає випадкові змінні, які мають певний розподіл ймовірностей.
2. **Детермінована модель**. Включає лише фактори, які чітко визначені та не змінюються.
3. **Змішана модель**. Поєднує елементи як випадкової, так і детермінованої моделі.

Побудова регресійних моделей

Для побудови регресійної моделі виконуються наступні кроки:

1. **Збір даних**. Необхідно зібрати дані для залежної та незалежних змінних.
2. **Попередній аналіз**. Перевірка даних на предмет наявності зв'язку, виявлення відхилень та пропущених даних.
3. **Вибір моделі**. Визначення типу регресії, яка найкраще відповідає даним.
4. **Оцінка параметрів моделі**. Використання методу найменших квадратів або інших методів для оцінки параметрів.
5. **Перевірка моделі**. Оцінка адекватності моделі за допомогою тестів на значущість, перевірка залишків.
6. **Прогнозування**. Використання моделі для прогнозування нових даних.

7. **Інтерпретація результатів**. Аналіз отриманих результатів для прийняття рішень.

Тема 5. Статистичний аналіз даних

Статистичний аналіз даних є основним інструментом для вивчення та інтерпретації великої кількості даних у різних галузях, включаючи економіку, соціальні науки, медицини та інші. Він охоплює широкий спектр методів, які дозволяють зібрати, описати, проаналізувати та інтерпретувати дані, щоб зробити висновки та прогнозування.

Основні етапи статистичного аналізу:

1. **Збір даних**. Включає вибір відповідних методів збору даних (опитування, експерименти, спостереження). Важливо забезпечити репрезентативність вибірки та якість зібраних даних.
2. **Попередня обробка даних**. Виявлення та усунення пропущених даних, аномалій, відхилень. Перетворення даних у відповідний формат для подальшого аналізу.
3. **Описова статистика**. Використовується для узагальнення та представлення даних за допомогою показників, таких як середнє значення, медіана, мода, дисперсія, стандартне відхилення. Описова статистика допомагає зрозуміти основні характеристики даних, такі як розподіл, центральна тенденція та варіабельність.
4. **Візуалізація даних**. Графічні методи, такі як гістограми, діаграми розсіювання, коробкові діаграми, використовуються для наочного представлення даних та виявлення можливих закономірностей. Візуалізація допомагає краще зрозуміти структуру та розподіл даних.
5. **Перевірка статистичних гіпотез**. Гіпотези формулюються для перевірки певних припущень про характеристики даних. Використовуються статистичні

тести, такі як t-тест, χ^2-тест, ANOVA, для перевірки гіпотез та визначення статистичної значущості результатів.

6. **Регресійний аналіз**. Дозволяє встановити зв'язок між залежною змінною та однією або кількома незалежними змінними. Регресійний аналіз використовується для побудови прогнозних моделей та оцінки впливу незалежних змінних на залежну.
7. **Кореляційний аналіз**. Визначає силу та напрямок зв'язку між змінними. Коефіцієнт кореляції використовується для вимірювання лінійної залежності між змінними.
8. **Аналіз дисперсії (ANOVA)**. Дозволяє порівняти середні значення між кількома групами та визначити, чи є між ними статистично значущі відмінності.
9. **Моделювання та прогнозування**. Побудова статистичних моделей на основі наявних даних для прогнозування майбутніх значень. Використовуються різні підходи, такі як лінійна регресія, часові ряди, машинне навчання.

Приклад застосування. Уявімо, що компанія проводить дослідження для виявлення факторів, що впливають на продажі продукту. Використовуючи регресійний аналіз, компанія може встановити, які змінні (ціна, реклама, сезонність) найбільше впливають на обсяги продажів. Далі можна використовувати цю модель для прогнозування майбутніх продажів за різних умов.

Класифікація методів статистичного аналізу охоплює різні підходи для обробки, узагальнення та інтерпретації даних. Методи можна класифікувати за різними критеріями, зокрема за типом даних, що аналізуються, та цілями аналізу.

Описова статистика

Описова статистика включає методи, які використовуються для узагальнення та представлення основних характеристик даних.

Частотний розподіл показує, як часто певні значення зустрічаються в наборі даних. Він зазвичай подається у вигляді таблиць або графіків (гістограми), де кожній групі значень (інтервалу) відповідає частота їх появи.

Відсоткові показники використовуються для представлення частот у вигляді відсотків від загальної кількості спостережень. Це допомагає краще зрозуміти відносну частоту значень у наборі даних.

Заходи центральної тенденції

Медіана – середнє значення, яке розташоване посередині впорядкованого ряду даних. Якщо кількість спостережень парна, медіана визначається як середнє двох середніх значень.

