УДК 378: 519.2

# ИССЛЕДОВАНИЕ ВЛИЯНИЯ ОСНОВНЫХ ФАКТОРОВ НА УСПЕВАЕМОСТЬ УЧЕБНЫХ ГРУПП ВУЗА

**Гранков М. В., доц., к.т.н., доц.; Аль-Габри В. М., аспирант**
*(Донской государственный технический университет, г. Ростов-на-Дону, РФ)*

**Введение.** Одной из проблем образования в Российской Федерации является значительное снижение уровня успеваемости студенческих групп, рассчитанного на конец сессии. В данном случае речь идет о проценте положительных оценок, полученных студентами некоторого множества учебных групп по результатам экзаменационной сессии. Снижение этого процента повышает общее количество задолженностей по вузу, что приводит к снижению уровня освоения студентами компетенций, предусмотренных образовательными программами. Проблеме повышения и анализу успеваемости студенческих групп посвящены научно-исследовательская работа [2]. В настоящем исследовании делается попытка анализа и обоснования основных факторов, влияющих на уровень успеваемости студенческих групп по результатам экзаменационных сессий.

**Цель.** Разработать метод, позволяющий проанализировать и формально обосновать степень влияния основных факторов, участвующих в формировании результатов экзаменов, на общую успеваемость студентов вуза.

**Исходные данные.** В исследовании использовалась база данных электронных деканатов Донского Государственного Технического Университета.

**Модель успеваемости учебных групп по результатам сессий.** На результаты экзаменов оказывают влияние большое число объектов, процессов, факторов и условий различного происхождения и природы. Рассматривая учебную группу как некий интегратор этих процессов, можно предположить, что такие группы, как пример социальных групп, могут быть классифицированы и иметь, в зависимости от внешних воздействий, достаточно предсказуемое поведение. Опираясь на это предположение, выделим основные типы объектов, влияющих на успеваемость: учебная группа. экзаменатор; дисциплина; расписание экзаменов (сессии).

Формально расписание экзаменов это отношение, построенное на прямом произведении множеств объектов: учебная группа, экзаменатор, дисциплина. Представим учебную группу некоторым объектом, преобразующим внешние воздействия экзаменатора, дисциплины и расписания сессии в значение успеваемости. Введем следующие обозначения:

$N = \{1, 2, ..., i, ...\}$ — множество натуральных чисел; (1)

$N_0 = \{0\} \bigcup N$ — расширенное множество натуральных чисел; (2)

$I = \{0,1\}$ — единичный отрезок действительных чисел; (3)

$$E_n = \left\{ r_i \middle| (i = 1, 2, ...n) \wedge (r_i = \frac{n1_i}{n2_i}) \wedge (n1_i \in N_0) \wedge (n2_i \in N) \wedge (n1_i \le n2_i) \wedge (n \in N) \right\}$$ —

множество неотрицательных рациональных чисел, представленных дробями, значение которых не превышает 1; (4)

$G = \{1, 2, ...n_g\}, n_g \in N$ — множество учебных групп, где $n_g$ – число учебных групп; (5)

$T = \{1, 2, ...n_t\}, n_t \in N$ — множество экзаменаторов (преподавателей), где $n_t$ – число экзаменаторов; (6)

$S = \{1, 2, ...n_s\}, n_s \in N$ — множество дисциплин (предметов), где $n_s$ – число дисциплин; (7)

$$TT \subset G \times N \times N \times N \times T \times S \times N_0 \times N \times U_n \quad (8)$$

где отношение $TT$ содержит результаты экзаменов групп вуза на конец соответствующих сессий. Кортеж $(g, y, i, j, t, s, n1, n2, u) \in TT$, если группа $g \in G$, год поступления в вуз которой равен $y \in N$, сдавала в сессию под номером $i \in N$ экзамен под номером $j \in N$ преподавателю $t \in T$ по дисциплине $s \in S$. Значение $n2$ соответствует общему числу студентов на экзамене, а $n1$ соответствует числу студентов из $n2$, получивших положительные оценки. Справочное значение $u\left(u \in E_n\right)$, является успеваемостью группы по результатам данного экзамена.

