
Э.Д. Товар Гарсиа
(Барселона, Испания)

ОБРАЗОВАТЕЛЬНЫЕ ТРАЕКТОРИИ ШКОЛЬНИКОВ ЯРОСЛАВСКОЙ ОБЛАСТИ: ДИСКРИМИНАНТНЫЙ АНАЛИЗ В СРАВНЕНИИ С ЛОГИСТИЧЕСКОЙ РЕГРЕССИЕЙ¹

В статье приведены результаты исследования образовательных траекторий школьников Ярославской области. Основные выводы заключаются в том, что успеваемость школьника, образовательные планы его друзей и уровень образования его отца служат ключевыми предикторами решения продолжить образование. Благодаря этой информации можно узнать, какие школьники рискуют не продолжить обучение. Далее сравниваются возможности логистической регрессии и дискриминантного анализа для бинарных зависимых переменных. При наличии необходимых предпосылок для использования методов обе стратегии позволяют классифицировать школьников с приемлемой точностью.

Ключевые слова: образовательная траектория, успеваемость, семейные ресурсы, человеческий капитал, логистическая регрессия, дискриминантный анализ, Ярославская область.

¹ Эдгар Деметрио Товар Гарсиа – Ph.D. по экономике Университета Барселоны, научный сотрудник Центра мониторинга качества образования Института образования НИУ ВШЭ. E-mail: beno09@yahoo.com.

Статья написана в рамках исследовательского проекта «Мониторинг образовательных и трудовых траекторий выпускников школ и вузов», разработанный НИУ ВШЭ. Автор выражает особую благодарность Е.В. Гусевой, Д.М. Янбарисовой, Д.С. Попову и анонимным рецензентам за комментарии к статье.

Введение

С одной стороны, для каждого человека образовательная траектория представляет собой главную детерминанту социальной мобильности и качества жизни. С другой стороны, в соответствии с теорией человеческого капитала Г.С. Беккера, для общества в целом высокий уровень образования населения служит определяющим фактором процесса экономического роста и развития. В России социально-экономический статус оказывает важное влияние на образовательные траектории, вследствие чего нередко высказывается рекомендация сконцентрироваться на разработке программ, ослабляющих негативный эффект от этого фактора [1]. Однако общее направление образовательной политики для страны в целом может оказаться неадекватным для некоторых ее регионов.

В связи со сказанным выше возникают два вопроса: какие факторы предсказывают образовательные траектории школьников в Ярославской области? Под траекторией в данном случае мы понимаем дихотомическое решение – продолжать обучение или нет. Отсюда второй вопрос: какая методология лучше подходит для определения этих предикторов?

В предыдущих эмпирических исследованиях для дихотомических зависимых переменных чаще всего применялись методы, основанные на оценке вероятности (логистическая и пробит регрессии). Но в этом исследовании предлагается также обратиться к методу дискриминантного анализа, который последние 20 лет сравнительно редко использовался для изучения образовательных траекторий.

Статья включает пять разделов. Первый – это короткое введение, с которым читатель уже знаком. Во втором разделе представлены теоретическая база и результаты предыдущих эмпирических исследований. Третий раздел – это обзор основных методов, используемых для предсказания образовательных траекторий. В четвертом разделе описываются использованные нами данные по

Ярославской области; в пятом – демонстрируются результаты применения логистической регрессии и дискриминантного анализа. В заключение представлены основные выводы, рекомендации и предложения, касающиеся возможных направлений дальнейших исследований.

Теоретическая база и результаты предыдущих эмпирических исследований

В литературе можно найти большой список показателей, которые используются для объяснения и предсказания успеваемости и образовательных траекторий школьников и студентов. Обобщая, можно выделить шесть основных подходов: с точки зрения теории человеческого капитала, культурного капитала, социального капитала, подход социально-экономического статуса, образовательный и психологический подходы [2].

Согласно теории *человеческого капитала*, разработанной Беккером, Шульцем и Минсером, образование, которое люди хотят получить сегодня, зависит от дохода, который они смогут получить на рынке труда в будущем. Иными словами, высокий уровень образования (знания и навыки) повышает ценность индивидуального человеческого капитала, который потом будет использован на рынке труда. Поэтому родители должны инвестировать в человеческий капитал детей в расчете на будущие доходы, и решение о начале или окончании образовательной траектории и полученные знания и навыки зависят от того, что сам человек планирует получить от работы в будущем.

Инвестиции в социальный и культурный капиталы, в соответствии с работами Бурдые и Коулмана, также повышают образовательные возможности детей. Разумеется, дети из семей с высоким уровнем культурного капитала и хорошей сетью социальных связей имеют больше возможностей при выборе образовательной траектории. Так, П. Димаджио, основываясь на

работах П. Бурдье, утверждал, что успеваемость сильно зависит от *культурного капитала* [3, р. 190]. Культурный капитал – «инструмент для присвоения символического социального богатства, которое достойно, чтобы его искали и им обладали», а также это инструмент для поддержания статуса на протяжении нескольких поколений (подход Вебера). Согласно теории культурного капитала, преподаватели лучше относятся к студентам или школьникам с высоким культурным капиталом, и благодаря этому ученики могут получать преимущества: они чувствуют себя комфортнее в школе, и их образовательные траектории лучше, чем у учеников с низким культурным капиталом.

