

МІЖРЕГІОНАЛЬНА
АКАДЕМІЯ УПРАВЛІННЯ ПЕРСОНАЛОМ



МАУП

НАВЧАЛЬНА ПРОГРАМА
дисципліни
“АНАЛІЗ ДАНИХ”
(для бакалаврів)

МАУП

Київ
ДП «Видавничий дім «Персонал»
2010

Підготовлено доцентом кафедри прикладної математики
та програмування *Н. Г. Кирилахою*

Затверджено на засіданні кафедри прикладної математики
та програмування (протокол № 5 від 18.01.06)

Перезатверджено на засіданні кафедри прикладної математики
та інформаційних технологій (протокол № 6 від 12.03.09)

Схвалено Вченою радою Міжрегіональної Академії управління персоналом

Кирилах Н. Г. Навчальна програма дисципліни “Аналіз даних” (для ба-
калаврів). — К.: ДП «Вид. дім «Персонал», 2010. — 15 с.

Навчальна програма містить пояснювальну записку, тематичний план,
зміст дисципліни “Аналіз даних”, питання для самоконтролю, а також список
літератури.

© Міжрегіональна Академія
управління персоналом (МАУП), 2010
© ДП «Видавничий дім «Персонал», 2010

ПОЯСНЮВАЛЬНА ЗАПИСКА

На сучасному етапі розвитку суспільних відносин прийняття рішення щодо функціонування будь-якої соціальної або виробничої системи пов'язано з обробкою значних масивів даних. В умовах жорсткої конкуренції рішення повинні прийматися на основі ретельного аналізу цих даних, бути обґрунтованими. Розповсюдженню методів аналізу даних сприяла ще можливість застосовувати до такого аналізу можливості обчислювальної техніки. Тепер коло замовників такого аналізу значно розширилося. Це вже не тільки великі підприємства і державні установи, а й представники середнього і малого бізнесу. Тому кожному майбутньому фахівцю необхідно володіти навичками аналізу даних, зокрема в числовому вигляді.

Предметом дисципліни “Аналіз даних” є задачі аналізу даних: побудови і перевірки гіпотез щодо властивостей розподілів, побудови математичних моделей у вигляді кривих регресії, різні види статистичного аналізу: дисперсійного, факторного, кластерного та інших; задачі на використання методів регуляризації, фільтрації і апроксимації даних.

Дисципліна “Аналіз даних” є вибірковою навчальною дисципліною. Її матеріал базується на знаннях курсів “Математичний аналіз”, “Аналітична геометрія та лінійна алгебра”, “Теорія ймовірностей та математична статистика” і забезпечує подальше виконання курсових, дипломних і магістерських робіт.

Форми занять з дисципліни: лекції, лабораторні роботи.

Робоча програма курсу містить тематичний план дисципліни із погодинним розподілом на лекційні, лабораторні заняття, самостійну роботу студентів, зміст лекцій курсу і лабораторних занять, розподіл матеріалу за модулями, список основної та додаткової літератури.

У результаті вивчення дисципліни студенти повинні *знати:*

- основні поняття теорії аналізу даних;
- методи опрацювання емпіричних даних та отримання можливих статистичних оцінок невідомих параметрів генеральної сукупності;
- методи побудови і перевірки статистичних гіпотез щодо характеру розподілу ознаки в генеральній сукупності;
- основні принципи використання методів дисперсійного аналізу;

- основні положення одно- і двофакторного аналізу;
- основні методи кореляційного та регресійного аналізу для виявлення взаємозв'язків між ознаками;
- основні принципи використання кластерного аналізу;
- основні методи вивчення часових рядів;
- основні принципи теорії обробки сигналів;
- основні методи регуляризації, фільтрації і апроксимації;

уміти:

- із опису предметної області і задач, що розв'язуються, приймати рішення щодо застосування того або іншого методу обробки даних;
- застосовувати різноманітні методи обробки даних;
- використовувати можливості обчислювальної техніки при обробці та аналізі даних експерименту;
- використовувати результати для формулювання висновків.

