

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ імені ІГОРЯ СІКОРСЬКОГО»
НАВЧАЛЬНО-НАУКОВИЙ КОМПЛЕКС
«ІНСТИТУТ ПРИКЛАДНОГО СИСТЕМНОГО АНАЛІЗУ»

Затверджено
Вченою радою ННК «ІПСА»
Протокол від 28.02.2017 № 2

Заступник голови

В.Д.Романенко

«28» 2017 р.



ПРОГРАМА

додаткового вступного випробування
на навчання за освітньо-професійною програмою підготовки «Магістр»
спеціальність 122 «Комп'ютерні науки та інформаційні технології»
спеціалізація «Інтелектуальний аналіз даних і управлінні проектами»

Програму рекомендовано
атестаційною комісією
кафедри математичних методів
системного аналізу
Протокол засідання кафедри
від 15.02.2017 р. № 5
В. о. завідувача кафедри

О.Л.Тимошук

«15» 02 2017 р.

Київ
2017

ВСТУП

Додаткове вступне випробування проводиться тільки для тих вступників, напрям підготовки (бакалаврат) яких не відповідає обраний для вступу на навчання за освітньо-професійною програмою підготовки «Магістр» спеціальності 122 «Комп'ютерні науки та інформаційні технології» (спеціалізація «Інтелектуальний аналіз даних в управлінні проектами»).

Мета додаткового вступного випробування — виявити достатність початкового рівня знань вступника в області напряму підготовки обраної для вступу спеціальності 122 «Комп'ютерні науки та інформаційні технології».

Програму додаткового вступного випробування розроблено атестаційною комісією кафедри математичних методів системного аналізу (ММСА) Інституту прикладного системного аналізу, ухвалено Вченою радою та затверджено заступником директора інституту.

Програма додаткового вступного випробування ґрунтується на програмах дисциплін нормативної частини заявленого напряму підготовки. В програму включено питання з дисциплін: «Прикладна статистика», «Дослідження операцій», «Бази даних та інформаційні системи».

СПИСОК ДИСЦИПЛІН ПРОГРАМИ

1. Прикладна Статистика
2. Дослідження операцій
3. Бази даних та інформаційні системи

Цей екзамен проводиться дві академічних години без перерви (90 хвилин). Фіксується час початку і закінчення роботи.

ЗМІСТ ПРОГРАМИ

1. Прикладна статистика

1.1. Два значення слова «статистика». Чотири етапи статистичного аналізу даних. Два типи даних: часові ряди і часові перерізи; два методи збору статистичних даних: суцільний і вибіркового. Планування експерименту і збір даних. Попередня обробка та дослідження (тестування) даних. Метод заповнення пропусків у часових рядах за допомогою оцінок прогнозів, нормування даних. Оптимальна фільтрація даних.

1.2. Поняття детермінованого і випадкового процесу. Випадкові процеси з детермінованими складовими. Формулювання гіпотез стосовно типів розподілів ймовірностей, можливих ситуацій. Деякі типи розподілів дискретних випадкових величин: нормальний, t –розподіл Стюдента, експоненціальний розподіл.

1.3. Математичні і статистичні моделі. Поняття структури математичної моделі. Стаціонарні та нестаціонарні процеси, тестування фактичних даних (тести на інтегрованість, гетероскедастичність, лінійність). Встановлення належності даних до дискретного розподілу конкретного типу: коефіцієнт асиметрії, ексцес, статистика Жак-Бера, статистика χ^2 , критерії Колмогорова-Смірнова та Андерсона-Дарлінга.

1.4. Загальна методика перевірки гіпотез. Нульова і альтернативна гіпотези. Поняття рівня значущості. Помилки першого і другого роду. Критерій перевірки нульової гіпотези. Перевірка гіпотези стосовно значущості оцінок параметрів моделей.