Позже Д.С. Коулман отмечал, что *социальный капитал* продуктивен, как и другие виды капитала, и способствует достижениям в образовательных сферах. Социальный капитал представляет собой совокупность разных сущностей с двумя общими элементами: все они состоят из каких-нибудь аспектов социальной структуры и облегчают определенные виды деятельности участников внутри структуры [4, р. 98]. Д.С.Коулман утверждает, что социальный капитал в семье (отношения между родителями и детьми) представляет собой ключевое средство для приобретения человеческого капитала от родителей и влияет на образовательные траектории [там же, р. 110].

Другие исследования показывают, что *образовательные ресурсы*, т.е., например, соотношение школьников и учителей, наличие компьютеров, оснащение школы, оказывают влияние на образовательные траектории, имеет значение и качество этих ресурсов, скажем, профессиональный уровень учителей, современность компьютеров и их программного обеспечения и эффективность школьного руководства [5]. Поэтому от качества школы во многом зависит и качество образования.

Психологический подход подчеркивает, что образовательные траектории зависят от факторов, влияющих на поведение людей, например от мотивации к получению определенного уровня обра-

зования. Если интеллект или когнитивные навыки (часто измеряемые с помощью экзаменов) не прогнозируют качество образования и его траектории достаточно верно, тогда другие индивидуальные характеристики должны объяснять результаты, полученные в ходе образовательного процесса. Из этого следует, что некогнитивные черты и поведение важнее, чем когнитивные навыки [6]. Для достижения успеха в школе нам нужны правильные привычки, черты и манеры поведения: общительность, добросовестность, лидерские качества, инициативность и др.

Подход с точки зрения *социально-экономического статуса* (СЭС) предполагает, что доход, образование и профессиональный статус родителей служат лучшими объясняющими переменными по сравнению с предыдущими предикторами образовательных траекторий и успеваемости. С.Р. Лукас описал три теории, объясняющие роль СЭС [7]. Согласно первой – «перспектива жизненного пути» – СЭС теряет свою значимость, так как в процессе взросления дети получают независимость от родителей. В соответствии со второй теорией – «максимально поддерживаемое неравенство» – молодые люди получают независимость от СЭС благодаря социальной политике и государственным инвестициям в образование, но переход к следующему уровню образования зависит еще и от социального происхождения, поскольку родители с высоким СЭС успешнее используют государственную поддержку [8]. Третья – «эффективно поддерживаемое неравенство» – утверждает, что, когда определенный уровень образования не является бесплатным и обязательным, социально и экономически обеспеченные люди используют свои ресурсы, чтобы гарантировать своим детям хороший уровень образования. Когда образование бесплатно и обязательно, тем не менее обеспеченные люди могут получить лучшее образование, разница будет не в количестве, но в качестве.

Важно отметить, что в развитых странах показатели СЭС (образование, доход и профессиональный статус родителей) часто не оказывают сильного влияния на успеваемость и траектории,

в особенности, если речь идет о низких уровнях образования. Например, в Австралии роль СЭС уменьшается [9]. В Германии профессиональный статус родителей не влияет на успеваемость детей [10]. Однако в развивающихся странах СЭС до сих пор остается важной объясняющей переменной [11; 12]. Так, в Индии обучение детей в школе напрямую коррелирует с уровнем образования и богатством родителей [13]. В Гвинее повышение уровня образования отцов повышает уровень образование детей [14]. В Эфиопии образование родителей напрямую связано с поступлением детей в школу [15].

Результаты эмпирических исследований в России

Результаты предыдущих эмпирических исследований по России показывают, что социально-экономический статус, особенно образование родителей, напрямую связаны с образовательными результатами [1; 16–22]. Я. Рощина на основе данных Мониторинга экономики образования изучила планы 1098 учащихся 8 и 9-х классов и 2098 учащихся 10 и 11-х классов (исследование проводилось с 2006 по 2008 гг.), где 76,3% опрошенных были из разных регионов России и 23,7% – из Москвы. По результатам исследования 76,6% школьников в 8 и 9-м классе планировали продолжать обучение в 10 и 11-м классе, 16,1 – не знали, что будут делать, и 7,3% не собирались в 10 и 11-й класс. Кроме того, 67% школьников 8 и 9-х классов и 87% школьников 10 и 11-ых классов планировали получить высшее образование [22, р. 22 – 25].

Для объяснения планов школьников о продолжении обучения в вузе Я.Рощина использовала пробит анализ [22, р. 45, *табл. А2.3*]. Теоретически самые главные объясняющие переменные – это семейный капитал, который включает показатели СЭС, социальный и культурный капитал (образование, профессиональный статус и доход родителей, домашняя библиотека, наличие компьютера, количество членов семьи), и личные способности (успеваемость

и дополнительное обучение). Но в модель также были включены пол, тип школы и тип населенного пункта.

По результатам анализа оказалось, что в большей степени с образовательными планами связана успеваемость школьника, однако тип школы, дополнительное обучение, семейный доход, наличие компьютера и домашняя библиотека также имеют положительную, статистически значимую связь. Такие переменные, как образование, профессиональный статус родителей и количество членов семьи оказались статистически не значимыми. Интересно отметить, что Я. Рощина включает в модели только дихотомические объясняющие переменные. Кроме того, полученные коэффициенты детерминации достаточно низки (самый большой псевдо $R^2 = 0,30$ – в модели для школьников 10 и 11-х классов). Несмотря на эти результаты, Рощина приходит к выводу, что получение высшего образования главным образом зависит от социального статуса родителей, и особенно от уровня их образования (как показано и в других эконометрических моделях).