ТЕМАТИЧНИЙ ПЛАН
дисципліни
“АНАЛІЗ ДАНИХ”

№ пор.	Назва змістового модуля і теми
1	2
	Змістовий модуль I. Оцінювання статистичних параметрів та основи перевірки статистичних гіпотез
1	Статистичне оцінювання
2	Основи перевірки статистичних гіпотез
3	Побудова гіпотез про характер розподілу
4	Критерії згоди
	Змістовий модуль II. Основні види статистичного аналізу
5	Однофакторний аналіз
6	Двофакторний аналіз
7	Регресійний аналіз
8	Часові ряди
9	Практичний аналіз часових рядів

1	2
	Змістовий модуль III. Інші методи аналізу даних. Застосування методів статистичного аналізу до прикладних задач
10	Багатовимірний аналіз та інші статистичні методи
11	Проблеми ідентифікації швидких процесів
12	Ознайомлення з теорією обробки сигналів
13	Ознайомлення з деякими оберненими задачами теорії управління, біофізики і механіки
14	Методи регуляризації, фільтрації та апроксимації
Разом годин: 81	

ЗМІСТ
дисципліни
“АНАЛІЗ ДАНИХ”

**Змістовий модуль I. Оцінювання статистичних параметрів
та основи перевірки статистичних гіпотез**

Тема 1. Статистичне оцінювання

1. Висновки із закону великих чисел.
2. Статистичні параметри.
3. Параметри розподілу.
4. Параметри моделі.
5. Оцінювання параметрів розподілу за даними вибірки.
6. Властивості оцінок.
7. Довірче оцінювання.
8. Точність вибіркової оцінки.
9. Метод найбільшої правдоподібності.
10. Застосування статистичних пакетів при знаходженні оцінок розподілу.

Література [1–4; 6]

Тема 2. Основи перевірки статистичних гіпотез

1. Статистичні моделі.
2. Перевірка статистичних гіпотез (загальні положення).
3. Приклади статистичних моделей і гіпотез.

4. Перевірка статистичних гіпотез про параметри розподілу генеральної сукупності (прикладні задачі).
5. Схема випробувань Бернуллі.
6. Критерій знаків для однієї вибірки.
7. Перевірка гіпотез у двовибіркових задачах.
8. Критерій Манна – Уїтні. Критерій Уїлкоксона.
9. Парні спостереження. Критерій знаків для аналізу парних повторних спостережень.
10. Аналіз повторних парних спостережень за допомогою знакових рангів (критерій знакових рангових сум Уїлкоксона).

Література [1–4; 7]

Тема 3. Побудова гіпотез про характер розподілу

1. Дослідження нормальних вибірок.
2. Оцінки параметрів нормального розподілу та їх властивості.
3. Побудова гіпотези про нормальний характер розподілу генеральної сукупності за однією вибіркою, за двома вибірками.
4. Парні дані.
5. Побудова гіпотези про рівномірний характер розподілу генеральної сукупності.
6. Гіпотези про показниковий характер розподілу.
7. Використання статистичних пакетів для аналізу характеру розподілу.

Література [1–4; 6]

Тема 4. Критерії згоди

1. Критерії згоди Колмогорова і омега-квадрат у випадку простої гіпотези. Критерій згоди χ^2 -квадрат К. Пірсона для простої гіпотези.
2. Критерії згоди для складних гіпотез.
3. Критерій згоди χ^2 -квадрат Фішера для складної гіпотези.
4. Інші критерії згоди.
5. Критерій згоди для пуасонівського розподілу.

Література [1–4; 6]

Змістовий модуль II. Основні види статистичного аналізу

Тема 5. Однофакторний аналіз

1. Постановка задачі однофакторного аналізу.