Мода – значення, яке зустрічається найчастіше в наборі даних. Мода може бути не єдиною, якщо є кілька значень, які мають однакову частоту.

Міри розкиду даних

Дисперсія – показник розсіювання даних навколо середнього значення. Вона розраховується як середнє арифметичне квадратів відхилень кожного значення від середнього.

Стандартне відхилення – квадратний корінь із дисперсії. Воно показує, наскільки значення в наборі даних відхиляються від середнього значення в середньому.

Квартилі та інтерквартильний розмах – квартилі розбивають набір даних на чотири рівні частини, а інтерквартильний розмах визначається як різниця між першим і третім квартилями, що відображає середнє розсіювання центральних 50% даних.

Методи вторинної статистичної обробки результатів

Методи вторинної обробки включають аналіз даних, який проводиться після первинного зведення та опису даних. Сюди входять:

Кореляційний аналіз: визначення взаємозв’язку між двома або більше змінними.

Регресійний аналіз: побудова моделі, яка описує зв'язок між залежною змінною та однією або більше незалежними змінними.

Параметричні та непараметричні статистичні методи

Параметричні методи передбачають, що дані відповідають певному розподілу (наприклад, нормальному розподілу). До параметричних методів відносяться:

- **t-тест** використовується для порівняння середніх значень двох груп.
- **ANOVA (дисперсійний аналіз)** використовується для порівняння середніх значень більше ніж двох груп.
- **Коефіцієнт кореляції Пірсона** вимірює силу лінійного зв'язку між двома змінними.

Непараметричні методи не передбачають певного розподілу даних і використовуються, коли параметричні припущення не можуть бути виконані. Вони включають:

- **Медіанний тест** використовується для перевірки відмінностей між медіанами кількох груп.
- **Критерій знаків** використовується для перевірки парних відмінностей між медіанами.
- **Коефіцієнт кореляції Спірмена** вимірює силу і напрямок монотонного зв'язку між двома змінними.

Класифікація методів статистичного аналізу включає описові методи, які дозволяють узагальнити дані, та методи вторинної обробки, які дозволяють досліджувати взаємозв'язки між змінними та робити висновки на основі даних. Вибір між параметричними та непараметричними методами залежить від характеру даних і припущень про їх розподіл.

Програмні засоби для статистичних досліджень відіграють важливу роль у проведенні аналізу даних, моделюванні та візуалізації результатів. Вони використовуються для виконання різних статистичних процедур, таких як описова статистика, регресійний аналіз, тестування гіпотез, аналіз часових рядів, і багато інших.

Основні програмні засоби, які широко використовуються у статистичних дослідженнях:

1. R є одним з найпотужніших і найпоширеніших інструментів для статистичного аналізу. Це мова програмування та програмне середовище для обчислень і графіки, яка має великий набір бібліотек для статистичних методів.

Основні можливості:

- Розширені можливості статистичного аналізу та моделювання.
- Потужна система для візуалізації даних.
- Широкий спектр пакетів для різних статистичних методів, включаючи часові ряди, регресію, машинне навчання, і т.д.

Застосування: Академічні дослідження, аналіз фінансових даних, біостатистика, соціологія.

2. SPSS (Statistical Package for the Social Sciences) є популярним програмним засобом для проведення статистичних аналізів у соціальних та поведінкових науках.

Основні можливості:

- Простий у використанні інтерфейс з підтримкою роботи через графічний інтерфейс користувача (GUI).
- Можливості для аналізу великих масивів даних, включаючи регресійний аналіз, кластерний аналіз, факторний аналіз і т.д.
- Інструменти для обробки даних, візуалізації результатів та автоматизації звітності.

Застосування: Соціологія, маркетингові дослідження, психологія, медицина.

3. SAS (Statistical Analysis System) є комплексним інструментом для аналізу даних і статистики, що використовується в різних галузях, від бізнесу до академії.

Основні можливості:

- Широкий спектр можливостей для обробки даних, включаючи маніпуляції з даними, аналіз часових рядів, багатовимірний аналіз.
- Потужні інструменти для прогнозування та моделювання.
- Інтеграція з іншими інструментами та мовами програмування.

Застосування: Біостатистика, фінанси, фармацевтика, аналітика бізнесу.

4. Stata є популярним інструментом для економетричного аналізу та статистики. Він використовується для аналізу даних у багатьох дисциплінах, включаючи економіку, соціологію та біостатистику.

Основні можливості:

- Потужні інструменти для роботи з панельними даними та часовими рядами.
- Інструменти для регресійного аналізу, тестування гіпотез, аналізу виживання.
- Підтримка командного рядка та GUI для користувачів.

Застосування: Економетрика, соціальні науки, медицина.

5. Python Python є універсальною мовою програмування, яка також широко використовується для статистичного аналізу завдяки численним бібліотекам для роботи з даними.