1.5. Задача оцінювання параметрів статистичних моделей, формулювання гіпотез стосовно значущості оцінок та якості (адекватності) статистичних моделей. Перевірка сформульованих гіпотез стосовно оцінок параметрів моделей, прийняття рішення стосовно адекватності моделі.

1.6. Кореляційний аналіз даних. Неперервні і дискретні кореляційні функції. Теоретична автокореляційна функція (АКФ), рівняння Юла-Уокера. Вибіркові автокореляційна і часткова автокореляційна функції, функція взаємної кореляції двох величин. Кореляційна матриця.

1.7. Регресійний аналіз. Типи регресійних рівнянь. Вимоги до даних, оцінок параметрів моделі та до моделі у цілому. Методика побудови регресійної моделі. Адекватність побудованої моделі, критерії адекватності моделі.

1.8. Метод найменших квадратів (МНК), властивості оцінок, обчислених за МНК. Рекурсивний метод найменших квадратів (РМНК). Можливі випадки застосування РМНК. Порівняння властивостей оцінок, отриманих за МНК, з властивостями оцінок, обчислених за РМНК.

1.9. Оцінювання параметрів за методами максимальної правдоподібності та Монте Карло для марковських ланцюгів.

1.10. Метод головних компонент: призначення, алгоритм реалізації.

Література

Бідюк П.І. Прикладна статистика / П. І. Бідюк, О.М. Терентьев, Т.І. Просянкін-Жарова. — Вінниця: «Едельвейс і К», 2013. — 288 с.

Мелник М. Основы прикладной статистики / Мелник М. — М.: Энергоатомиздат, 1983. — 415 с.

Сигел Э.Ф. Практическая бизнес-статистика / Э.Ф. Сигел. — СПб: Вишьямс, 2002. — 1052 с.

Сеньо П.С. Теорія ймовірностей та математична статистика / С. С. Сеньо. — Київ: Центр навчальної літератури, 2004. — 448 с.

Gelman A. Bayesian Data Analysis / A. Gelman, J.B. Carlin, H.S. Stern, D.B. Rubin. — New York: Chapman & Hall, 2000. — 670 с.

Чернов Г. Элементарная теория статистических решений / Г. Чернов, Л. Мозес. — М.: «Советское Радио», 1982. — 407 с.

Хей Дж. Введение в методы байесовского статистического вывода / Дж. Хей. — М.: «Финансы и статистика», 1987. — 336 с.

Андерсон Т. Статистический анализ временных рядов / Т. Андерсон. — М.: Мир, 1976. — 755 с.

Бідюк П.І. Моделі і методи математичної статистики / П. І. Бідюк. — Київ: НТУУ «КПІ» (електронне видання), 2014. — 710 с.

Casella G. Statistical Inference / G. Casella, R.L. Berger. — Duxbury (USA): Thomson Learning, 2002. — 660 p.

Приклад екзаменаційного завдання

Знайти внутрішньогрупову, міжгрупову і загальну дисперсії сукупності, що складається з двох груп, наведених нижче:

$x_1(i)$	5,5	8,4	12,2	20,5
$n_1(i)$	8	10	10	4

$x_2(i)$	15,0	8,0	10,5	0,5
$n_2(i)$	5	8	10	12

У цих таблицях $n_1(i), n_2(i)$ — частоти значень відповідно для $x_1(i), x_2(i)$.

Розв'язок:

Груповою дисперсією називають дисперсію значень змінної, що належить групі, відносно групової середньої:

$$D_{\text{групи}} = \frac{1}{N_j} \sum_{i=1}^{l_j} n_{x_i} (x_i - \bar{x}_j)^2,$$

де n_{x_i} — частота значень x_i у групі j ; j — номер групи; $\bar{x}_j$ — групова середня; $N_j = \sum_{i=1}^{l_j} n_{x_i}$ — обсяг групи; l_j — кількість підгруп у групі j .