2. Непараметричні критерії перевірки однорідності: критерій Краскела – Уолліса (довільні альтернативи).
3. Критерій Джонкхієра (альтернативи з упорядкуванням).
4. Оцінювання ефектів обробки (непараметричний підхід).
5. Дисперсійний аналіз.
6. Оцінювання ефектів обробки в нормальній моделі: довірчі інтервали.
7. Метод Шеффе множинних порівнянь.
8. Використання статистичних пакетів при проведенні однофакторного аналізу.

Література [1–4; 6; 7; 16]

Тема 6. Двофакторний аналіз

1. Зв'язок задач дво- і однофакторного аналізу.
2. Таблиця двофакторного аналізу.
3. Адитивна модель даних двофакторного експерименту при незалежній дії факторів.
4. Непараметричні критерії перевірки гіпотези про відсутність ефектів обробки: критерій Фрідмана (довільні альтернативи).
5. Критерій Пейджа (альтернативи з упорядкуванням).
6. Двофакторний дисперсійний аналіз.

Література [1–4; 6; 7; 16]

Тема 7. Регресійний аналіз

1. Модель лінійного регресійного аналізу.
2. Стратегія, методи і проблеми регресійного аналізу.
3. Проста лінійна регресія.
4. Перевірка передумов у задачі регресійного аналізу.
5. Метод найменших квадратів.
6. Непараметрична лінійна регресія.
7. Нелінійна регресія.
8. Обґрунтування і застосування методу найменших квадратів при побудові нелінійної регресії.
9. Шкали вимірювання.
10. Інструменти і стратегія дослідження зв'язків ознак. Зв'язки номінальних ознак (таблиці спряженості).
11. Зв'язок ознак, що вимірюються у шкалі порядків.
12. Зв'язок ознак у кількісних шкалах.

13. Коефіцієнт кореляції, нормальна кореляція.

14. Зв'язок ознак, вимірних у різних шкалах.

Література [1–4; 6; 7; 16]

Тема 8. Часові ряди

1. Аналіз часових рядів, його розділи.
2. Мета, етапи та методи аналізу часових рядів.
3. Детермінована і випадкова складові часового ряду.
4. Тренд. Сезонна і циклічна компоненти. Моделі тренду.
5. Моделі випадкової компоненти. Числові характеристики часових рядів. Процеси, стаціонарні в широкому смислі.
6. Оцінки числових характеристик часових рядів.

Література [1–4; 6; 7; 16]

Тема 9. Практичний аналіз часових рядів

1. Порядок аналізу часових рядів.
2. Графічний метод аналізу.
3. Методи зведення до стаціонарності: виділення тренду; виділення сезонних ефектів; метод сквозних середніх; сезонні різниці оператори; перетворення шкали.
4. Аналіз часових рядів за допомогою прикладних пакетів.
5. Методи дослідження стаціонарного часового ряду.
6. Побудова та інтерпретація графіка корелограми, графіка частинної автокореляційної функції.

Література [1–4; 6; 7; 16]

Змістовий модуль III. Інші методи аналізу даних.

Застосування методів статистичного аналізу до прикладних задач

Тема 10. Багатовимірний аналіз та інші статистичні методи

1. Факторний аналіз.
2. Дискримінантний аналіз.
3. Кластерний аналіз.
4. Практичні задачі, що потребують використання методів кластерного аналізу.
5. Методи контролю якості.

Література [1–4; 6; 7; 14; 16]

Тема 11. Проблеми ідентифікації швидких процесів

1. Характеристика процесів, які можна віднести до класу швидких і постановка задачі ідентифікації швидких процесів.
2. Моделювання швидких процесів у вигляді експоненціальних моделей.
3. Дослідження і розробка методів оцінювання параметрів моделей, представлених у вигляді функцій швидкого зростання, спадання.
4. Оцінка числа нулів і областей опуклості отриманих залежностей.
5. Багатоекстремальність і обґрунтованість оцінок методом найменших квадратів експоненціальних моделей.