Основні можливості:

- Бібліотеки NumPy, Pandas для обробки даних.

- SciPy, StatsModels для статистичних методів і аналізу.
- Matplotlib, Seaborn для візуалізації даних.
- Інтеграція з іншими інструментами машинного навчання.

Застосування: Дослідження даних, наука про дані, машинне навчання, аналітика бізнесу.

6. Excel є простим і доступним інструментом для базового статистичного аналізу, який особливо підходить для початківців і невеликих досліджень.

Основні можливості:

- Простота у використанні для базових статистичних розрахунків, включаючи середні, медіани, дисперсії та регресії.
- Вбудовані функції для статистичного аналізу та обробки даних.
- Інструменти для візуалізації, такі як діаграми та графіки.

Застосування: Бізнес-аналітика, навчання, маркетингові дослідження.

Вибір програмного засобу для статистичного аналізу залежить від специфіки завдання, складності даних та потреб користувача. Для складних статистичних аналізів, таких як моделювання та прогнозування, найкраще підходять R, Python, SAS та SPSS, тоді як для простіших завдань можна використовувати Excel.

Статистичний аналіз даних є потужним інструментом для обробки, інтерпретації та прогнозування на основі наявних даних. Він дозволяє робити обґрунтовані висновки та приймати рішення на основі доказів, що є критично важливим у багатьох галузях.

t-критерій Стьюдента – це статистичний тест, що використовується для порівняння середніх значень двох вибірок. Він допомагає визначити, чи є різниця між середніми статистично значимою, що важливо для перевірки гіпотез в умовах невеликого обсягу вибірки.

Дисперсійний аналіз (ANOVA) – це метод статистичного аналізу, що використовується для визначення, чи існують статистично значимі відмінності між середніми значеннями більше ніж двох груп. Він допомагає виявити фактори, які впливають на змінну.

U-критерій Манна-Уїтні – це непараметричний тест для визначення, чи відрізняються дві незалежні вибірки одна від одної. Цей тест є альтернативою t-критерію Стьюдента, коли дані не відповідають припущенню про нормальний розподіл.

Критерій серій Вальда-Вольфовиця – використовується для перевірки випадковості послідовності даних. Він допомагає визначити, чи є дані випадковими або ж вони мають певну структуру, яка порушує випадковість.

Двовибірковий критерій Колмогорова-Смірнова – цей тест використовується для порівняння двох вибірок і визначення, чи мають вони однаковий розподіл. Він не робить припущень про форму розподілу і тому є досить універсальним.

Ранговий дисперсійний аналіз Фрідмана – непараметричний тест, який використовується для аналізу варіацій у випадках, коли дані є ранговими. Він є альтернативою дисперсійному аналізу для залежних вибірок.

Q-критерій Кохрена – використовується для перевірки гомогенності варіацій у багатьох групах. Він допомагає виявити, чи є дисперсія в групах однаковою, що є важливим передумовою для багатьох статистичних тестів.

Коефіцієнт кореляції Браве-Пірсона – це міра лінійного зв'язку між двома змінними. Він варіюється від -1 до 1, де 1 означає повний позитивний зв'язок, -1 – повний негативний, а 0 – відсутність лінійного зв'язку.

Коефіцієнт кореляції рангів Спірмена – це непараметрична альтернатива коефіцієнту Пірсона, що використовується для вимірювання зв'язку між ранговими

змінними. Він також варіюється від -1 до 1, але не вимагає нормального розподілу даних.

Роль і значення графічного методу в статистиці – графічні методи є важливими інструментами в статистиці, що дозволяють візуалізувати дані, виявляти аномалії, закономірності, а також полегшують інтерпретацію результатів аналізу. Вони включають гістограми, діаграми розсіювання, ящики з вусами та інші візуальні представлення даних, які допомагають у прийнятті рішень на основі статистичних досліджень.

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Основна

1. Стрелковська І.В. Вища математика для фахівців в галузі зв’язку. Ч.5 / І.В. Стрелковська, В.М. Паскаленко. – Одеса: ВМВ, 2018. – 508 с.

2. Стрелковська І.В. Теорія ймовірностей та випадкові процеси (для фахівців у галузі ІТ-технологій) / І.В. Стрелковська, В.М. Паскаленко. – Одеса: ОНАЗ, 2018. – 384 с.

3. Стрелковська І.В. Математична статистика / І.В. Стрелковська, В.М. Паскаленко. – Одеса: ОНАЗ, 2019. – 110 с.

4. Стрелковська І.В., Григор’єва Т.І. Математичні методи в наукових дослідженнях: методичні рекомендації для самостійної роботи здобувачів / 2023, МГУ, Одеса, 22с.