Внутрішньогруповою дисперсією називають середню арифметичну дисперсію, зважену за обсягами груп:

$$D_{\text{внутрігруп}} = \frac{1}{N} \sum_{i=1}^m n_i D_{i \text{ групи}},$$

де n_i — обсяг групи i ; $N = \sum_{i=1}^m n_i$ — обсяг усієї сукупності даних; m — кількість груп.

Міжгруповою дисперсією називають дисперсію групових середніх відносно загальної середньої:

$$D_{\text{міжгрупова}} = \frac{1}{N} \sum_{j=1}^m n_j (\bar{x}_j - \bar{x})^2,$$

де $\bar{x}_j$ – групове середнє групи j ; n_j – обсяг групи j ; $\bar{x}$ – загальне середнє;
 $N = \sum_{j=1}^m n_j$ – обсяг усієї сукупності даних.

Спочатку обчислимо групові дисперсії за виразом $D_{\text{груп}} = \bar{x}_1^2 - (\bar{x}_1)^2$:

$$\begin{aligned} \bar{x}_1^2 &= \frac{1}{8+10+10+4} (5,5^2 \cdot 8 + 8,4^2 \cdot 10 + 12,2^2 \cdot 10 + 20,5^2 \cdot 4) = \\ &= \frac{1}{32} (30,25 \cdot 8 + 70,56 \cdot 10 + 148,84 \cdot 10 + 420,25 \cdot 4) = 128,66. \end{aligned}$$

$$\begin{aligned} \bar{x}_2^2 &= \frac{1}{5+8+10+12} (15^2 \cdot 5 + 8^2 \cdot 8 + 10,5^2 \cdot 10 + 0,5^2 \cdot 12) = \\ &= \frac{1}{35} (225 \cdot 5 + 64 \cdot 8 + 110,25 \cdot 10 + 0,25 \cdot 12) \approx 78,36. \end{aligned}$$

$$\begin{aligned} \bar{x}_3^2 &= \frac{1}{6+5+6+9} (14^2 \cdot 6 + 5^2 \cdot 5 + 15,5^2 \cdot 6 + 3^2 \cdot 9) = \\ &= \frac{1}{26} (196 \cdot 6 + 25 \cdot 5 + 240,25 \cdot 6 + 9 \cdot 9) = \frac{470,25}{26} \approx 18,1. \end{aligned}$$

$$\begin{aligned} \bar{x}_4^2 &= \frac{1}{11+8+8+4} (7^2 \cdot 11 + 12,5^2 \cdot 8 + 6,5^2 \cdot 8 + 25,5^2 \cdot 4) = \\ &= \frac{1}{31} (49 \cdot 11 + 156,25 \cdot 8 + 42,25 \cdot 8 + 650,25 \cdot 4) = \frac{4728}{31} \approx 152,52. \end{aligned}$$

$$\begin{aligned} D_{1\text{груп}} &= \bar{x}_1^2 - (\bar{x}_1)^2 = 128,66 - \left[\frac{1}{8+10+10+4} (5,5 \cdot 8 + 8,4 \cdot 10 + 12,2 \cdot 10 + 20,5 \cdot 4) \right]^2 = \\ &= 128,66 - \left(\frac{332}{32} \right)^2 = 128,66 - (10,38)^2 = 21,02. \end{aligned}$$

$$\begin{aligned} D_{2\text{груп}} &= \bar{x}_2^2 - (\bar{x}_2)^2 = 78,36 - \left[\frac{1}{5+8+10+12} (15 \cdot 5 + 8,0 \cdot 8 + 10,5 \cdot 10 + 0,5 \cdot 12) \right]^2 = \\ &= 78,36 - \left(\frac{250}{35} \right)^2 = 78,36 - (7,14)^2 = 27,38. \end{aligned}$$

Внутрішньогрупова дисперсія для двох перших груп:

$$D_{\text{внутрішньогруп}} = \frac{1}{N} \sum_{i=1}^m n_i D_{i\text{групи}} = \frac{1}{30+35} (21,02 \cdot 32 + 27,38 \cdot 35) = \frac{672,64 + 958,31}{67} = \frac{1630,95}{67} = 24,34.$$

