

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**  
**“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ”**  
**ФІЗИКО-ТЕХНІЧНИЙ ІНСТИТУТ**

**Кафедра математичних методів захисту інформації**

**ЗВІТ**

**З ВИРОБНИЧОЇ ПРАКТИКИ**

Напрямок підготовки: 6.040301 «Прикладна математика»

Тема: «Розробка автоматизованого тестуючого комплексу,  
що враховує психологічні особливості студентів»

Виконав студент 4 курсу

групи ФІ-13

Кригін Валерій Михайлович

Науковий керівник:

Доктор фізико-математичних наук, професор

Дороговцев Андрій Анатолійович

---

*(підпис)*

## ЗМІСТ

1 Вступ . . . . .	3
1.1 Обґрунтування та актуальність роботи . . . . .	3
1.2 Мета та завдання . . . . .	3
2 Основна частина . . . . .	4
2.1 Теоретичні відомості . . . . .	4
2.1.1 Метод головних компонент . . . . .	4
2.1.2 Гістограма . . . . .	7
2.1.3 Критерій узгодженості Пірсона $\chi^2$ . . . . .	8
Перелік посилань . . . . .	11

## **1 ВСТУП**

### **1.1 Обґрунтування та актуальність роботи**

Існуючі на даний момент системи тестування недостатньо гнучкі: вони аналізують лише відповіді на запитання, відносячи їх до вірних або невірних, а на цій базі роблять кінцевий висновок щодо знань студента. Стрімкий розвиток комп'ютерної техніки й інформаційних технологій надає можливість визначати ритм складання тесту, а також індивідуальні особливості людини. Дані психологічних досліджень допоможуть правильно трактувати отримані значення, а добре вивчені та перевірені часом математичні методи надають великі можливості для систематизації та обробки результатів вимірювання.

### **1.2 Мета та завдання**

Завдання наступні:

- 1) Вивчити математичні методи та розділи психології, що дозволять розв'язати поставлену задачу, пояснити та обґрунтувати отримані результати
- 2) Ознайомитися з правилами побудови тестових завдань для найбільш ефективної та об'єктивної процедури оцінки знань студентів
- 3) Розробити програмний комплекс тестування й обробки результатів
- 4) Моделювання

За мету поставлено збільшення об'єктивності тестування, а також покращення якості навчання за допомогою порад студентам і викладачам практичних занять.

## 2 ОСНОВНА ЧАСТИНА

### 2.1 Теоретичні відомості

#### 2.1.1 Метод головних компонент

Метод головних компонент (Principal component analysis) — метод, що дозволяє зменшити розмірність досліджуваної вибірки з мінімальними втратами інформації.

Маємо  $m$  об'єктів, з яких треба зняти по  $n$  певних властивостей. На вході в нас є виборки  $\vec{X}_k$ , кожна з яких відповідає сукупності властивостей  $k$ -го об'єкту

$$\vec{X}_k = \begin{bmatrix} x_k^1 \\ x_k^2 \\ \vdots \\ x_k^n \end{bmatrix}, \quad k = \overline{1, m}$$

Згрупуємо всі вимірювання в одну матрицю  $X$

$$X = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_m^1 \\ x_1^2 & x_2^2 & \dots & x_m^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^n & x_2^n & \dots & x_m^n \end{bmatrix}$$

Спочатку нам знадобиться знайти вибірккові середні значення для кожної властивості

$$a_i = \frac{1}{m} \cdot \sum_{k=1}^m x_k^i, \quad i = \overline{1, n}$$

Маємо вектор вибірових середніх значень

$$\vec{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

Центруємо отримані дані, що містяться в матриці  $X$ , віднявши від кожного стовбця вектор вибірових середніх  $\vec{a}$

$$\tilde{X} = \begin{bmatrix} \tilde{x}_1^1 & \tilde{x}_2^1 & \dots & \tilde{x}_m^1 \\ \tilde{x}_1^2 & \tilde{x}_2^2 & \dots & \tilde{x}_m^2 \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{x}_1^n & \tilde{x}_2^n & \dots & \tilde{x}_m^n \end{bmatrix} = \begin{bmatrix} x_1^1 - a_1 & x_2^1 - a_1 & \dots & x_m^1 - a_1 \\ x_1^2 - a_2 & x_2^2 - a_2 & \dots & x_m^2 - a_2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^n - a_n & x_2^n - a_n & \dots & x_m^n - a_n \end{bmatrix}$$

