

Міністерство освіти і науки України
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ”

Кафедра прикладної математики

МЕТОДИЧНІ РЕКОМЕНДАЦІЇ

до виконання лабораторних робіт
з кредитного модуля "Аналіз даних"

Затверджено на засіданні
кафедри прикладної математики

Протокол №

від « » червня 2017 р.

Завідувач кафедри
прикладної математики

_____ О.Р. Чертов

« ____ » _____ 2017 р.

Методичні рекомендації до виконання лабораторних робіт з кредитного модуля "Аналіз даних" для студентів за напрямом підготовки 6.040301 Прикладна математика освітньо-кваліфікаційного рівня бакалавр за денною формою навчання складена відповідно до програми навчальної дисципліни «Аналіз даних».

Розробники методичних рекомендацій:

доцент кафедри прикладної математики, д.ф.-м.н., доцент Пашко
Анатолій Олексійович

Методичні рекомендації затверджено на засіданні кафедри прикладної математики.

Протокол № ____ від «____» _____ 2017 р.

Завідувач кафедри

_____ О.Р. Чертов

«____» _____ 2017 р.

Мета та завдання лабораторних робіт

Метою лабораторного практикуму є формування у студентів навичок і вмінь пов'язаних з аналізом та обробкою даних:

аналізувати дані, що використовуються в інформаційних системах;
обирати інформаційну технологію аналізу даних відповідно до визначених вимог до інформаційної системи;
забезпечувати обробку даних з використанням сучасних інформаційних технологій;

виконувати розрахунок основних характеристик вибіркового даних.

Під час викладання лекційного матеріалу з кредитного модуля «Аналіз даних» потрібно взяти до уваги наступне:

- в лекційному матеріалі слід зосередитись на теоретичних аспектах теорії аналізу прикладних даних,
- вивчення програмного забезпечення для аналізу даних на лекційних заняттях є недоцільним.

При проведенні лабораторних робіт слід акцентувати увагу студентів на прикладні аспекти аналізу даних, а саме аналіз даних для різних прикладних задач і оцінювання основних характеристик.

Основні завдання циклу лабораторних робіт: аналіз характеристик статистичних даних та їх реалізація в інформаційних системах для конкретних практичних задач.

Методичні рекомендації з виконання лабораторних робіт

Лабораторне заняття, як одна із форм навчальних занять, розрахована на виконання студентами в електронному вигляді певної задачі з використанням ПК. Викладач організує індивідуальну роботу студентів на ПЕОМ з метою формування умінь та навичок практичного використання певних оболонок, програм тощо. На лабораторному занятті студент під керівництвом викладача проводить експерименти або дослідження з метою практичного підтвердження окремих теоретичних положень, набуває умінь роботи з сучасним програмним забезпеченням, обчислювальною технікою, вимірювальною апаратурою, оволодіває методикою експериментальних досліджень в конкретній предметній галузі та обробки отриманих результатів.

Під час виконання лабораторної роботи створюються умови для перевірки та виявлення інтелектуального рівня студентів.

Навчальні програми з переліком тем та питань дисципліни «Аналіз даних» студенти отримують на першому практичному занятті. Для самостійного опанування тем предмета студенти можуть використовувати не тільки зазначений список основної літератури, а також інші джерела інформації, можливості Internet та додаткову літературу.

На першому занятті викладач вказує на основні теми предмета для практичного засвоєння, роз'яснює загальні положення, надає рекомендації по вивченню та опануванню всіх розділів, загострює увагу на найбільш

важливих «вузлових» питаннях. На заключній лабораторній роботі розглядаються теми, які стали найбільш важкими для самостійного опанування.

ТЕМИ І ПЛАНИ ЛАБОРАТОРНИХ ЗАНЯТЬ

Лабораторна робота 1. Організація та проведення вибіркового обстеження і оцінювання основних параметрів.

Завдання: Сплануйте власне обстеження. Сформулюйте мету обстеження, визначте популяцію та фрейм даних. Складіть анкету вибіркового обстеження. Які величини будуть вивчатись і які параметри популяції підлягають оцінюванню. Які методи збору даних можуть бути застосовані. Проведіть обстеження і підготуйте звіт.