Дискриминантный анализ в сравнении с логистической регрессией

Последние 20 лет для анализа образовательных траекторий школьников и студентов (в том числе для их классификации по образовательным результатам) исследователи в основном использовали методы, основанные на оценке вероятности: логистическую, пробит, мультиномиальную и порядковую регрессии (когда зависимая переменная является категориальной). Развитие компьютерных технологий облегчило использование этих моделей, и их главный конкурент – дискриминантный анализ – в современных работах отошел на второй план, хотя он имеет свои преимущества.

Логистическая регрессия (ЛР)

При наличии дихотомической зависимой переменной ($Y = 0$ или 1) и множества независимых переменных (X) логистическая регрессия (ЛР) является вероятностной нелинейной функцией (1), поэтому ее прирост из-за увеличения независимой переменной на единицу не одинаков для всех значений X .

$$\Pr(Y = 0|X) = \frac{1}{1 + e^{(\beta'X)}}, \quad (1)$$
$$\Pr(Y = 1|X) = \frac{e^{(\beta'X)}}{1 + e^{(\beta'X)}}$$

В отличие от линейной регрессии, предсказанное значение зависимой переменной (Y) в этой модели всегда будет принадлежать интервалу от 0 до 1. Но эту логистическую функцию можно сформулировать как линейную функцию с помощью логарифмического преобразования (2). Предполагается, что зависимую переменную можно рассматривать в терминах вероятности (p), лежащей между 0 и 1, и преобразовать эту вероятность в логарифмическую: $p = \log_e p / (1 - p)$.

$$\log_e \frac{\Pr(Y = 1|X)}{\Pr(Y = 0|X)} = \beta'X = \beta_0 + \beta_1 X_1 + \beta_2 X_2 \dots + \beta_n X_n \quad (2)$$

Для оценки параметров β в уравнении (1) используется метод максимального правдоподобия. Такая модель позволяет распределить индивидов по двум группам в зависимости от значения вероятности (чаще всего за пороговую точку принимают значение $Y = p = 0,5$). Иными словами, в первую группу попадают индивиды с низкой вероятностью принадлежности к $Y \leq 0,5$, во вторую – с высокой вероятностью ($Y > 0,5$). Это обуславливает предсказательные возможности ЛР. Важно заметить, что ЛР является одной из моделей регрессионного анализа, т.е. позволяет изучать статистическую связь одной переменной (Y) с множеством объясняющих ее факторов (X) и оценить, какую роль играет одна или несколько независимых переменных в объяснении зависимой.

Как и в линейном регрессионном анализе, качество полученной модели ЛР можно проверить по критерию согласия, для чего часто используется Макфадден R^2 (псевдо R^2), который может принимать значение между 0 и 1. Но важно отметить, что псевдо R-квадрат позволяет лишь сравнивать модели с одной и той же зависимой переменной. Кроме того, можно просто рассчитать количество верно классифицированных индивидов, разделенное на общее количество наблюдений (*Count R2*).

С более подробной информацией о ЛР можно ознакомиться в [23; 24]. Интересно отметить, что пробит регрессия отличается от ЛР лишь другим выбором кумулятивной функции распределения: нормального распределения вместо логистической функции, т.е. уравнение (1) заменяется на (3).

$$P(X) = \int_{-\infty}^{x_i} \frac{1}{\sqrt{2\sigma^2\pi}} e^{-(X-\mu)^2/2\sigma^2} \quad (3)$$

Дискриминантный анализ (ДА)

Для множественных измерений ($X: X_1, X_2, \dots, X_n$) с нормальным распределением из двух совокупностей (Y_1, Y_2), которые имеют разные средние величины внутри каждой совокупности (μ_1, μ_2), но однородность ковариаций (Σ), можно сформулировать следующую функцию как линейную комбинацию множественных измерений (независимых переменных):

$$Y = \beta' X = \beta_1 X_1 + \beta_2 X_2 \dots + \beta_n X_n \quad (4)$$

$$\beta = (\mu_1 - \mu_2)' \Sigma^{-1} \quad (5)$$

$$Y = \beta' X = (\mu_1 - \mu_2)' \Sigma^{-1} X, \quad (6)$$

где параметры β представляют собой коэффициенты дискриминантной функции, которая преобразует многомерные совокупности (Y_1, Y_2) в одномерные для того, чтобы максимально разделить средние в одномерных совокупностях относительно дисперсии

внутри группы на линейную составную функцию. Эти средние величины используются для классификации после определения порогового значения (*cutoff point*). Тогда средние величины таковы:

$$\mu_{1Y} = \beta' \mu_1, \quad \mu_{2Y} = \beta' \mu_2, \quad (7)$$

а пороговое значение (m):

$$m = \frac{1}{2}(\mu_{1Y} + \mu_{2Y}) \quad (8)$$

$$m = \frac{1}{2}(\beta' \mu_1 + \beta' \mu_2) \quad (9)$$

$$m = \frac{1}{2} \beta' (\mu_1 + \mu_2). \quad (10)$$

Для каждого нового случая с измерениями X_i правило классификации выглядит следующим образом:

$$X_i \text{ к совокупности } Y_1 \text{ если } \beta' X_i - m \geq \log_e \frac{p_2}{p_1}, \quad (11)$$

где p_2 и p_1 – априорные вероятности каждой совокупности.