Обчислюємо вибірову коваріаційну матрицю властивостей. Вибіркову коваріацію  $i$  та  $j$  властивості рахуємо за формулою

$$\sigma_i^j = \frac{1}{m} \cdot \sum_{k=1}^m \tilde{x}_k^i \cdot \tilde{x}_k^j = \frac{1}{m} \cdot \sum_{k=1}^m \left[ (x_k^i - a_i) \cdot (x_k^j - a_j) \right], \quad i, j = \overline{1, n}$$

Маємо вибірову коваріаційну матрицю

$$K = \begin{bmatrix} \sigma_1^1 & \sigma_1^2 & \dots & \sigma_1^n \\ \sigma_2^1 & \sigma_2^2 & \dots & \sigma_2^n \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_n^1 & \sigma_n^2 & \dots & \sigma_n^n \end{bmatrix}$$

Щоб отримувати лише потрібну інформацію, ми хочемо знайти таке ортогональне лінійне перетворення  $L$  вхідної матриці  $\tilde{X}$ , щоб отримати матрицю

$Y = L \cdot \tilde{X}$ , яка має діагональну вибірку коваріаційну матрицю  $K'$  з незростаючими зверху вниз значеннями. Діагональна вибірка коваріаційна матриця гарантує той факт, що отримані значення  $Y$  будуть некорельованими. Рангування значень діагональних елементів матриці  $K'$  за величиною дасть більш наглядне представлення про будову досліджуваних об'єктів, адже діагональні елементи — вибірккові дисперсії; а чим більше дисперсія, тим більше відповідна властивість змінюється від об'єкту до об'єкту і тим більше корисної інформації вона нам надає.

Вибіркова коваріаційна матриця  $K'$  для  $Y = L \cdot \tilde{X}$  має вигляд

$$K' = L \cdot K \cdot L^* = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

З лінійної алгебри відомо, що матриця  $L$  складається з координат власних векторів матриці  $K$ , а елементи  $\lambda_k$  — її власні числа, які існують і є невід'ємними через невід'ємну означеність матриці  $K$ . Вважаємо що числа  $\lambda_1, \dots, \lambda_n$  впорядковані від більшого до меншого для зручності подальших дій. Позначимо власний вектор матриці  $K$ , що відповідає власному числу  $\lambda_k$ , як  $\vec{l}_k$ . Тоді

$$\vec{l}_k = [l_k^1, l_k^2, \dots, l_k^n], \quad k = \overline{1, n}$$

Матриця  $L$  має вигляд

$$L = \begin{bmatrix} l_1^1 & l_1^2 & \dots & l_1^n \\ l_2^1 & l_2^2 & \dots & l_2^n \\ \vdots & \vdots & \ddots & \vdots \\ l_n^1 & l_n^2 & \dots & l_n^n \end{bmatrix}$$

Треба зменшити розмірність простору досліджуваних параметрів системи з  $n$  до  $p < n$ , але при цьому втратити якомога менше відомостей про досліджувані об'єкти. Введемо міру інформації, що залишається при зменшенні кількості компонент, що розглядаються

$$I = \frac{\lambda_1 + \dots + \lambda_p}{\lambda_1 + \dots + \lambda_n}$$

Будемо вважати, що діємо продуктивно, тому починаємо обирати з перших компонент, адже саме вони є найбільш інформативними. Також бачимо, що інформативність змінюється в межах від 0 (нічого не дізнаємось) до 1 (зберегли усю інформацію).

### 2.1.2 Гістограма

Для подальшого аналізу потрібно здобути щільність розподілу головних компонент. Оскільки маємо справу з вибіркою і вибірковими характеристиками, потрібно побудувати гістограму, адже це і є вибіркова характеристика, що відповідає щільності.

Побудуємо  $i$ -й стовбець гістограми для виборки з  $k$ -ї строки матриці  $Y$

$$h_i^k = \frac{1}{p} \cdot \sum_{j=1}^p \mathbb{1}(y_j^k \in I_i^k)$$

де  $I^k$  — набір напівінтервалів, що розбиває відрізок  $\left[ \min_{j=1,p} y_j^k; \max_{j=1,p} y_j^k \right]$  на  $N$  рівних частин. Для вибору  $N$  можна скористатися досить відомою формулою Стюрджеса (Sturges' formula) [1]

$$N = \lfloor \log_2 p \rfloor + 1$$

Маємо матрицю гістограм

$$H = \begin{bmatrix} h_1^1 & h_2^1 & \dots & h_N^1 \\ h_1^2 & h_2^2 & \dots & h_N^2 \\ \vdots & \vdots & \ddots & \vdots \\ h_1^n & h_2^n & \dots & h_N^n \end{bmatrix}$$

і напівінтервалів, що відповідають кожному стовбчику кожної гістограми

$$I = \begin{bmatrix} I_1^1 & I_2^1 & \dots & I_N^1 \\ I_1^2 & I_2^2 & \dots & I_N^2 \\ \vdots & \vdots & \ddots & \vdots \\ I_1^n & I_2^n & \dots & I_N^n \end{bmatrix}$$

### 2.1.3 Критерій узгодженості Пірсона $\chi^2$

Гістограма може використовуватися не тільки для графічної інтерпретації отриманих даних, але й для віднесення вибірки до якогось відомого розподілу. Відповідь на питання “Чи дійсно вибірка  $y_1^k, \dots, y_p^k$  має розподіл  $F^k$ ?” може надати критерій узгодженості Пірсона.