Література: 1,2.

Завдання на СРС: Обстеження проводиться за допомогою Інтернет (знайдіть приклад). Опишіть план обстеження. Проаналізуйте помилки, що виникають при проведенні таких обстежень.

Методичні вказівки до проведення вибіркового обстеження.

В якості популяції для дослідження виберіть:

- м. Київ
- студентська спільнота м. Києва
- студентська спільнота НТУУ КПІ
- студентська спільнота ФПМ
- студентська спільнота 3 курсу ФПМ.

Сплануйте вибірконе обстеження для дослідження параметрів:

- середній зріст;
- підтримка європейського вибору України (в середньому і всього);
- вболівальники "Динамо" Київ (в середньому і всього);
- індекс IQ;
- час знаходження в мережі Інтернет.

Оформіть звіт, в якому відобразіть:

- ціль вибіркового обстеження
- визначення популяції
- планування вибірки
- побудова вибіркової схеми
- одержання вибірки для обстеження
- зібрані дані
- результати обробки даних
- точність результатів.

Лабораторна робота 2. Оцінювання параметрів статистичної вибірки.

Завдання: Для заданої вибірки оцінити основні параметри популяції. Розглянути випадки, коли вибірка – випадкова вибірка без повернення, систематична вибірка, стратифікована вибірка. Розробити програмне забезпечення. Порівняти отримані результати. Оформити звіт.

Література: 1-2.

Завдання на СРС: Сплануйте вибіркове обстеження для оцінювання якості обслуговування в супермаркетах. Проведіть його та оцініть результати.

Методичні вказівки до виконання.

Довжина популяції N . Деякий параметр популяції $\{y_i\}, i = 1, 2, \dots, N$ змодельуйте як рівномірний розподіл на відрізку $[10 + M, MG + 2M]$, M – номер студента в списку групи, MG – номер групи.

Вибрати число M – об'єм вибірки.

Простим випадковим вибором сформууйте вибірку $\{x_i\}, i = 1, 2, \dots, M$ із популяції.

А). Обчисліть середнє значення і дисперсію для популяції

$$Y = \sum_{i=1}^N y_i, \quad Y_s = \frac{1}{N} \sum_{i=1}^N y_i, \quad DY^2 = \frac{\sum_{i=1}^N (y_i - Y_s)^2}{N - 1}.$$

Б). Оцініть суму і середнє значення для популяції за вибіркою

$$\hat{Y} = N \sum_{i=1}^M \frac{x_i}{M}, \quad \hat{Y}_s = \frac{1}{M} \sum_{i=1}^M x_i, \quad S^2 = \frac{\sum_{i=1}^M (x_i - \hat{Y}_s)^2}{M - 1}.$$

Дисперсія оцінок суми і середнього в залежності від об'єму випадкової вибірки:

$$D(\hat{Y}_s) = \frac{S^2}{M} \left(1 - \frac{M}{N}\right) \text{ та } D(\hat{Y}) = \frac{S^2 N^2}{M} \left(1 - \frac{M}{N}\right)$$

Верхній і нижній $100(1-\alpha)\%$ довірчі інтервали для середнього значення і загальної суми популяції відповідно дорівнюють

$$\hat{Y}_s \pm \frac{z_{1-\frac{\alpha}{2}} S}{M} \sqrt{1 - \frac{M}{N}} \quad \text{та} \quad N\hat{Y}_s \pm \frac{z_{1-\frac{\alpha}{2}} SN}{\sqrt{M}} \sqrt{1 - \frac{M}{N}}$$

де z є відповідним $1-\alpha/2$ квантилем нормального розподілу $N(0,1)$.

Найбільш часто використовуються наступні квантилі:

Рівень довіри (%)	50	80	90	95	99
z	0,67	1,28	1,64	1,96	2,58

В). Написати програму для обчислення основних показників вибіркового обстеження. Провести розрахунки для

$N = 10000$, $M = 100$ та $M = 1000$.

Порівняти результати.

Г). Сформуувати систематичну вибірку таких же розмірів. Порівняти отримані результати.