В противном случае X_i относится к совокупности Y_2 .

ДА – линейная модель, поэтому для оценки ее коэффициентов можно использовать метод наименьших квадратов (МНК). Таким образом, линия регрессии (по результатам МНК) служит границей, которая разделяет индивидов по группам, т.е. линия регрессии дискриминирует и предсказывает принадлежность к группе (индивиды-наблюдения, оказавшиеся ниже линии принадлежат совокупности Y_1 , выше линии – к совокупности Y_2). ДА хорошо описывает ключевые различия между совокупностями и как независимые переменные объясняют эти различия, но характер взаимосвязи переменных, ее вид и параметры нельзя определить с помощью этого анализа, потому что это не регрессионный анализ.

Насколько точно ДА способен предсказывать, к какой совокупности принадлежит конкретный индивид, можно оценить при сравнении результатов априорной или апостериорной клас-

сификации. Количество верно классифицированных индивидов в выборке, по которой была проведена оценка дискриминирующей функции (апостериорная классификация), обычно предпочтительнее, чем проверка на новой выборке (априорная классификация). Поэтому, когда есть возможность, рекомендуется тестировать модель ДА на новых наблюдениях.

Представленная модель позволяет распределить наблюдения только по двум группам и получить только одну дискриминантную функцию. Р.А. Фишер первым разработал эту модель, поэтому она также известна как линейный дискриминантный анализ Фишера [25]. Возможности модели можно также расширить на случай большего числа групп. Подробную информацию об этом методе можно найти в [24; 26].

Какой метод лучше использовать?

Очевидно, что прежде чем использовать тот или иной метод, нужно убедиться, что данные отвечают необходимым предпосылкам его использования (см. *табл. 1*). Если используется метод наименьших квадратов или максимального правдоподобия, также необходимо обратить внимание на их предпосылки и ограничения (мультиколлинеарность, гетероскедастичность и т.д.). Для дискриминантного анализа особенно важны однородность ковариаций и нормальное распределение множества независимых переменных, однако неясно насколько сильно и негативно игнорирование данных предпосылок влияет на результаты, и сам Р.А. Фишер не уделил достаточного внимания этим предположениям [25]. Однородность ковариаций и нормальное распределение не имеют значения для логистической регрессии, поэтому при их невыполнении рекомендуется использовать ЛР.

Когда объемы получившихся групп сильно разнятся, например, 10% принадлежат одной группе и 90% – другой, нельзя ожидать правильную линию для разделения индивидов, поэтому

рекомендуется использовать ЛР, так как ее функция нелинейна. В целом, с точки зрения качества классификации оба метода приводят к схожим результатам¹, но когда априорные вероятности неравны, логистическая регрессия минимизирует ошибку для меньшей группы, а дискриминантный анализ – для большей группы.

Таблица 1

РЕКОМЕНДАЦИИ ПО ИСПОЛЬЗОВАНИЮ ЛР И ДА

Предположение	ЛР	ДА
Отсутствие однородности ковариаций множества независимых переменных	X	
Отсутствие нормального распределения множества независимых переменных	X	
Большой размер выборки и отсутствие однородности ковариаций и нормального распределения	X	
Объемы получившихся групп сильно разнятся	X	
Ни одно из предшествующих предположений не выполняется		X

Примечание: Такие рекомендации можно найти в [27–30].

Исследование в Ярославской области

По данным Всероссийской переписи 2010 г. в Ярославской области проживает 1 272 485 жителей, хотя в последние годы численность населения несколько сократилась. Область находится недалеко от Москвы, и большинство жителей являются русскими (96%), в отличие от многих других регионов, где проживают и другие национальности. Отметим, что национальность служит важной объясняющей переменной образовательных результатов [12], но такая переменная фактически не может оказывать сильного влияния в Ярославской области. Несмотря на отсутствие

¹ Для обсуждения возможностей этих методов в сравнении с другими, см. [30].

национальной компоненты, результаты исследования в Ярославской области интересно сравнивать с результатами предыдущих исследований по регионам России, обладающими отличными демографическими характеристиками.

Для исследования образовательных траекторий школьников в рамках проекта «Мониторинг образовательных и трудовых траекторий выпускников школ и вузов» была разработана анкета, которая позволяла собрать информацию о демографических и семейных характеристиках, планах на будущее (учебу и работу) и личностных особенностях школьников. Применялась модель выборки, используемая в Международном исследовании образовательных достижений учащихся (*Programme for International Student Assessment, PISA*) [12]. Случайным образом было отобрано 65 школ, при этом учитывалось их местоположение (областной центр, город, село) и тип школы (средняя общеобразовательная, гимназия, лицей и т.д.). Анкету заполнили 2003 ученика 9-х классов в феврале и марте 2010 г., средний возраст – 15 лет.

