

УДК 519.2: 613

ИСПОЛЬЗОВАНИЕ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ ДЛЯ КЛАССИФИКАЦИИ СОСТОЯНИЯ ПСИХИЧЕСКОГО ЗДОРОВЬЯ СТУДЕНТОВ НФИ КЕМГУ

С. П. Казаков, Г. К. Отрубейникова, Л. С. Данилова

В работе применяется метод главных компонент для оценки состояния психического здоровья студентов первого курса НФИ КемГУ, который дает возможность разделить их на три классификационных группы.

In work a component the method of the main things is applied to an estimation of a condition of mental health of first-year students Nfi KemSU which gives the chance to divide them into three classification groups.

Состояние психического здоровья студентов анонимно может быть оценено с помощью следующих показателей.

Таблица 1

**Таблица показателей состояния психического
здоровья студентов**

<i>Показатель</i>	<i>Градация</i>	<i>Оцифровка</i>
1. Адаптация к учебному процессу	• высокая • средняя • низкая	• 1 • 2 • 3
2. Нервно-психическое напряжение	• слабое • умеренное • чрезмерное	• 1 • 2 • 3
3. Тревожность личностная	• низкая • средняя • высокая	• 1 • 2 • 3
4. тревожность ситуативная	• низкая • средняя • высокая	• 1 • 2 • 3
5. Субдепрессия	• нормальное настроение • незначительно сниженное • значительно сниженное • субдепрессия	• 1 • 2 • 3 • 4
6. Качество жизни	• хорошее • соответствует желаемому • удовлетворительное • неудовлетворительное	• 1 • 2 • 3 • 4
7. Психологический климат в коллективе	• зрелый • достаточно зрелый • недостаточно зрелый • незрелый	• 1 • 2 • 3 • 4
8. Самооценка	• адекватная • завышенная • заниженная • завышенная по невротическому типу	• 1 • 2 • 3 • 4
9. Акцентуация	• есть • нет	• 1 • 0

10. Мотивация обучения в вузе	• получение знаний • овладение профессией • получение диплома	• 1 • 2 • 3
11. Стрессоустойчивость	• высокая • средняя • низкая • состояние стресса	• 1 • 2 • 3 • 4

С помощью соответствующих методик количественные факторы оцифрованы (последняя графа таблицы 1). После этого каждый объект выборки (студент) может рассматриваться в виде вектора:

$$x_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(p)}), \quad (1)$$

где i – номер студента в выборке,
 p – число показателей здоровья.

Для выявления неоднородности совокупности объектов x_i , т. е. принадлежности их к отдельным группам, например, группам больных и здоровых, часто используется метод главных компонент, который состоит в следующем. Главными компонентами $z_i^{(1)}, \dots, z_i^{(p)}$ многомерного признака $X = (x^{(1)}, \dots, x^{(p)})$ называют систему ортонормированных линейных комбинаций исходных признаков.

$$\left\{ \begin{aligned} z^{(i)}(X) &= u_{1j} \frac{(x^{(1)} - m^{(1)})}{\sigma_1} + \dots + u_{pj} \frac{(x^{(p)} - m^{(p)})}{\sigma_p}; \\ \sum_{i=1}^p u_{ij}^2 &= 1 (j = \overline{1, p}); \\ \sum_{i=1}^p u_{ij} u_{ik} &= 0 (j; k = \overline{1, p}, j \neq k), \end{aligned} \right. \quad (2)$$

где $m^{(i)}$ – математическое ожидание признака $x^{(i)}$;
 σ_j – корень из дисперсии $x^{(j)}$.

При этом линейные комбинации выбираются так, чтобы первая главная компонента обладала наибольшей дисперсией. Геометрически это означает, что первая главная компонента ориентирована вдоль направления наибольшей вытянутости эллипсоида рассеивания исследуемой совокупности. Вторая главная компонента имеет наибольшую дисперсию среди всех линейных преобразований вида (2), некоррелированных с первой и второй главными компонентами и т. д.

Вычисление коэффициентов главных компонент основано на том факте, что векторы $U_1, \dots, U_p$ являются собственными характеристическими векторами (2) матрицы ковариаций исследуемой совокупности, т. е. удовлетворяют системе уравнений

$$\sum U = \lambda U. \quad (3)$$

Соответствующие собственные числа равны дисперсиям главных компонент. Если собственные векторы упорядочить в порядке убывания собственных чисел $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$, то первой главной компоненте соответствует собственный вектор с максимальным собственным числом λ_1 и $D_z^{(1)} = \lambda_1$, второй главной компоненте соответствует собственный вектор со следующим по значению собственным числом λ_2 и $D_z^{(2)} = \lambda_2$ и т. д.

Всего существует p собственных векторов матрицы Σ и, следовательно, p главных компонент. Ковариационная матрица главных компонент $z^{(1)}, \dots, z^{(p)}$ является диагональной.

Отношение суммы первых q главных компонент к общей их сумме можно интерпретировать как долю информации о выборке, в них содержащейся. Чем ближе это значение к 1, тем больше информации они несут о совокупности экспериментальных точек.

Процесс классификации объектов по состоянию здоровья складывается из нескольких этапов:

1. Вычисляются собственные числа уравнения (3).
2. Вычисляются главные компоненты выборки.
3. На плоскость первых двух главных компонент, содержащих основную информацию о

выборке, наносятся точки, характеризующие объекты.

4. Осуществляется визуальная интерпретация классификации.

5. Построение гистограммы распределения значений параметров главных компонент.