Розглянемо величину

$$S = p \cdot h_i^k = \sum_{j=1}^p \mathbb{1}(y_j^k \in I_i^k)$$

Вона є сумою  $p$  результатів бернулівських експериментів, ймовірність успіху якого заздалегіть невідома — саме ця характеристика й визначається припущенням щодо розподілу. Отже, треба визначити ймовірність  $\rho^k$  того, що випадкова



величина  $\xi^k$  з функцією розподілу  $F^k$  потрапить у напівінтервал  $I_i^k$

$$\rho^k = \mathbb{P}(\xi^k \in I_i^k), \quad \mathbb{P}(\xi^k \leq x) = F^k(x), \quad k = \overline{1, p}$$

Маємо наступні характеристики випадкової величини  $S$  [2]

$$M S = p \cdot \rho^k, \quad D S = p \cdot \rho^k$$

Застосуємо центральну граничну теорему

$$\frac{S - M S}{\sqrt{D S}} = \frac{p \cdot h_i^k - p \cdot \rho^k}{\sqrt{p \cdot \rho^k}} = \sqrt{p} \cdot \frac{h_i^k - \rho^k}{\sqrt{\rho^k}} \xrightarrow{p \rightarrow \infty} N(0, 1)$$

Розглянемо величину  $R^k$ , що рахується за формулою

$$R^k = p \cdot \sum_{i=1}^N \frac{(h_i^k - \rho^k)^2}{\rho^k}, \quad k = \overline{1, p}$$

Маємо суму  $p$  квадратів **залежних** між собою випадкових величин, що приблизно мають стандартний гаусовський розподіл. Хотілося б мати розподіл суми квадратів **незалежних** стандартних гаусовських випадкових величин — розподіл Пірсона  $\chi^2$ . Виявляється, що залежність отриманих випадкових величин забирає лише один ступінь вільності, і на виході отримуємо розподіл Пірсона з  $N - 1$  ступенями вільності [3]

$$R^k \xrightarrow{p \rightarrow \infty} \chi_{N-1}^2$$

Це не дивно: якщо відомо, що  $t$  значень виборки потрапило в  $N - 1$  напівінтервал, то ми знаємо, що в останній напівінтервал потрапило рівно  $p - t$  елементів. Лінійне обмеження, що губить один ступінь вільності, в даній задачі виглядає наступним чином

$$\sum_{i=1}^N h_i^k = 1$$

З таблиці для функції розподілу  $\chi_{N-1}^2$  обираємо рівень значущості  $\alpha$  і шукаємо відповідне  $r_\alpha$ . Рівень значущості — ймовірність помилки першого роду —

ймовірність того, що буде відкинуто вірну гіпотезу

$$\mathbb{P}(\chi_{N-1}^2 \leq r_\alpha) = \alpha$$

Якщо  $R^k \leq r_\alpha$ , то гіпотеза про те, що вибірка  $Y^k$  дійсно має розподіл  $F^k$ , не відхиляється; інакше значення  $R^k$  буде поводитись як  $\sqrt{p}$  — достатньо швидко зростати внаслідок центральної граничної теореми — гіпотезу буде відхилено.

Чим більше рівень значущості, тим менше значення  $r_\alpha$  — тим менше проміжок, в який дозволяється потрапити значенню  $R^k$  і більша ймовірність відхилити вірну гіпотезу щодо розподілу; але при цьому є більше впевненості в правильності результату.

## ПЕРЕЛІК ПОСИЛАНЬ

1. *Sturges, Herbert A.* The Choice of a Class Interval / Herbert A. Sturges // *J-AM-STAT-ASSOC.* — 1926. — March. — Vol. 21, no. 153. — Pp. 65–66.
2. *Weatherburn, C.E.* A First Course Mathematical Statistics / C.E. Weatherburn. — Cambridge University Press, 1963.
3. *Hudson, Derek J.* Lectures on elementary statistics and probability / Derek J Hudson. — Geneva: CERN, 1963.