Література [1–7; 15; 16]

Тема 12. Ознайомлення з теорією обробки сигналів

1. Постановка задачі.
2. Типи обробки сигналів.
3. Методи вторинної обробки сигналів.
4. Методи компенсації локальних сигналів-перешкод.
5. Методи адаптації. Спектральний аналіз.
6. Типи спектрів.
7. Типи спектрального аналізу.
8. Алгоритм інтегральної апроксимації.

Література [1–3; 9–11; 18]

Тема 13. Ознайомлення з деякими оберненими задачами теорії управління, біофізики і механіки

1. Загальна постановка оберненої задачі.
2. Відновлення сигналу в динамічній системі.
3. Відновлення зображень біологічних об'єктів. Томографія. Обернена задача мовної акустики.
4. Спектральні задачі механіки. Застосування методу Фур'є.

Література [1–2; 4; 8–11; 18]

Тема 14. Методи регуляризації, фільтрації та апроксимації

1. Метод регуляризації Тихонова.
2. Рівняння типу згортки.
3. Метод оптимальної фільтрації Кальмана — Бюссі.

4. Метод оптимальної лінійної фільтрації Вінера. Інтерполяція, екстраполяція, згладжування та апроксимація.
5. Лінійна інтерполяція та екстраполяція, квадратична інтерполяція та екстраполяція.
6. Поліном Лагранжа, сплайни, кубічні інтерполюючі сплайни, лінійна та нелінійна апроксимація (згладжування), кубічні сплайни.

Література [1; 2; 9–11; 16–18]

ПИТАННЯ ДЛЯ САМОКОНТРОЛЮ

1. Поняття вибірки. Репрезентативність вибірки.
2. Основні теоретичні розподіли.
3. Що таке статистичні параметри? Які ви знаєте параметри розподілу?
4. Що таке параметри моделі? Які ви знаєте методи оцінювання параметрів розподілу за даними вибірки?
5. Властивості оцінок. Що таке довірче оцінювання? Поняття довірчого інтервалу.
6. Основні типи довірчих оцінок. Точність і надійність довірчої оцінки.
7. Метод найбільшої правдоподібності.
8. Які ви знаєте функції пакету EXCEL для знаходження оцінок розподілу?
9. Поняття статистичної моделі.
10. Поняття статистичної гіпотези. Нульова та альтернативна гіпотези.
11. Приклади статистичних моделей і гіпотез.
12. Критерії перевірки статистичних гіпотез про параметри розподілу генеральної сукупності (прикладні задачі).
13. Перевірка гіпотез у двовибіркових задачах.
14. Критерій Манна – Уїтні. Критерій Уїлкоксона.
15. Парні спостереження. Критерій знаків для аналізу парних повторних спостережень.
16. Аналіз повторних парних спостережень за допомогою знакових рангів (критерій знакових рангових сум Уїлкоксона).
17. Про дослідження нормальних вибірок. Як оцінюються параметри нормального розподілу? Властивості оцінок.