6. Разделение гистограммы и плоскости главных компонент на три области. $\bar{X} \pm \sigma$ – параметры студентов со средним состоянием здоровья, $\bar{X} - \sigma$ и менее – параметры студентов с хорошим состоянием здоровья, $\bar{X} + \sigma$ и более – параметры студентов с неудовлетворительным состоянием здоровья.

Собственными числами системы (3) для базы данных первого курса является:

$$2,129 \geq 1,286 \geq 1,069 \geq 1,008 \geq 0,974 \geq 0,860 \geq 0,764 \geq 0,743 \geq 0,629 \geq 0,538;$$

Поскольку $\frac{\lambda_1 + \lambda_2}{\sum \lambda_i} = 0,73$, т. е. в первых двух

главных компонентах содержится около 73 % информации о выборке, элементы выборки можно располагать в плоскости двух главных компонент.

На рис. 1 представлены параметры выборки, расположенных в плоскости первых двух главных компонент.

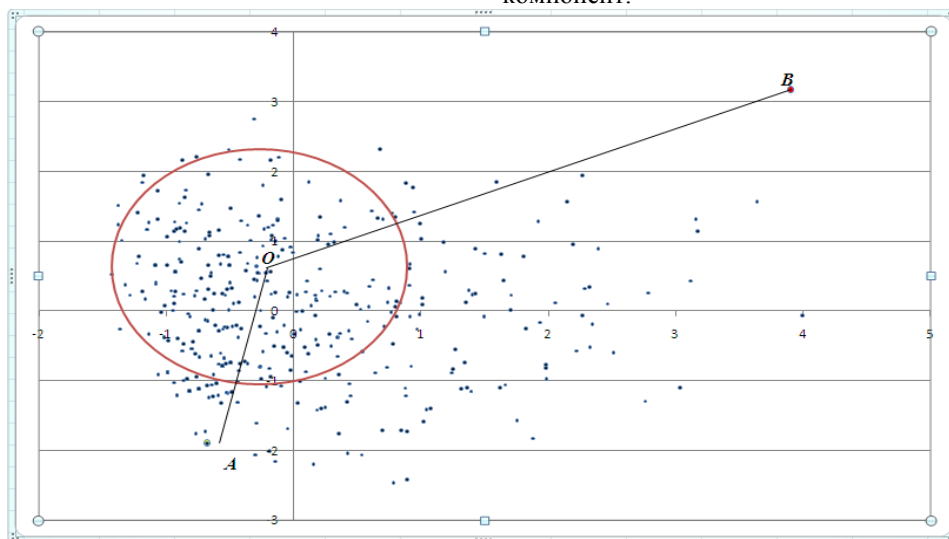


Рис. 1. Расположение выборки в плоскости первых двух главных компонент

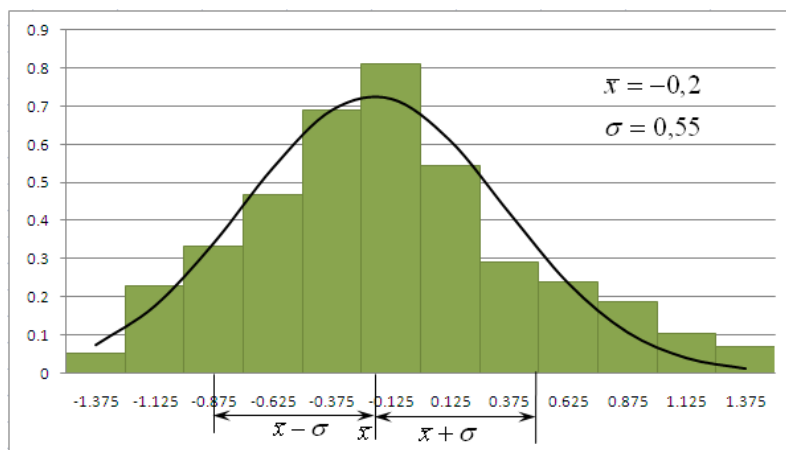


Рис. 2. Гистограмма распределения параметров здоровья студентов в выпрямленной системе координат

Направление прямой вниз от центра овала характеризует улучшение состояния здоровья, вправо – ухудшение состояния здоровья. Для построения областей значений параметров студентов с различным состоянием здоровья на рис. 2 построена гистограмма распределений параметров в выпрямленной системе координат АОВ.

Область параметров студентов со средним состоянием здоровья ($\bar{X} \pm \sigma$) находится внутри овала. Ниже параметры студентов с хорошим состоянием здоровья, справа – с плохим состоянием здоровья. Полученные результаты используются далее как обучающая выборка для анализа изменения состояния здоровья студентов во времени.

Литература

1. Айвазян, С. А. Прикладная статистика. Классификация и снижение размерности / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков и др. – М.: Финансы и статистика, 1989. – 607 с.
2. Айвазян, С. А. Вероятностно-статистическое моделирование распределительных отношений в обществе: ученые записки по статистике / С. А. Айвазян. – М.: Статистика, 1980. – Т. 37.
3. Айвазян, С. А. Классификация многомерных наблюдений / С. А. Айвазян, З. И. Бежаева, О. В. Староверов. – М.: Статистика, 1974.
4. Кендалл, М. Многомерный статистический анализ / М. Дж. Кендалл, А. Стюарт. – М.: Наука, 1976.
5. Харман, Г. Современный факторный анализ / Г. Харман. – М.: Статистика, 1972.

Рецензент – В. И. Байзакин, ГОУ ВПО «Сибирский государственный индустриальный университет».