Зауваження. Наведені формули для довірчих інтервалів використовуються при розмірах вибірок $M > 50$, при менших розмірах величину z беруть з таблиць розподілу Ст'юдента з $(n-1)$ степенями свободи.

Д). Провести аналогічні розрахунки для стратифікованої вибірки.

Використаємо наступні позначення:

y_{hi} - значення i -го елемента зі страти h ,

$w_h = \frac{N_h}{n_h}$ - вибіркова вага,

$\bar{y}_h = \frac{\sum_{i=1}^{N_h} y_{hi}}{N_h}$ - середнє значення для страти h ,

$Y_h = \sum_{i=1}^{N_h} y_{hi}$ - загальна сума для страти h ,

$Y = \sum_{h=1}^H Y_h$ - загальна сума для всієї популяції,

$\bar{y}_U = \frac{Y}{N} = \frac{\sum_{h=1}^H \sum_{i=1}^{N_h} y_{hi}}{N}$ - середнє значення для всієї популяції,

$S_h^2 = \frac{\sum_{i=1}^{N_h} (y_{hi} - \bar{y}_h)^2}{N_h - 1}$ - дисперсія в страті h ,

$W_h = \frac{N_h}{N}$ - вага страти h .

$\hat{\bar{y}}_h = \frac{\sum_{i \in b_h} y_{hi}}{n_h}$ - оцінка середнього для страти h , $h=1, \dots, N$

$\hat{Y}_h = \frac{N_h}{n_h} \sum_{i \in b_h} y_{hi}$ - оцінка загальної суми для страти h , $h=1, \dots, N$

$s_h^2 = \frac{\sum_{i \in b_h} (y_{hi} - \hat{\bar{y}}_h)^2}{n_h - 1}$ - оцінка вибіркової дисперсії для страти h , $h=1, \dots, N$.

Оцінки суми та середнього популяції при використанні стратифікованої вибірки будуть відповідно мати вигляд

$$\hat{Y}_{st} = \sum_{h=1}^H \hat{Y}_h = \sum_{h=1}^H N_h \hat{\bar{y}}_h,$$

$$\hat{\bar{y}}_{st} = \frac{\hat{Y}_{st}}{N} = \sum_{h=1}^H \frac{N_h}{N} \hat{\bar{y}}_h = \sum_{h=1}^H W_h \hat{\bar{y}}_h.$$

З останньої формули видно, що середнє значення популяції є зваженим середнім від середніх для кожної страти з вагами, які дорівнюють відносним розмірам страт.

Незміщена оцінка для дисперсії:

$$\hat{D}(\hat{Y}_{st}) = \sum_{h=1}^H \left(1 - \frac{n_h}{N_h}\right) N_h^2 \frac{s_h^2}{n_h}$$

$$\hat{D}(\hat{y}_{st}) = \frac{1}{N^2} D(\hat{Y}_{st}) = \sum_{h=1}^H \left(1 - \frac{n_h}{N}\right) \left(\frac{N_h}{N}\right)^2 \frac{s_h^2}{n_h}$$

100(1- α)% довірчий інтервал для оцінки середнього значення визначається за формулою $\hat{y}_{st} \pm z_{\frac{\alpha}{2}} \sqrt{\hat{D}(\hat{y}_{st})}$, де $z_{\frac{\alpha}{2}}$ є відповідним квантилем стандартного нормального розподілу $N(0,1)$.

Лабораторна робота 3. Алгоритми одно факторного і двох факторного дисперсійного аналізу.

Завдання: Провести дисперсійний аналіз даних, відповідно до варіанту, при довірчій ймовірності $\alpha=0.95$.

Провести двох факторний дисперсійний аналіз даних, відповідно до варіанту, при довірчій ймовірності $\alpha=0.95$.

Література: 5,10.

Завдання на СРС розробити програмне забезпечення для дисперсійного аналізу вибірки.

Методичні вказівки до виконання.

Номер студента в списку групи N .