Пусть $TT_1$ – некоторое подмножество отношения $TT\left(TT_1 \subset TT\right)$, которое соответствует некоторой выборке из последнего отношения. В зависимости от запросов в подмножество $TT_1$ могут включаться результаты экзаменов одной или нескольких групп. Рассмотрим проекцию множества $TT_1$ по атрибутам 7 и 8 (число положительных оценок и общее число студентов, соответственно).

$$\pi_{7,8}\left(TT_1\right) = \left\{(n1_1, n2_1), (n1_2, n2_2), \ldots, (n1_k, n2_k)\right\}, \quad (9)$$

где $k = \left|TT_1\right|$ – общее число результатов экзаменов, включенных в выборку.

В этом случае среднюю успеваемость на экзаменах групп, включенных в выборку, можно найти по формуле:

$$u_1 = \left(\sum_{i=1}^{n} n1_i \middle/ \sum_{i=1}^{n} n2_i\right), u_1 \in E_n. \quad (10)$$

**Задача и методы кластеризации объектов.** Одним из стандартных аппаратов решения проблем, подобных сформулированной в цели настоящего исследования, является дисперсионный анализ. Основная и популярная у исследователей часть этого аппарата базируется на гипотезах о нормальности распределений, однородности математических ожиданий и дисперсий, что очень проблематично в исследуемой предметной области. Как известно, множество полученных отдельным студентом оценок подчиняется непараметрической статистике. В работе [2] рассмотрены проблемы, которые возникают при переходе от непараметрической статистики успеваемости каждого студента к параметрической статистике успеваемости в больших выборках студентов. В этих исследованиях под успеваемостью понималось среднее значение оценок (баллов), которые студент набрал за некоторый интервал времени обучения. Было обнаружено, что на границах допустимых значений оценок возникают функциональные зависимости между средней успеваемостью в локальном интервале и средней дисперсией, рассчитанной в том же интервале. При этом отмечалось влияние на данную зависимость процедуры пересдач неудовлетворительных оценок. В результате, средние успеваемости, полученные для больших выборок студентов, могут не подчиняться нормальному закону распределения случайной величины, даже если для данного распределения выполнился критерий Хи-квадрат.

В настоящем исследовании, в отличие от работы [2]:

1. понятие успеваемости рассматривается только как доля студентов, получивших положительные оценки на конец сессии, к общему числу студентов, которые должны были сдавать экзамены;

2. значение показателя успеваемости регистрируется на конец сессии, без учета пересдач;

3. основным объектом исследования является не студент, а учебная группа. При этом, результатом измерения является успеваемость учебной группы, показанная по дисциплине на экзамене, проведенным в сессию (в соответствии с расписанием) некоторым преподавателем.

В общем случае на результат измерения успеваемости оказывают влияние: учебная группа, дисциплина, преподаватель, а, возможно, и расписание экзаменов [1]. Такое количество факторов значительно усложняет задачу анализа успеваемости. Воспользуемся аппаратом дисперсионного анализа средних, корректность которого доказана, в том числе для случая неравенства дисперсий и неравных чисел измерений в ячейках. Нами предложено получать значения факторов, которые необходимы для реализации дисперсионного анализа средних, с помощью методов кластеризации. Для этого, из выделенных нами ранее основных типов объектов образуем три типа качественных факторов (группа, преподаватель, предмет), влияющих на число задолженностей в вузе по результатам сессии. Соответствующие факторам объекты могут распадаться на подмножества, влияющие близким образом на результаты экзаменов. Для оценки влияния основных факторов на среднюю успеваемость с помощью дисперсионного анализа формализуем механизм деления факторов (объектов) на классы, используя методы нечеткой кластеризации. Для этого, каждому объекту из типов: группа, преподаватель, предмет, поставим в соответствие количественное значение фактора. В качестве такого значения используем значение средней успеваемости. В нашем случае измерение значения средней успеваемости реализуется проведением экзаменов в соответствии с расписанием сессии. Так как число студентов в учебных группах могут различаться, то в общем случае мы имеем дело с неравноточными измерениями. Тогда, под средней успеваемостью следует понимать средневзвешенную успеваемость. Формула (10) является рациональным вариантом расчета средневзвешенной успеваемости. Очевидно, что полученные средние значения для соответствующих факторов (объектов) могут проявляться с той, или иной степенью стабильности. Для оценки стабильности средней успеваемости будем использовать средневзвешенную выборочную дисперсию. Следовательно, каждый рассматриваемый объект можно описать двумерным вектором **x:**

$$\boldsymbol{x} = \left(u, D_u\right), \tag{11}$$

где $u$ – выборочная средняя успеваемость; $D_u$ – выборочная средняя дисперсия успеваемости.