Допоміжна

5. Стрелковська І.В., Золотухін Р.В., Григор’єва Т.І. Узагальнена модель оцінки показників функціонування низькошвидкісних мереж зв’язку автоматизованих систем управління. Інфокомунікаційні та комп’ютерні технології. – 2022. – № 1 (03). – С. 138-153.

6. Мішура Ю. С. Випадкові процеси: теорія, статистика, застосування : підручник / Ю. С. Мішура, К. В. Ральченко, Г. М. Шевченко. – 2-ге вид., випр. і допов. – К. : ВПЦ "Київський університет", 2021. – 496 с.

7. Випадкові процеси: теорія, статистика, застосування : підручник / Ю. С. Мішура, К. В. Ральченко, Л. М. Сахно, Г. М. Шевченко. – 2-ге вид., випр. і допов. – К. : ВПЦ "Київський університет", 2023. – 496 с.

8. Крилик, Л. В. Обчислювальна математика. Інтерполяція та апроксимація табличних даних : навчальний посібник / Л. В. Крилик, І.В. Богач, М. О. Прокопова. – Вінниця : ВНТУ, 2013. – 111 с.

9. Л.П.Вакал Апроксимація функцій багатьох змінних із застосуванням алгоритму диференціальної еволюції / Л.П. Вакал. // Математичні машини і системи. – 2017. – с. 90–96с.

10. Методика та організація наукових досліджень : Навч. посіб. / С. Е. Важинський, Т І. Щербак. – Суми: СумДПУ імені А. С. Макаренка, 2016. – 260 с.

11. Стрелковська І.В. Операційне числення для фахівців у галузі зв'язку (для студентів та аспірантів) / І.В. Стрелковська, В.М. Паскаленко. – Одеса: ОНАЗ, 2017. – 120 с.

12. Strelkovskaya I., Solovskaya I., Makoganiuk A. Different extrapolation methods in Problems of Forecasting. Advances in Information and Communication Technology and Systems. Lecture Notes in

Networks and Systems. 2020. Vol. 152. Springer. P. 217-228. https://doi.org/10.1007/ 978-3-030-58359-0_12

13. Strelkovskaya I., Solovskaya I., Strelkovska J. Fingerprinting/Indoor positioning using complex planar splines. Journal of Electrical Engineering. Vol. 72 (2021), N06, pp. 401-406. https://doi.org/10.2478/jee-2021-0057

14. Павленко П.М. Основи математичного моделювання систем і процесів: Навчальний посібник. – К.: Книжкове вид-во НАУ, 2010.– 198 с.

15. Маценко В.Г. Математичне моделювання: навчальний посібник– Чернівці: Чернівецький національний університет, 2014.–519 с.

16. Пасічник В.В., Виклюк Я.І., Камінський Р.М. Моделювання складних систем. Посібник. Львів: Видавництво "Новий Світ - 2000". 2017. 404 с.

Інформаційні ресурси

17. Національна бібліотека України імені В. І. Вернадського [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – Київ: НБУВ, 2013-2015. – Режим доступу: www.nbuv.gov.ua – Назва з екрана. Дата звернення: 07.08.2023.

18. Електронний каталог Національної парламентської бібліотеки України [Електронний ресурс]: [політемат. база даних містить відом. про вітчизн. та зарубіж. кн., брош., що надходять у фонд НПБ України]. – Електронні дані (803 438 записів). – Київ: Нац. парлам. б-ка України, 2002-2015. – Режим доступу: catalogue.nplu.org . – Назва з екрана. Дата звернення: 07.08.2023.

19. Український інститут інтелектуальної власності [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – Київ: УІІВ, 2017. – Режим доступу: http://www.uipv.org – Назва з екрана. Дата звернення: 07.08.2023.

ЗМІСТ

1. Опис навчальної дисципліни……..……...3
2. Заплановані результати навчання за навчальною дисципліною ……4
3. Програма навчальної дисципліни……..5
4. Структура навчальної дисципліни……..6
5. Питання до практичних занять……..…….…7
6. Самостійна робота……..8
7. Теми доповідей ……..9
8. Види та методи контролю……..……..….…11
9. Питання до заліку…….…11
10. Критерії підсумкової оцінки знань студентів…….….13
11. План – конспект лекцій дисципліни «**Сучасні проблеми теорії ймовірностей та математичної статистики**»……15
12. Рекомендована література……44

Навчальне видання

Григор'єва Тетяна Ігорівна

Соловська Ірина Миколаївна

СУЧАСНІ ПРОБЛЕМИ ТЕОРІЇ ІМОВІРНОСТЕЙ ТА МАТЕМАТИЧНОЇ СТАТИСТИКИ

Методичні рекомендації для самостійної роботи здобувачів

Українською мовою