1. Однофакторний аналіз

Таблиця однофакторного експерименту

n (j)	Рівні фактора A (i)				
	A ₁	A ₂	A ₃	A ₄	A ₅
1000	N+rnd(1)	0.5*N+rnd(1)	0.8*N+rnd(1)	1.4*N+rnd(1)	2*N+rnd(1)

Для кожного фактору знаходимо

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 = \frac{1}{n-1} \left[\sum_{j=1}^n x_{ij}^2 - \frac{1}{n} \left(\sum_{j=1}^n x_{ij} \right)^2 \right].$$

За припущенням дисперсійного аналізу - повинна мати місце рівність дисперсій. Перевірити рівність дисперсій за критерієм порівняння.

$$g = \frac{\max_{1 \leq i \leq k} s_i^2}{\sum_{i=1}^k s_i^2}.$$

Критерій порівняння подується за формулою

При $g > g_{\alpha}(k, n)$ нульова гіпотеза про рівність дисперсій відхиляється.

Значення статистики $g_{\alpha}(k, n)$ для $\alpha=0,95$

k	n					
	8	9	10	11	17	37
2	0,899	0,882	0,867	0,854	0,795	0,707
3	0,733	0,711	0,691	0,673	0,606	0,515
4	0,613	0,590	0,570	0,554	0,488	0,406
5	0,526	0,504	0,485	0,470	0,409	0,335
6	0,461	0,440	0,423	0,408	0,353	0,286
7	0,410	0,391	0,375	0,362	0,310	0,249
8	0,370	0,352	0,337	0,325	0,278	0,221
9	0,338	0,321	0,307	0,295	0,251	0,199
10	0,311	0,294	0,281	0,270	0,230	0,181
12	0,268	0,253	0,242	0,232	0,196	0,153
15	0,223	0,210	0,200	0,192	0,161	0,125
20	0,175	0,165	0,157	0,150	0,125	0,096
30	0,123	0,116	0,110	0,105	0,087	0,066

При виконанні припущення про рівність дисперсій, знаходимо оцінку дисперсії, що характеризує розсіювання поза впливом фактора,

$$S_0^2 = \frac{1}{k} \sum_{i=1}^k S_i^2 \frac{1}{k(n-1)} \sum_{i=1}^k \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 = \frac{1}{k(n-1)} \left[\sum_{i=1}^k \sum_{j=1}^n x_{ij}^2 - \frac{1}{n} \sum_{i=1}^k \left(\sum_{j=1}^n x_{ij} \right)^2 \right]$$

Знаходимо вибірккову дисперсію всіх спостережень

$$S^2 = \frac{1}{kn-1} \left[\sum_{i=1}^k \sum_{j=1}^n x_{ij}^2 - \frac{1}{kn} \left(\sum_{i=1}^k \sum_{j=1}^n x_{ij} \right)^2 \right]$$

Знаходимо оцінку дисперсії, що характеризує зміни параметра, пов'язані з фактором

$$S_A^2 = \frac{n}{k-1} \sum_{i=1}^k (\bar{x}_i - \bar{\bar{x}})^2. \quad \bar{\bar{x}} = \frac{1}{k} \sum_{i=1}^k \bar{x}_i; \quad \bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}.$$

Оцінка впливу фактора на зміни середнього значення визначається відношенням (вплив значущий з ймовірністю $1-\alpha$)

$$\frac{S_A^2}{S_0^2} > F_\alpha[k-1; k(n-1)]$$

де $F_\alpha(f_1, f_2)$ - α -квантиль F-розподілу з f_1 та f_2 степенями свободи.

2. Двохфакторний аналіз

Таблиця двофакторного експерименту

Рівні фактора B(j)	Рівні фактора A (i)				
	A ₁	A ₂	A ₃	A ₄	A ₅
B ₁	N+rnd(1)	3.5*N+rnd(1)	3.8*N+rnd(1)	1.4*N+rnd(1)	2*N+rnd(1)
B ₂	N+rnd(1)	2.5*N+rnd(1)	2.8*N+rnd(1)	2.4*N+rnd(1)	3*N+rnd(1)
B ₃	N+rnd(1)	1.5*N+rnd(1)	1.8*N+rnd(1)	3.4*N+rnd(1)	4*N+rnd(1)
B ₄	N+rnd(1)	0.5*N+rnd(1)	0.8*N+rnd(1)	4.4*N+rnd(1)	5*N+rnd(1)

В кожній клітинці таблиці зберігається масив із n=100 значень.