Каждый, рассматриваемый нами объект (фактор) представим вектором (11) в двумерном пространстве свойств объектов. Наиболее популярным способом измерения подобия объектов считается метрика Евклида. Одним из распространенных алгоритмов нечеткой кластеризации является алгоритм FCM (*Fuzzy C-Means*), который позволяет приписать одни и те же объекты к разным классам с соответствующими степенями принадлежности.

**Результаты кластеризации.** Разделим учебные группы на пять классов по параметру «средняя успеваемость». Первоначальные эксперименты с одномерной декомпозицией учебных групп на классы по первому параметру вектора $\boldsymbol{x}$ (11) показали, что добиться однородности по оценкам успеваемости и ее дисперсии для всех групп, входящих в один класс, не представляется возможным. Удовлетворительный результат был получен после применения двухэтапной кластеризации. На первом этапе – одномерная кластеризация на пять классов по средней успеваемости, а на втором этапе – двумерная кластеризация на два – пять подклассов по параметрам вектора **x** (11), проведенная для множеств групп каждого класса, полученных на первом этапе. Объекты типов «Преподаватели» и «Предметы» были разбиты на пять классов каждый. Проверка однородности по оценке дисперсии в полученных классах не производилась. Классы всех типов объектов упорядочены по убыванию успеваемости.

**Анализ влияния факторов на успеваемость учебных групп.** Приведем модифицированный алгоритм двухфакторного дисперсионного анализа средних:

1. разобьем каждое из множеств преподавателей и предметов на пять классов;
2. заполним таблицу средних следующим образом: число $u_{ij}$, стоящее на пересечении i-й строки и j-го столбца таблицы– это средняя взвешенная успеваемость, показанная в учебных группах по результатам экзаменов по предмету класса i, и проводимых

преподавателями класса j. Удалим из таблицы пятую строку и пятый столбец, т.к. в них содержались нулевые ячейки, получая таблицу 1.

3. Обработаем данные из таблицы 1, получая общую дисперсию $D_{общ}$ и групповые дисперсии $D_1$ и $D_2$ по факторам «Преподаватели» «Предметы».

4. Высказываем две нулевых гипотезы $H_{01}$ и $H_{02}$ о несущественности отличий групповых математических ожиданий успеваемостей соответственно по первому фактору («Преподаватели») и по второму («Предметы»). При проверке гипотез используются критерии (12) для гипотезы $H_{01}$ и (13) гипотезы $H_{02}$.

5. Высказываем нулевую гипотезу $H_{03}$ о несущественности различий в степени влияния значений первого и второго фактора на результаты измерений. При проверке гипотезы используется критерий (14).

*Таблица 1 – Таблица средних значений успеваемостей в зависимости от уровней двух факторов: «Преподаватели» и «Предметы»*

| Класс предмета | Класс преподавателя | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | $\overline{u1_i}^*$ |
| 1 | 0.889756 | 0.8183 | 0.7883 | 0.62727 | 0.7809065 |
| 2 | 0.834071 | 0.679 | 0.5564 | 0.44228 | 0.627938 |
| 3 | 0.791342 | 0.616 | 0.4742 | 0.34155 | 0.555773 |
| 4 | 0.719231 | 0.3986 | 0.3504 | 0.23529 | 0.42588 |
| $\overline{u2_j}^*$ | 0.8086 | 0.627975 | 0.542325 | 0.4115975 | **0.59762** |

$f_1 = \frac{D_1^*}{D_{ост}}$ — значение критерия для нулевой гипотезы $H_{01}$; (12)

$f_2 = \frac{D_2^*}{D_{ост}}$ — значение критерия для нулевой гипотезы $H_{02}$; (13)

$f_3 = \frac{\max(D_1^*, D_2^*)}{\min(D_1^*, D_2^*)}$ — значение критерия для нулевой гипотезы $H_{03}$. (14)

Результаты выполненных расчетов, сведены в таблицу 2.