18. Алгоритм побудови гіпотези про нормальний характер розподілу генеральної сукупності за однією вибіркою, за двома вибірками.
19. Побудова гіпотези про рівномірний характер розподілу генеральної сукупності.
20. Гіпотези про показниковий характер розподілу.
21. Критерії згоди (Колмогорова і χ^2 -квадрат К. Пірсона, омега-квадрат) у випадку простої гіпотези.
22. Критерії згоди для складних гіпотез. Критерій згоди χ^2 -квадрат Фішера для складної гіпотези.
23. Які ви знаєте інші критерії згоди?
24. Критерій згоди для пуассонівського розподілу.
25. Постановка задачі однофакторного аналізу.
26. Непараметричні критерії перевірки однорідності.
27. Критерій Краскела — Уолліса (довільні альтернативи).
28. Критерій Джонкхієра (альтернативи з упорядкуванням).
29. Оцінювання ефектів обробки (непараметричний підхід).
30. Дисперсійний аналіз. Кореляційне відношення, критерій Фішера.
31. Оцінювання ефектів обробки в нормальній моделі: довірчі інтервали; метод Шеффе множинних порівнянь.
32. Як пов'язані задачі дво- і однофакторного аналізу? Як скласти таблицю двофакторного аналізу?
33. Алгоритм побудови адитивної моделі даних двофакторного експерименту при незалежній дії факторів.
34. Які ви знаєте непараметричні критерії перевірки гіпотези про відсутність ефектів обробки?
35. Сформулюйте критерій Фрідмана (довільні альтернативи). В чому полягає критерій Пейджа (альтернативи з упорядкуванням)?
36. Які види регресійних моделей ви знаєте?
37. Сутність основної задачі регресійного аналізу. Сформулюйте модель простої лінійної регресії. Які передумови повинні виконуватися при використанні методів регресійного аналізу? Сутність основної ідеї методу найменших квадратів.
38. Основні види нелінійної регресії. Обґрунтування і застосування методу найменших квадратів при побудові нелінійної регресії.
39. Які ви знаєте інструменти і стратегії дослідження зв'язків ознак?

40. Зв'язки номінальних ознак (таблиці спряженості). Зв'язок ознак, що вимірюються у шкалі порядків. Зв'язок ознак у кількісних шкалах.
41. Що таке коефіцієнт кореляції, нормальна кореляція?
42. Як виявити зв'язок ознак, виміряних у різних шкалах?
43. Що таке часовий ряд? Мета, етапи та методи аналізу часових рядів.
44. Як ви розумієте поняття детермінованої і випадкової складової часового ряду?
45. Що таке тренд? Сезонна і циклічна компоненти. Моделі тренду.
46. В якому вигляді будуються моделі випадкової компоненти?
47. Які ви знаєте числові характеристики часових рядів? Як отримати оцінки числових характеристик часових рядів?
48. Порядок аналізу часових рядів. Графічний метод аналізу. Які ви знаєте методи зведення до стаціонарності: виділення тренду?
49. Виділення сезонних ефектів, метод скожзких середніх, сезонні різницеві оператори, перетворення шкали.
50. Що таке графік корелограми, графік частинної автокореляційної функції, які їх особливості?
51. Як ви розумієте задачі дискримінантного аналізу?
52. Коли застосовуються методи кластерного аналізу, в чому їх особливості?
53. Які ви знаєте методи контролю якості? Можливості статистичних пакетів з виконання кластерного та дискримінантного аналізу.
54. Характеристика процесів, які можна віднести до класу швидких. Як може бути поставлено задачу ідентифікації швидких процесів?
55. Які види моделей швидких процесів ви знаєте? Які ви знаєте методи оцінювання параметрів моделей, представлених у вигляді функцій швидкого зростання, спадання.
56. Багатоекстремальність і обґрунтованість оцінок методом найменших квадратів експоненціальних моделей.
57. Постановка задачі обробки сигналів. Які ви знаєте типи обробки сигналів?
58. Які ви знаєте методи вторинної обробки сигналів? Назвіть методи компенсації локальних сигналів-перешкод. Методи адаптації.
59. Що таке спектральний аналіз? Які ви знаєте типи спектрів? Які ви знаєте типи спектрального аналізу?
60. Сутність алгоритму інтегральної апроксимації.

61. Які ви знаєте задачі застосування методів обробки сигналів у задачах механіки, біології, медицини тощо?
62. Застосування методу Фур'є при дослідженні задач обробки сигналів.
63. Сутність методу регуляризації Тихонова.
64. Що таке інтерполяція, екстраполяція, згладжування та апроксимація? Як ви розумієте сутність лінійної інтерполяції та екстраполяції, квадратичної інтерполяції та екстраполяції.
65. Що таке сплайни, кубічні інтерполюючі сплякни? Сутність лінійної та нелінійної апроксимації (сгладжування).