Якщо фактори A і B незалежні:

Знаходимо середнє значення в кожній клітинці x_{ij} .

Обчислюємо основні показники

$$Q_1 = \sum_{i=1}^k \sum_{j=1}^m x_{ij}^2; \quad Q_2 = \frac{1}{m} \sum_{i=1}^k X_i^2; \quad Q_3 = \frac{1}{k} \sum_{j=1}^m X_{j'}^2;$$

$$Q_4 = \frac{1}{mk} \left(\sum_{i=1}^k X_i \right)^2 = \frac{1}{mk} \left(\sum_{j=1}^m X_{j'} \right)^2.$$

X_i - сума по стовпчиках (рівень фактора A_i), $X_{j'}$ - сума по рядках (рівень фактора B_j).

Знаходимо оцінки дисперсій

$$S_0^2 = \frac{Q_1 + Q_4 - Q_2 - Q_3}{(k-1)(m-1)}; \quad S_A^2 = \frac{Q_2 - Q_4}{k-1}; \quad S_B^2 = \frac{Q_3 - Q_4}{m-1}.$$

Якщо $\frac{S_A^2}{S_0^2} > F_\alpha(f_1, f_2)$ для $f_1=k-1$, $f_2=(k-1)(m-1)$ то фактор A є значущим. Аналогічно для фактора B при $f_1=m-1$, $f_2=(k-1)(m-1)$.

Якщо фактори A і B залежні:

Знаходимо додатково

$$Q_5 = \sum_{i=1}^k \sum_{j=1}^m \sum_{\nu=1}^n x_{ij\nu}^2. \quad S_{AB}^2 = \frac{Q_5 - nQ_1}{mk(n-1)},$$

взаємодію факторів перевіряємо за критерієм

$$\frac{nS_0^2}{S_{AB}^2} > F_\alpha(f_1, f_2) \quad , \text{ де } f_1 = (k-1)(m-1), \quad f_2 = mk(n-1).$$

Лабораторна робота 4. Регресійний аналіз. Виділення тренду.

Завдання: Результати спостереження за деякою функціональною залежністю задані таблицею (відповідно варіанту). Методом найменших квадратів знайти найкращу функціональну залежність: лінійна, поліноміальна(другого та третього порядків).

Провести статистичний аналіз отриманих коефіцієнтів для довірчої ймовірності $\alpha=0.95$.

Література: 5,6.

Завдання на СРС: розробити програмне забезпечення для оцінювання функціональної залежності заданої вибірки.

Методичні вказівки до виконання.

А) Задано: $n=1000$, $\delta=0.95$

$$x_i = i + (rnd(1) * N) / NG$$

$$y_i = N * rnd(1) * x_i + NG * rnd(1) + N, \quad i = 1, 2, \dots, n$$

N - номер студента в списку групи,

NG - номер групи.

Знайти оцінки параметрів лінійної регресії $y = \alpha + \beta x$ методом найменших квадратів.

Перевірити наявність викидів у регресії - якщо $R > R_\delta$, то y_i , що відповідає максимальному значенню відношення $\frac{e_i}{S_i}$ є викидом з

достовірністю δ , де $\hat{y}_i = a + bx_i$

$$e_i = y_i - \hat{y}_i,$$

$$R = \max \left| \frac{e_i}{S_i} \right|$$

$$S_i^2 = \frac{\sum_{i=1}^n e_i^2}{n-2} \left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right]$$

Значення $R_\delta \approx 4$.

Якщо є викиди, то їх видалити і провести оцінки спочатку.

Перевірити гіпотези про відповідність оцінок коефіцієнтів істинним значенням та адекватність моделі.