Для обчислення між групової дисперсії знайдемо спочатку загальне середнє значення:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^M x_i n_i = \frac{1}{32+35} \left(5,5 \cdot 8 + 8,4 \cdot 10 + 12,2 \cdot 10 + 20,5 \cdot 4 + 15 \cdot 5 + 8,0 \cdot 8 + 10,5 \cdot 10 + 0,5 \cdot 12 \right) = \frac{582}{67} \approx 8,7,$$

де $M = \sum_{l=1}^m n_l$ – загальна кількість різних значень в усіх групах (тобто загальна кількість підгруп). Тепер

$$D_{\text{міжгрупова}} = \frac{1}{N} \sum_{j=1}^m n_j (\bar{x}_j - \bar{x})^2 = \frac{1}{32+35} [32(10,38 - 8,7)^2 + 35(7,14 - 8,7)^2] = \frac{1}{67} [32 \cdot 2,82 + 35 \cdot 2,43] = \frac{175,42}{67} = 2,62.$$

2. Дослідження операцій

2.1. Лінійне програмування (ЛП)

Метод розв'язання задач ЛП з довільним видом обмежень, оснований на штучних змінних.

Двоїста задача ЛП. Двоїстий симплекс-метод.

Метод оберненої матриці.

Дослідження моделей ЛП-задач на чутливість.

Транспортні задачі. Метод потенціалів.

2.2. Дискретне програмування (ДП)

Метод відсікаючих площин Гоморі.

Метод гілок та меж.

Метод гілок та меж в задачі комівояжера.

Метод послідовного аналізу та відсіву варіантів (ПАВ) в задачі ЛЦП.

Метод послідовного аналізу та відсіву варіантів (ПАВ) в задачі булево програмування.

2.3. Нелінійне програмування

Метод множників Лагранжа.

Задача квадратичного програмування. Умови оптимальності Куна-Таккера для задач квадратичного програмування.

Метод розв'язання загальної задачі геометричного програмування ГП з ступенем складності $d > 0$.

Методи можливих напрямків. Метод Зойтендейка у випадку лінійних обмежень.

Методи можливих напрямків. Метод Зойтендейка у випадку нелінійних обмежень-нерівностей.

Прямі методи пошуку.

2.4. Динамічне програмування

Послідовні задачі прийняття рішень.

Багатовимірні задачі динамічного програмування.

Динамічні задачі управління запасами.

Література

Зайченко Ю. П. Дослідження операцій / Ю. П. Зайченко. — К.: Слово, 2001. — 688 с.

Лэдсон Л. Оптимизация больших систем / Л. Лэдсон. — М.: Наука. — 431 с.

Михайлович В. С. Методы последовательной оптимизации в дискретных сетевых задачах оптимального распределения ресурсов / В. С. Михайлович, А. И. Кукса. — М.: Наука. — 208 с.

Зайченко Ю. П. Исследование операций / Ю. П. Зайченко. — К.: Вища школа. — 552 с.

Даффин Р. Геометрическое программирование / Р. Даффин. — М.: Мир. — 308 с.

Хедли Дж. Нелинейное и динамическое программирование / Дж. Хедли. — М.: Мир. — 352 с.

Навчально-методичний посібник до практичних занять з курсу «Математичні методи оптимізації» для студентів магістратури усіх спеціальностей / Уклад. О.Ю.Зайченко. — К.: Політехніка. — 88 с.

Приклад екзаменаційного завдання

1. Динамічні задачі керування запасами. Постановка задачі та опис алгоритму.

2. Дослідити таку задачу лінійного програмування на чутливість.

Підприємство випускає вироби двох видів, для випуску яких використовуються 2 види ресурсів. Нехай прибуток від продажу виробів складає відповідно $C_1=2$, $C_2=3$, а об'єми ресурсів складають відповідно $b_1=32$, $b_2=24$.