СПИСОК ЛІТЕРАТУРИ

Основна

1. *Торин Ю. Н., Макаров А. А., Фигурнов Н. Р.* Анализ данных на компьютере: Учебник. — М: ИНФРА-М, 2003. — 544 с.
2. *Гурман В. Е.* Теория вероятностей и математическая статистика. — М, 1972. — 345 с.
3. *Фишер Р. А.* Статистические методы для исследователей. — М., 1958. — 537 с.
4. *Айвазян С. А., Енюков И. С., Мешалкин Л. Д.* Прикладная статистика: исследование зависимостей. — М.: Финансы и статистика, 1985.
5. *Семенов М. И., Трубилин И. Т., Лойко В. И., Барановская Т. П.* Автоматизированные информационные технологии в экономике: Учебник. — М.: Финансы и статистика, 2002. — 416 с.
6. *Андронов А. М.* Теория вероятностей и математическая статистика: Учебник. — СПб.: Питер, 2004. — 461 с.
7. *Минько А. А.* Статистический анализ в MS Excel. — М.: Диалектика, 2004. — 448 с.
8. *Бойко И. И.* Маркетинговые исследования (минимальный маркетинговый эксперимент): Учеб. пособие. — К.: Кондор, 2005. — 280 с.
9. *Петренко П. А.* Обработка данных в вычислительных системах и сетях / П. А. Петренко, Г. С. Теслер. — К.: Техника, 1980. — 232 с.
10. *Прокопенко І. Ф.* Комп'ютеризація економічного аналізу (теорія, практика): Навч. посіб. / І. Ф. Прокопенко, В. І. Ганін, В. В. Москаленко. — К.: ЦНЛ, 2005. — 340 с.

11. *Бабак В. П., Хандецький В. С., Шрюфер Е.* Обробка сигналів: Підруч. для студ. техн. спец. вищ. закл. освіти. — 2-ге вид., переробл. і допов. — К.: Либідь, 1999. — 496 с.

Додаткова

12. *Гнеденко Б. П.* Беседы о математической статистике. — М., 1965.
13. *Зинченко Г. П.* Социология для менеджеров. — Ростов н/Д: Феникс, 2001. — 352 с.
14. *Телетов О. С.* Маркетинг у промисловості: Підручник. — К.: ЦНЛ, 2004. — 248 с.
15. *Дацюк О. А.* Аналіз швидких процесів: — 2-ге вид., переробл. і допов. — К.: Політехніка, 2004. — 256 с.
16. *Шеффе Г.* Дисперсионный анализ. — М.: Наука, 1980.
17. *Холлендер М., Вулф Д.* Непараметрические методы статистики. — М.: Издат. дом “Вильямс”, 2003.
18. *Френкс Л.* Теория сигналов: Пер. с англ. / Под ред. Д. Е. Вакмана. — М.: Сов. радио, 1974. — 344 с.

МАУП

ЗМІСТ

Пояснювальна записка	3
Тематичний план дисципліни “Аналіз даних”	4
Зміст дисципліни “Аналіз даних”	5
Питання для самоконтролю	10
Список літератури	13

Відповідальний за випуск	<i>Ю. В. Нешкуренко</i>
Редактор	<i>Т. М. Тележенко</i>
Комп’ютерне верстання	<i>Т. Г. Замура</i>

Зам. № ВКЦ-4673

Формат 60×84/₁₆. Папір офсетний.

Друк ротатійний трафаретний. Наклад пр.

Міжрегіональна Академія управління персоналом (МАУП)
03039 Київ-39, вул. Фрометівська, 2, МАУП

ДП «Видавничий дім «Персонал»
03039 Київ-39, просп. Червонозоряний, 119, літ. XX

*Свідоцтво про внесення до Державного реєстру
суб’єктів видавничої справи ДК № 3262 від 26.08.2008*