1. $H_0: \beta = b$. Значення коефіцієнта є значимим з достовірністю δ , якщо $|b| > \frac{t_{1+\delta}}{2} S_\beta$,

де $\frac{t_{1+\delta}}{2}$ – коефіцієнт розподілу Стюдента з $(n-2)$ степенями свободи,

$$S_\beta = \frac{S}{S_x \sqrt{n-1}}; \quad S^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - a - bx_i)^2;$$

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2; \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i;$$

2. $H_0: \alpha = a$. Значення коефіцієнта є значимим з достовірністю δ , якщо $|a| > \frac{t_{1+\delta}}{2} S_\alpha$,

де $\frac{t_{1+\delta}}{2}$ – коефіцієнт розподілу Стюдента з $(n-2)$ степенями свободи,

$$S_\alpha = S \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{(n-1) S_x^2}},$$

3. H_0 : модель адекватна. Модель адекватна з достовірністю δ , якщо $\frac{S^2}{S_y^2} < F_\delta$,

де F_δ – квантиль розподілу Фішера з $(n-2)$ та $(n-1)$ степенями свободи.

В) Методом найменших квадратів знайти найкращу функціональну залежність: лінійна, поліноміальна (другого та третього порядків). Для вибору найкращої скористатись значенням S^2 .

Результати оформити у вигляді звіту.

Лабораторна робота 5. Алгоритми перетворення Фур'є.

Завдання: Для заданої реалізації деякого сигналу (відповідно до варіанту) оцінити коваріаційну функцію, спектральну щільність. Використати вагові вікна Хеммінга і Блекмана. Порівняти результати.

Література: 4.

Завдання на СРС: розробити програмне забезпечення для реалізації швидкого перетворення Фур'є.

Методичні вказівки до виконання.

А) Задано: $n = NG/2$,

$$s2_i := 2 \cdot \text{md}(1) + NG \cdot \cos\left(2M\pi \cdot \frac{i}{N}\right) \cdot (1 + 0.1 \text{md}(1)) + 17 \cdot \cos\left(\frac{4M\pi i}{N} + \text{md}(1)\right) + 3 \cos\left(\frac{5M\pi i}{N}\right) \cdot (\text{md}(1) + NG)$$

$$N = 2^n,$$

M - номер студента в списку групи,

NG - номер групи.

Провести нормування, оцінити кореляційну функцію.

Виконати перетворення Фур'є за формулами

$$A_0 := \frac{1}{N} \cdot \sum_{i=0}^{N-1} \left(s1_i \cdot \cos\left(\frac{2\pi \cdot i \cdot 0}{N}\right) \right) \quad A_{\frac{N}{2}} := \frac{1}{N} \cdot \sum_{i=0}^{N-1} \left(s1_i \cdot \cos\left(\frac{\pi \cdot i}{1}\right) \right)$$

$$A_l := \frac{2}{N} \cdot \sum_{i=0}^{N-1} \left(s1_i \cdot \cos\left(\frac{2\pi \cdot i \cdot l}{N}\right) \right) \quad l := 1, 2, \dots, \frac{N}{2} - 1$$

$$B_j := \frac{2}{N} \cdot \sum_{i=0}^{N-1} \left(s1_i \cdot \sin\left(\frac{2\pi \cdot i \cdot j}{N}\right) \right) \quad j := 0, 1, \dots, \frac{N}{2}$$

$$C_j := \sqrt{(A_j)^2 + (B_j)^2} \quad j := 0, 1, \dots, \frac{N}{2}$$

Знайти спектр

Побудувати графік. Обчислити частоту першої синусоїди (або крок по частоті).

Розрахунки виконати для варіантів

$$s1_i := s2_i,$$

$$s1_i := s2_i \cdot va_i \quad va_i := 0.42 - 0.5 \cos\left(\frac{2\pi i}{N}\right) + 0.08 \cos\left(\frac{4\pi i}{N}\right)$$

$$s1_i := s2_i \cdot vb_i \quad vb_i := 0.54 - 0.46 \cdot \cos\left(\frac{2\pi i}{N}\right)$$

Порівняти отримані результати.

Виконати обернене перетворення Фур'є

$$d1_i := \sum_{j=0}^{\frac{N}{2}-1} \left(A_j \cdot \cos\left(\frac{2\pi j \cdot i}{N}\right) \right) + \sum_{j=0}^{\frac{N}{2}-1} \left(B_j \cdot \sin\left(\frac{2\pi j \cdot i}{N}\right) \right)$$

Порівняти початковий і результуючий масиви. Співпали?

*) Запрограмувати швидке перетворення Фур'є (або використати готову програму). Провести аналогічні розрахунки. Порівняти отримані результати.