Матриця технологічних витрат має вигляд $A = \begin{pmatrix} 3 & 2 \\ 1 & 2 \end{pmatrix}$.

Нехай підприємство може додатково випускати третій вид продукції, для якого $c_3=4$, а норми витрат ресурсів дорівнюють $a_{13}=3$; $a_{23}=2$. Знайти оптимальний план за цієї умови.

3. Знайти мінімальну довжину шляху для задачі комівояжера:

X	5	8	2	7	3
2	X	9	7	1	5
5	3	X	2	6	7
C=7	8	8	X	2	10
9	3	7	9	X	2
10	1	5	2	7	X

3 Бази даних та інформаційні системи

3.1. Основні поняття. Інформація, дані, знання. Аспекти роботи з даними. Поняття про інформаційні технології (ІТ). Особливості та завдання інформаційних систем (ІС). Файл, запис, поле. Файлові інформаційні системи (ФІС). Ідея СКБД, відміни від ФІС. Визначення банку даних (БнД). Вимоги до БнД. Переваги централізації керування даними.

3.2. Технології створення ІС. Життєвий цикл інженерного виробу. Моделі життєвого циклу розробки ІС. Задачна модель. Каскадна модель. Спіральна модель. Загальна технологія створення ІС та АС. Основи побудови банків даних БнД.

3.3. Моделі даних. Поняття моделі даних (МД). Сильно- та слабкоструктуровані МД. Структури, операції, обмеження МД. Модель «сутність-зв'язок». Типи зв'язків. Степені зв'язку, залежність по коду. N-зв'язки. Композиція зв'язків. Типи і підтипи (ролі). Поняття життєвого циклу об'єкта (екземпляр сутності). Початок, закінчення, координація ЖЦ. Обмеження цілісності, бізнес-правила. Локальні інфологічні моделі. Побудова глобальної інфологічної моделі.

3.4. Реляційна МД. Реляційні алгебра та числення. Базові поняття реляційних баз даних. Реляційна модель даних. Реляційна алгебра та її операції. Реляційне числення на кортежах. Реляційне числення на доменах.

3.5. Нормалізація БД. Групування атрибутів у відношення. Аномалії та їхні види. 1НФ, 2НФ, 3НФ, BCNF, 4НФ, 5НФ.

3.6. Дореляційні МД. Ієрархічна МД. Мережева МД.

3.7. Мова БД SQL. Мови опису даних і маніпулювання даними. Реляційні операції як команди мови маніпулювання даними. Віртуальні атрибути і таблиці. Приклади використання операторів Insert, Update та Delete.

3.8. Зміна порядку виведення рядків (ORDER BY). Усунення дублювання (модифікатор DISTINCT). З'єднання (JOIN). Використання агрегатних функцій з угрупованням. Порядок виконання оператора SELECT.

3.9. Тригери та цілісність посилення. Збережені процедури. Використання курсорів. Динамічний SQL.

3.10. Фізична організація ІС. Структура зовнішньої пам'яті. Зберігання таблиць. Індеси та В-дерева. Хеш-функція. Службова інформація.

3.11. Система «клієнт—сервер». Модель файлового серверу. Модель віддаленого доступу до даних. Модель сервера бази даних. Модель сервера додатків. Відкриті системи.

3.12. Транзакції. Транзакції і цілісність БД. Рівні ізоляваності транзакцій (PIT). Серіалізація транзакцій. Конфлікти між транзакціями. Синхронізаційні захоплення. Гранульовані синхронізаційні захоплення. Предикатні синхронізаційні захоплення. Поняття глухого кута (Dead End). Граф очікування транзакцій. Метод тимчасових міток.

3.13. Журналізація змін БД. Журналізація і буферизація. Індивідуальний відкат транзакції. Відновлення після м'якого збою. Фізична узгодженість бази даних. Відновлення після жорсткого збою.