Траектории

Были сконструированы две дихотомические переменные, описывающие образовательные планы респондентов, на основе вопросов о том, что школьники собираются делать после 9-го класса и спустя три года. Переменной ТРАЕК1 присваивался 1, если через год школьник планировал продолжать обучение в 10-м классе (56,2%), и 0 в другом случае (42%): профессиональное обучение, работу или, наконец, не имел определенных планов на будущее. Переменная ТРАЕК2 описывала планы школьников через три года. Ей присваивалась 1, если школьник отмечал, что собирается через три года поступать в вуз (57,2%), и 0 в другом случае (39,3%). В анализ не были включены школьники, которые не ответили на данные вопросы, 1,8% опрошенных для переменной ТРАЕК1 и 3,5% – ТРАЕК2.

Независимые переменные

Преыдушие эмпирические исследования, проводившиеся в России, показали, что успеваемость и уровень образования родителей служат основными предикторами образовательных траекторий (см. ниже). В нашем исследовании задействованы эти ключевые объясняющие переменные, а также 8 других независимых переменных, которые в литературе часто используются для предсказания, а в данном случае также работают как контрольные переменные.

Для измерения успеваемости (ОЦЕНКА) в качестве критерия была принята средняя годовая оценка за предыдущий год обучения по алгебре, геометрии, русскому языку, литературе, истории, физике, химии и биологии (школьники сами указывали свои оценки).

Отдельно был включен уровень образования матерей (ОБРАЗОВАНИЕ_М) и отцов (ОБРАЗОВАНИЕ_О). Обычно эти переменные коррелируют, но в данном случае коэффициент корреляции достаточно низкий (см. табл. 2), что не должно вести к проблеме мультиколлинеарности. Интересно отметить, что при включении этих переменных также можно определить, кто из родителей сильнее влияет на образовательные траектории детей.

Уровень образования родителей был измерен по шкале от 1 до 7, где 1 – самый низкий уровень образования – 9 классов школы или меньше (2,1% матерей и 3% отцов), 7 – самый высокий уровень – ученая степень или два высших образования (5,1% матерей и 4,2% отцов). Между ними в порядке возрастания располагались: среднее общее образование (6,4% матерей и 6,8% отцов); начальное профессиональное образование (3% матерей и 5,3% отцов); среднее профессиональное образование (30,9% матерей и 29,5% отцов); незаконченное высшее образование (3,8% матерей и 2,9% отцов) и высшее образование (31,9% матерей и 21,7% отцов)¹.

¹ Образование родителей также можно измерить количеством лет обучения, но в данном случае предпочтительна порядковая шкала, так как она отражает прямую связь между зависимыми и независимыми переменными [11, с. 226].

Первая независимая контрольная переменная – субъективная оценка учеником материального положения своей семьи (ДОХОД), которое измерялось с помощью вопроса о том, на что в семье хватает денег. Переменной присваивался 1 балл, если респондент отмечал, что денег хватает только на еду (0,7%), и 6 баллов, если у школьника нет денежных проблем, при желании его семья может купить дорогой автомобиль (10,8%). Промежуточные значения включали следующие варианты: на еду денег хватает, но есть серьезные проблемы с покупкой одежды (2 балла, 2,8%); денег хватает на еду и одежду, но было бы трудно купить телевизор (3 балла, 16,2%); денег хватает на еду, одежду и можем позволить себе купить телевизор (4 балла, 22%); на имеющиеся деньги можем купить все, кроме дорогих вещей (5 баллов, 31,8%)¹.

Для измерения второй контрольной переменной был использован вопрос, сколько книг есть дома у школьников (КНИГИ)? В зависимости от количества книг переменной присваивались баллы в порядке возрастания: 0–10 книг (1 балл, 7,4%); 11–25 книг (2 балла, 13,5%); 26–100 книг (3 балла, 29,5%); 101–200 книг (4 балла, 21%); 201–500 книг (5 баллов, 16%) и более 500 книг (6 баллов, 11,4%). Предполагается, что школьники, семьи которых владеют большим количеством книг, имеют более высокий культурный капитал.

Третья контрольная переменная – информация о том, с кем живет школьник. Была построена дихотомическая переменная (СЕМЬЯ), которой присваивалось значение 1, если школьник указывал, что живет с матерью и отцом (70,9%), и 0 в другом случае (29,1%). Мы предполагаем, что ребенок, который растет в полной семье, тесно окружен социальными связями и получает больший социальный капитал.

В качестве четвертой контрольной переменной использована информация о наличии специализации у школы. Была построена

¹ Важно заметить, что вопрос о доходе семей – чувствительный, и мы не смогли бы получить более точную информацию о доходе семьи, поэтому используется порядковая переменная.

дихотомическая переменная (ШКОЛА), которой присваивалось значение 1, если ученик указывал, что его школа предлагает особую специализацию: языковую, математическую, спортивную и проч. (37,5%), и 0 в другом случае (62,5%).

Как пятая контрольная переменная была использована информация о том, в какой степени школьник склонен согласиться с суждением: «Как правило, я старательно выполняю все, чем мне приходится заниматься». Была построена переменная (СТАРАТЕЛЬНОСТЬ), которой присваивались баллы от 1 (полностью не согласен, 4,9%) до 5 (полностью согласен, 16,4%).