Лабораторна робота 6. Алгоритми вейвлет – аналізу.

Завдання: Побудувати вейвлет - базис для заданого батьківського вейвлету.

Для заданого сигналу обчислити вейвлет – коефіцієнти. Відновити сигнал. Проаналізувати отримані результати.

Література: 3,7.

Завдання на СРС: розробити програмне забезпечення для розрахунку коефіцієнтів Хаара.

Методичні вказівки до виконання.

А) Для батьківських вейвлетів Хаара, Шенона, Гауссового побудувати вейвлет - базис. Представити базис графічно.

Б) Задано: $n = NG/2$,

$$s1_i := 2 \cdot md(1) + NG \cdot \cos\left(2M\pi \cdot \frac{i}{N}\right) \cdot (1 + 0.1 \cdot md(1)) + 17 \cdot \cos\left(\frac{4M\pi i}{N} + md(1)\right) + 3 \cos\left(\frac{7M\pi i}{N}\right) \cdot (md(1) + NG)$$

$$N = 2^n,$$

M - номер студента в списку групи,

NG - номер групи.

Виконати пряме і обернене вейвлет - перетворення для кожного базису.

$$M1 = N,$$

$$f2(j, k, x) := 2^{\frac{j}{2}} \cdot g1(2^j \cdot x - k)$$

, $g1(x)$ - батьківський вейвлет,

$$W(1, j) := \sum_{i=0}^{N-1} (s1_i \cdot f2(1, j, i))$$

$$d_i := \sum_{l=0}^M \sum_{j=0}^{M1} \left(w(l,j) \frac{f2(l,j,i)}{2^{2l}} \right)$$

Порівняти початковий і відновлений масиви.

Рекомендована література

Базова література

1. Пархоменко В.М. Методи вибірових обстежень / В.М. Пархоменко. – Київ. – 2001. – 148 с.
2. Джессен Р.Д. Методы статистических обследований / Р.Д. Джессен. - М.: Финансы и статистика. - 1985.
3. Добеши И. Десять лекций по вейвлетам / И. Добеши. – Москва-Ижевск: НИЦ «Регулярная и хаотическая динамика». – 2004. – 464 с.
4. Бендат Дж. Прикладной анализ случайных данных / Дж. Бендат, А. Пирсол. М.: Мир. - 1989. – 540 с.
5. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников / А.И. Кобзарь. – М.: ФИЗМАТЛИТ. - 2006. – 816 с.
6. Джонсон Н. Статистика и планирование эксперимента в технике и науке. Методы обработки данных / Н. Джонсон, Ф. Аннон. – М.: Мир. – 1980. - 610 с.
7. Малла С. Вейвлеты в обработке сигналов / С. Малла. – М.: Мир.- 2005. – 672 с.

Допоміжна література

8. Гнеденко Б.В. Курс теории вероятностей / Б.В. Гнеденко. - М.: Наука. - 1988.
9. Закс Л. Статистическое оценивание / Л. Закс. –М.: СТАТИСТИКА. – 1976. – 598 с.
10. Шеффе Г. Дисперсионный анализ / Г. Шеффе. – М.: Наука. – 1980. – 512 с.
11. Закс Ш. Теория статистических выводов / Ш. Закс. – М.: Мир. – 1975. – 776 с.
12. Кокрен У. Методы выборочного исследования / У. Кокрен. – М.: Финансы и статистика. – 1976.
13. Блаттер К. Вейвлет – анализ. Основы теории / К. Блаттер. – М.: Техносфера. – 2004. 276 с.
14. Бахтин В.И. Введение в прикладную статистику / В.И. Бахтин. – Минск: БГУ. – 2011. – 91 с.
15. Спиридонов А.А. Планирование эксперимента при исследовании технологических процессов / А.А. Спиридонов. – М.: Машиностроение. – 1981. – 184 с.

16. Столниц Э. Вейвлеты в компьютерной графике. Теория и приложения / Э. Столниц, Т. ДеРоуз, Д. Салезин. – Ижевск: НИЦ "Регулярная и хаотическая динамика". – 2002. – 272 с.
17. Чуи К. Введение в вэйвлеты / К. Чуи. – М.: Мир. – 2001. – 412 с.
18. Новиков Л.В. Основы вейвлет-анализа сигналов / Л.В. Новиков. – С-Пб.: ООО"МОДУС". – 1999. – 152 с.