Література

Хомоненко А.Д. Базы данных: учебник для вузов по техн. и экон. спец. / А.Д.Хомоненко, В.М.Цыганков, М.Г. Мальцев [под ред. А.Д.Хомоненко]. — С-Пб: Корона принт, 2002. — 672 с.

Дейт К. Введение в системы баз данных [8-е изд.: пер. с англ.] / Дейт К. — М.: Издательский дом «Вильямс», 2005. — 1328 с.

Виейра Р. Программирование баз данных Microsoft SQL Server 2005. Базовый курс [пер. с англ.] / Виейра Р. — М.: Диалектика, 2008. — 832 с.

Форта Б. Освой самостоятельно SQL. 10 минут на урок [пер. с англ.] / Форта Б. — М.: Вильямс, 2006. — 288 с.

Исаченко А.Н. Модели данных и системы управления базами данных: пособие для студентов / Исаченко А.Н., Бондаренко С.П. — Минск: БГУ, 2007. — 220 с.

Грабер М. Введение в SQL / Грабер М. — М.: Лори, 2010. — 228 с.

Маклаков С.В. BPwin, ERwin CASE-средства разработки информационных систем / Маклаков С.В. — М.: Диалог-МИФИ, 2001. — 304 с.

Пасічник В.В. Організація баз даних та знань / Пасічник В.В., Резніченко В.А. — К.: Видавнича група BVH, 2006. — 384 с.

Коннолли Т. Базы данных: проектирование, реализация и сопровождение. Теория, практика [3-е изд.: пер. с англ.] / Коннолли Т., Бегг К. — М.: Издательский дом «Вильямс», 2003. — 1440 с.

Приклад екзаменаційного завдання

1. Продажа топлива на АЗС

Задана схема БД, яка складається з двох таблиць: «Типи палива» [fuel (id_fuel (PK), fuel_name, fuel_price)] та «Заправка авто» [filling (id_filling (PK), id_fuel (FK), filling_size, date_filling)]. Таблиця fuel містить перелік палив (дизельне паливо, А-92, А-95) та ціну fuel_price за літр. В таблиці filling фіксується заправка авто, в полі date_filling зберігається дата заправки, в полі filling_size – кількість придбаних літрів.

Відповідно до заданої схеми написати запит до бази даних для обчислення сумарної вартості проданого палива за типом, які реалізовані

водіям з конкретної дати D0 до конкретної дати D1 включно, тобто $\text{date_filling} \geq D0$ і $\text{date_filling} \leq D1$. Результат повинен мати вигляд: назва палива, розрахована сума вартості за паливо.

Розв'язок:

```
SELECT f.fuel_name, sum(f.fuel_price*i.filling_size)
FROM fuel f, filling i
WHERE f.id_fuel = i.id_fuel
and (i.date_filling >= D0 and i.date_filling <= D1)
GROUP BY f.fuel_name
```

ПРИКІНЦЕВІ ПОЛОЖЕННЯ

Користування допоміжним матеріалом на екзамені

— забороняється

Критерії оцінювання і принципи виставлення кінцевої оцінки (за системою ECTS, 100-бальна шкала)

Розв'язання кожної задачі оцінюється за такими критеріями:

95—100	—	задачу розв'язано повністю, вірно
85—94	—	задачу розв'язано вірно, відповідь правильна, але наявними є один-два недоліки (наявними є деякі методичні помилки, порушено послідовність викладок тощо)
75—84	—	задачу розв'язано вірно, але відповідь неправильна (наявними є арифметичні помилки)
65—74	—	задачу розв'язано неповністю, але намічено правильний хід розв'язування
60—64	—	задачу не розв'язано, але наведено формули або твердження, що можуть бути використані при розв'язуванні задачі
менше 60	—	задачу не розв'язано

Результат роботи обчислюється як середнє арифметичне оцінок, що їх отримано за кожну задачу і заокруглюється до цілих.

Розробники програми

д.т.н., проф. Бідюк П.І.

д.т.н., доц. Зайченко О.Ю.

к.т.н. Діденко Д.Г.