Здоровье школьников также может влиять на их успехи в учебе. Поэтому в качестве шестой независимой переменной был использован вопрос: возникали ли у школьника сложности с учебной из-за состояния здоровья, например хронических заболеваний, проблем со зрением, слухом, опорно-двигательным аппаратом. Была построена переменная (ЗДОРОВЬЕ), которой присваивался 1 балл, если школьник указывал «часто» (5,1%); 2 балла «время от времени» (12,8%); 3 балла «редко» (35,9%) и 4 балла «никогда не возникали» (45,6%).

Считается, что хорошая компания, т.е. друзья, может позитивно воздействовать на поведение школьника. В нашей анкете школьники охарактеризовали отношение своих друзей к учебе. С помощью этой информации была построена седьмая контрольная переменная (ДРУЗЬЯ), которая принимала значение 1, если школьник указывал, что все или почти все его друзья собираются идти в 10 класс или уже учатся в 10 (или 11) классе (23,9%) и 0 – в остальных случаях (63,3%)¹.

Последняя контрольная переменная – пол школьника, было опрошено 55,1% девушек и 44,9% юношей, и затем построена дихотомическая переменная (ПОЛ): 1 = юноша и 0 = девушка.

¹ Мы предположили, что школьник неплохо знает учебные планы своих друзей. Кроме того, он мог выбрать вариант: затрудняюсь ответить (12,9%).

Эконометрическая модель

В ряде работ [27–30] для сравнения ДА и ЛР использовали метод Монте-Карло, именно на результатах такого сравнения основаны рекомендации, приводимые в *табл. 1*. Здесь же задача проще: сравниваются результаты на конкретных примерах при анализе детерминант образовательных траекторий в Ярославской области, учитывая упомянутые выше свойства методов.

В данном случае трудно определить с помощью описанных выше рекомендаций (см. *табл. 1*), какой метод следует использовать. Очевидно, что мы используем категориальные переменные, отчего у них отсутствует нормальное распределение. Но на практике не так легко оценить, насколько это влияет на способность ДА по сравнению ЛР [28]. Кроме того, размер выборки достаточно велик, а благодаря центральной предельной теореме известно, что это помогает приблизиться к нормальному распределению. Объемы получившихся групп также не сильно разнятся, поэтому данный критерий не помогает выбрать метод.

При корреляционном анализе между предыдущими зависимыми и независимыми переменными были получены низкие коэффициенты корреляции (самая большая корреляция 0,45 между образованием матери и отца). Вследствие этого можно сказать, что проблема мультиколлинеарности отсутствует (см. *табл. 2*).

Таким образом, чтобы узнать, какие переменные объясняют, предсказывают и различают (дискриминируют) образовательные планы школьников на будущее в Ярославской области, была построена модель (12):

$$\begin{aligned} \text{ТРАЕК}_i = & \beta_0 + \beta_1 \text{ОЦЕНКА}_i + \beta_2 \text{ОБРАЗОВАНИЕ}_i + \beta_4 \text{ДОХОД}_i + \\ & \beta_5 \text{КНИГИ}_i + \beta_6 \text{СЕМЬЯ}_i + \beta_7 \text{ШКОЛА}_i + \beta_8 \text{СТАРАТЕЛЬНОСТЬ}_i + \\ & \beta_9 \text{ДРУЗЬЯ}_i + \beta_{10} \text{ЗДОРОВЬЕ}_i + \beta_{11} \text{ПОЛ}_i, \end{aligned} \quad (12)$$

где ТРАЕК – дихотомические переменные ТРАЕК1 или ТРАЕК2 (продолжать ли обучения в 10-м классе и в вузе).

МАТРИЦА

Переменная	ТРАЕК1	ТРАЕК2	ОЦЕНКА	ОБРАЗОВАНИЕ_М	ОБРАЗОВАНИЕ_О
ТРАЕК1	1				
ТРАЕК2	0,526**	1			
ОЦЕНКА	0,471**	0,445**	1		
ОБРАЗОВАНИЕ_М	0,183**	0,185**	0,214**	1	
ОБРАЗОВАНИЕ_О	0,223**	0,238**	0,227**	0,456**	1
ДОХОД	0,042	0,044	0,025	0,137**	0,129**
КНИГИ	0,169**	0,147**	0,222**	0,230**	0,218**
СЕМЬЯ	0,071**	0,036	0,023	0,032	0,020
ШКОЛА	0,121**	0,116**	0,048*	0,100**	0,116**
СТАРАТЕЛЬНОСТЬ	0,122**	0,135**	0,219**	0,046	0,025
ДРУЗЬЯ	0,294**	0,229**	0,313**	0,184**	0,202**
ЗДОРОВЬЕ	0,080**	0,063**	0,072**	-0,010	-0,004
ПОЛ	0,154**	0,154**	0,232**	0,041	0,012

Примечание. Коэффициенты корреляции Пирсона:

* – статистически значимо на уровне 1%;

** – статистически значимо на уровне 5%.

Таблица 2

КОРРЕЛЯЦИЙ

ДОХОД	КНИГИ	СЕМЬЯ	ШКОЛА	СТАРАТЕЛЬНОСТЬ	ДРУЗЬЯ	ЗДОРОВЬЕ	ПОЛ
1							
0,102**	1						
0,084**	0,020	1					
0,040	0,011	0,016	1				
0,060*	0,051*	0,042	0,053*	1			
0,045	0,177**	0,005	0,198**	0,063**	1		
0,072**	0,023	0,029	0,003	0,052*	0,062**	1	
0,012	0,056*	0,035	0,029	-0,085**	0,102**	0,089**	1

Результаты логистической регрессии

Сначала была использована логистическая регрессия, чтобы рассчитать коэффициенты модели (12). В табл. 3 и 4 отображены основные результаты. Была применена стратегия Юбера-Уайта и получены робастные оценки стандартных ошибок (*robust estimators*), данная стратегия часто используется для того, чтобы сделать поправку на корреляцию случайных ошибок, что помогает исправить ситуацию с гетероскедастичностью. В выборку попали разные школьники, из этого следует, что существует возможность наличия гетероскедастичности, но при графическом анализе случайных ошибок не наблюдалось.