*Таблица 2 – Результаты двухфакторного дисперсионного анализа успеваемости от классов объектов «Преподаватели» и «Предметы»*

| Название | $D_1^*$ | $D_2^*$ | $D_{ост}$ | $f_1$ | $F_1(\infty,3)$ | $f_2$ | $F_2(\infty,3)$ | $f_3$ | $F_3(3,3)$ |
|---|---|---|---|---|---|---|---|---|---|
| Значение | 0.332 | 0.263 | 0.03 | 11.1 | 2.6 | 8.87 | 2.6 | 1.26 | 9.28 |

В таблице 2 значения $F_1(\infty,3)$, $F_2(\infty,3)$ и $F_3(3,3)$ – критические значения $f_{0.95}$ для распределения Фишера. Значение первого аргумента $F_1(\infty,3)$и $F_2(\infty,3)$ соответствует значению степени свободы, связанное с числом измерений успеваемости каждой группы на каждом экзамене. Число таких измерений в проведенном исследовании значительно превышает максимальное число степеней свободы, предусмотренной в таблицах распределения Фишера. В связи с этим, значение первого параметра для таблиц распределения Фишера было принято равным бесконечности.

Из анализа Таблицы 2 следует:

1. Значение критерий $f_1$ и $f_2$ значительно превышает соответствующее ему критическое значение $F_1(\infty,3)$ и $F_2(\infty,3)$ из таблиц распределения Фишера. Следовательно, нулевые гипотезы $H_{01}$ и $H_{02}$ не могут быть приняты.

2. Значение критерия $f_3$ значительно меньше соответствующего ему критическое значение $F_3\left(3,3\right)$ из таблиц распределения Фишера. Следовательно, нулевая гипотеза $H_{03}$ может быть принята.

Для проверки гипотезы о влиянии расписания на общее число положительных оценок, полученных группой по результатам экзаменов сессии [1], был реализован однофакторный дисперсионный анализ. В результате получаем задачу с четырьмя уровнями фактора (4 экзамена) «Расписание». Применим в этой задаче дисперсионный анализ средних значений с различным числом измерений в ячейках. В настоящей работе исследовалось влияния расписания экзаменов на успеваемость групп второго класса, которые сдавали предметы второго класса преподавателям, отнесенных при кластеризации ко второму классу. Опуская промежуточные расчеты, приведем в Таблице 3 окончательные результаты.

*Таблица 3 – Результаты анализа влияния расписания на успеваемость*

| Название | $D^*_{мг}$ | $D^*_{вг}$ | $f$ | $F\left(N-1,k-1\right)$ |
|---|---|---|---|---|
| Значение | 0.0893 | 0.025 | 3.59 | 2.6 |

где $N$=1359 - общее число измерений; $K=4$ – число уровней факторов; $D^*_{мг}$- оценка межгрупповой дисперсии; $D^*_{вг}$ - оценка внутригрупповой дисперсии; $f$ - значение критерия; $F$ –критическое значение из распределения Фишера.

Нулевая гипотеза в отношении влияния расписания экзаменов на успеваемость групп не подтверждается ( $f > F\left(N-1,k-1\right)$). Следовательно, имеет место влияния номера экзамена на успеваемость групп по результатам этого экзамена.

**Выводы.**

1. Использование оригинального программного стенда позволило перейти от успеваемости отдельных студентов к успеваемости учебных групп и создать условия формализации анализа влияния факторов на общее число положительных оценок, полученных по результатам сессий.
2. Применение двухуровневой декомпозиции по средней успеваемости и оценке ее дисперсии позволило предложить технологию разделения множества объектов на подмножества статистически однородных объектов. Это позволит применять параметрические статистики внутри полученных однородных подмножеств.
3. Предложено в качестве уровня фактора использовать номер класса объектов, полученный в результате кластеризации. Такой подход позволил для достижения цели настоящего исследования корректно применять дисперсионный анализ средних успеваемостей групп.
4. Проведенный двухфакторный дисперсионный анализ доказал зависимость результатов экзаменов от классов предмета и преподавателя. Следовательно, доказано, что результаты экзамена зависят от средней успеваемости по предмету и преподавателю.
5. Проведенный однофакторный дисперсионный анализ показал наличие примеров влияния расписания на результаты экзаменов.

Перечень ссылок

1. Гранков, М. В. Постановка задачи автоматизации построения расписания экзаменов студентов вуза. Системный анализ, управление и обработка информации / М. В. Гранков, В. М. Аль-Габри // Труды 5-го Международного семинара ; под общ. ред. Р. А. Нейдорфа. – Ростов-на-Дону: ДГТУ, 2014. — 472 с.
2. Сосницкий, В. Н. Проблемы статистического анализа средней успеваемости студентов / В. Н. Сосницкий, Н. И. Потанин, Л. В. Шевелева // Журнал «Фундаментальные исследования». - №10(часть 2.). – Москва: ИД «Академия естествознания», 2013. - С.316 – 320.