Інформаційні ресурси

1. Web-портал ФПМ. Архів матеріалів. – Режим доступу : http://fpm.kpi.ua/archive/dir.do?sys_id=obj_2
2. Електронний кампус НТУУ «КПІ». Матеріали з дисципліни «Аналіз даних». – Режим доступу : <http://login.kpi.ua>

Таблица II.

$\nu_1 \backslash \nu_2$	1	2	3	4	5	6	7	8	9
1	161	200	216	225	230	234	237	239	241
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88

$\alpha = 0,05$

10	12	15	20	24	30	40	60	120	∞
242 19,4 8,79 5,96 4,74	244 19,4 8,74 5,91 4,68	246 19,4 8,70 5,86 4,62	248 19,4 8,66 5,80 4,56	249 19,5 8,64 5,77 4,53	250 19,5 8,62 5,75 4,50	251 19,5 8,59 5,72 4,46	252 19,5 8,57 5,69 4,43	253 19,5 8,55 5,66 4,40	254 19,5 8,53 5,63 4,36
4,06 3,64 3,35 3,14 2,98	4,00 3,57 3,28 3,07 2,91	3,94 3,51 3,22 3,01 2,85	3,87 3,44 3,15 2,94 2,77	3,84 3,41 3,12 2,90 2,74	3,81 3,38 3,08 2,86 2,70	3,77 3,34 3,04 2,83 2,66	3,74 3,30 3,01 2,79 2,62	3,70 3,27 2,97 2,75 2,58	3,67 3,23 2,93 2,71 2,54
2,85 2,75 2,67 2,60 2,54	2,79 2,69 2,60 2,53 2,48	2,72 2,62 2,53 2,46 2,40	2,65 2,54 2,46 2,39 2,33	2,61 2,51 2,42 2,35 2,29	2,57 2,47 2,38 2,31 2,25	2,53 2,43 2,34 2,27 2,20	2,49 2,38 2,30 2,22 2,16	2,45 2,34 2,25 2,18 2,11	2,40 2,30 2,21 2,13 2,07
2,49 2,45 2,41 2,38 2,35	2,42 2,38 2,34 2,31 2,28	2,35 2,31 2,27 2,23 2,20	2,28 2,23 2,19 2,16 2,12	2,24 2,19 2,15 2,11 2,08	2,19 2,15 2,11 2,07 2,04	2,15 2,10 2,06 2,03 1,99	2,11 2,06 2,02 1,98 1,95	2,06 2,01 1,97 1,93 1,90	2,01 1,96 1,92 1,88 1,84
2,32 2,30 2,27 2,25 2,24	2,25 2,23 2,20 2,18 2,16	2,18 2,15 2,13 2,11 2,09	2,10 2,07 2,05 2,03 2,01	2,05 2,03 2,01 1,98 1,96	2,01 1,98 1,96 1,94 1,92	1,96 1,94 1,91 1,89 1,87	1,92 1,89 1,86 1,84 1,82	1,87 1,84 1,81 1,79 1,77	1,81 1,78 1,76 1,73 1,71
2,22 2,20 2,19 2,18 2,16	2,15 2,13 2,12 2,10 2,09	2,07 2,06 2,04 2,03 2,01	1,99 1,97 1,96 1,94 1,93	1,95 1,93 1,91 1,90 1,89	1,90 1,88 1,87 1,85 1,84	1,85 1,84 1,82 1,81 1,79	1,80 1,79 1,77 1,75 1,74	1,75 1,73 1,71 1,70 1,68	1,69 1,67 1,65 1,64 1,62
2,08 1,99 1,91 1,83	2,00 1,92 1,83 1,75	1,92 1,84 1,75 1,67	1,84 1,75 1,66 1,57	1,79 1,70 1,61 1,52	1,74 1,65 1,55 1,46	1,69 1,59 1,50 1,39	1,64 1,53 1,43 1,32	1,58 1,47 1,35 1,22	1,51 1,39 1,25 1,00