В модели с переменной ТРАЕК1 (см. табл. 3) коэффициенты независимых переменных ОЦЕНКА, ОБРАЗОВАНИЕ_О и ДРУЗЬЯ статистически значимы на 1%-м уровне, и коэффициенты ШКОЛА, ЗДОРОВЬЕ и ПОЛ – на 5%-м уровне. Коэффициенты при переменных ОБРАЗОВАНИЕ_М, ДОХОД, КНИГИ, СЕМЬЯ и СТАРАТЕЛЬНОСТЬ статистически не значимы.

Это значит, что школьник, получающий более высокие оценки, учащийся в школе со специализацией и чей отец обладает более высоким уровнем образования, с большей вероятностью будет планировать продолжать обучение в 10-м классе. Школьники, имеющие хорошее здоровье и друзей, планирующих продолжать обучение, также с большей вероятностью последуют их примеру. Негативный коэффициент при переменной ПОЛ означает, что девушки чаще, по сравнению с юношами, планируют продолжить обучение в 10-м классе.

В модели с переменной ТРАЕК2 (табл. 4) результаты схожие. Коэффициенты независимых переменных ОЦЕНКА, ОБРАЗОВАНИЕ_О и ДРУЗЬЯ статистически значимы на 1%-м уровне, коэффициент при переменной ШКОЛА – лишь на уровне 0,1. Коэффициенты ОБРАЗОВАНИЕ_М, ДОХОД, КНИГИ, СЕМЬЯ, СТАРАТЕЛЬНОСТЬ, ЗДОРОВЬЕ и ПОЛ статистически не значимы.

Таблица 3
ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ ДЛЯ ЗАВИСИМОЙ ПЕРЕМЕННОЙ
ТРАЕК1 – ПРОДОЛЖАТЬ ОБУЧЕНИЕ В 10-М КЛАССЕ

	Коэффициенты	Стандартная ошибка	z-критерий	Значимость
КОНСТАНТА	-8,208	0,809	-10,149	0,000
ОЦЕНКА	1,820	0,178	10,242	0,000
ОБРАЗОВАНИЕ_М	-0,018	0,055	-0,321	0,748
ОБРАЗОВАНИЕ_О	0,152	0,054	2,808	0,005
ДОХОД	0,066	0,071	0,930	0,353
КНИГИ	0,063	0,059	1,062	0,288
СЕМЬЯ	0,279	0,184	1,517	0,129
ШКОЛА	0,330	0,157	2,096	0,036
СТАРАТЕЛЬНОСТЬ	-0,027	0,071	-0,381	0,703
ДРУЗЬЯ	0,872	0,207	4,216	0,000
ЗДОРОВЬЕ	0,188	0,091	2,051	0,040
ПОЛ	-0,291	0,159	-1,835	0,067

Примечание.

Метод: Ньютон-Рафсон

Стандартные ошибки и ковариация Юбера-Уайта

$N = 1023$; 679 (1) и 344 (0)

Псевдо R^2 Макфаддена = 0,21

LR-критерий = 278,94; значимость LR = 0,00

Таблица 4
ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ ДЛЯ ЗАВИСИМОЙ
ПЕРЕМЕННОЙ ТРАЕК2 – ПРОДОЛЖАТЬ ОБУЧЕНИЕ В ВУЗЕ

	Коэффициенты	Стандартная ошибка	z-критерий	Значимость
КОНСТАНТА	-6,841	0,763	-8,966	0,000
ОЦЕНКА	1,501	0,169	8,897	0,000
ОБРАЗОВАНИЕ_М	0,033	0,058	0,575	0,565
ОБРАЗОВАНИЕ_О	0,193	0,055	3,535	0,000
ДОХОД	0,035	0,073	0,482	0,630
КНИГИ	0,017	0,056	0,298	0,766

Окончание табл. 4

	Коэффициенты	Стандартная ошибка	z-критерий	Значимость
СЕМЬЯ	0,142	0,182	0,779	0,436
ШКОЛА	0,264	0,156	1,688	0,091
СТАРАТЕЛЬНОСТЬ	0,048	0,070	0,685	0,493
ДРУЗЬЯ	0,551	0,199	2,772	0,006
ЗДОРОВЬЕ	0,107	0,092	1,165	0,244
ПОЛ	-0,223	0,162	-1,375	0,169

Примечание.

Метод: Ньютон-Рафсон

Стандартные ошибки и ковариация Юбера-Уайта

N = 1023; 704 (1) и 319 (0)

Псевдо R² Макфаддена = 0,16

LR-критерий = 208,72; значимость LR = 0,00

Иными словами, обладатели более высоких оценок, хороших друзей и учащиеся в школах со специализацией опять же с большей вероятностью будут планировать продолжить обучение в вузе.

Результаты дискриминантного анализа

Для модели (12) была рассчитана дискриминантная функция¹. В табл. 5 представлены главные результаты для ТРАЕК1. Дискриминантная функция статистически значима, лямбда Уилкса равна 0,76. Но F-критерии и его вероятности показывают, что переменные ДОХОД и СЕМЬЯ статистически не значимы при дискриминации между совокупностями, поэтому их можно исключить из анализа. Переменные ОЦЕНКА (0,78), ДРУЗЬЯ (0,26), и ОБРАЗОВАНИЕ_О (0,17) имеют самые большие стандартизованные коэффициенты, т.е. они вносят самый большой, по сравнению с другими переменными, вклад в различие

¹ Были использованы априорные вероятности, пропорциональные объемам совокупностей в выборке.

Таблица 5
ДИСКРИМИНАНТНЫЙ АНАЛИЗ ДЛЯ ЗАВИСИМОЙ ПЕРЕМЕННОЙ ТРАЕК1 – ПРОДОЛЖАТЬ
ОБУЧЕНИЕ В 10-М КЛАССЕ

	Лямбда Уилкса	F- критерий	Значи- мость	Стандарти- зованные коэффи- циенты	Коэффициенты классификации		Структурные коэффициенты
					0	1	
ОЦЕНКА	0,807	244,628	0,000	0,783	13,321	15,073	0,887
ОБРАЗОВАНИЕ_М	0,977	23,906	0,000	0,001	0,647	0,648	0,277
ОБРАЗОВАНИЕ_О	0,964	37,655	0,000	0,177	0,727	0,867	0,348
ДОХОД	0,998	1,763	0,185	0,056	3,248	3,309	0,075
КНИГИ	0,977	24,325	0,000	0,080	0,627	0,695	0,280
СЕМЬЯ	0,999	0,559	0,455	0,104	5,389	5,681	0,042
ШКОЛА	0,989	11,555	0,001	0,130	1,242	1,550	0,193
СТАРАТЕЛЬНОСТЬ	0,994	6,247	0,013	-0,040	1,596	1,553	0,142
ДРУЗЬЯ	0,928	79,563	0,000	0,267	-4,135	-3,426	0,506
ЗДОРОВЬЕ	0,995	5,568	0,018	0,127	3,692	3,868	0,134
ПОЛ	0,977	24,070	0,000	-0,149	3,596	3,242	-0,278

Примечание.

$N = 1023$; 679 (1) и 344 (0)

Собственное (характеристическое) значение = 0,305

Каноническая корреляция = 0,483

Лямбда Уилкса = 0,766; $\chi^2 = 270,13$; Значимость = 0,00

M-критерий Бокса = 156,84; значимость = 0,00

74,9% образцов верно классифицированы

между группами, особенно по результатам успеваемости. Такой результат подтверждается коэффициентами факторной структуры, которые показывают, что ОЦЕНКА (0,88) и ДРУЗЬЯ (0,50) определяют дискриминантную функцию. Кроме того, коэффициенты классификации демонстрируют, что ДРУЗЬЯ и ПОЛ лучше предсказывают, какие школьники не планируют продолжать обучение в 10-м классе. Таким образом, 74,9% респондентов были верно классифицированы по группам.

В табл. 6 представлены основные результаты для ТРАЕК2. Дискриминантная функция статистически значима, лямбда равна 0,82. *F*-критерии показывают, что переменные ДОХОД, СЕМЬЯ и ЗДОРОВЬЕ статистики не значимы и их можно исключить. Переменные ОЦЕНКА (0,75), ОБРАЗОВАНИЕ_О (0,27) и ДРУЗЬЯ (0,19) снова вносят наибольший вклад в дискриминацию. Коэффициенты факторной структуры показывают, что ОЦЕНКА (0,88), ДРУЗЬЯ (0,47) и ОБРАЗОВАНИЕ_О (0,45) маркируют дискриминантную функцию. Как и в ситуации с предсказыванием планов продолжить обучение в школе, коэффициенты классификации демонстрируют, что переменные ДРУЗЬЯ и ПОЛ точнее предсказывают, какие школьники не планируют продолжать обучение в вузе. Метод верно классифицировал 74,9% респондентов.

Обе модели ДА показали большие значения *M*-критерия Бокса, что можно интерпретировать как отсутствие однородности ковариаций. Но были рассчитаны и модели с использованием отдельных ковариационных матриц, и результаты практически не изменились: второй подход верно классифицировал 74,3% респондентов для ТРАЕК1 и 73,8% для ТРАЕК2. Поэтому нет смысла использовать отдельную ковариацию, *M*-критерий Бокса может быть слишком чувствителен к большой выборке и к отклонению от многомерной нормальности, которая имеется в нашем случае.

Таблица 6
ДИСКРИМИНАНТНЫЙ АНАЛИЗ ДЛЯ ЗАВИСИМОЙ ПЕРЕМЕННОЙ ТРАЕК2 – ПРОДОЛЖАТЬ
ОБУЧЕНИЕ В ВУЗЕ

	Лямбда Уилкса	F- критерий	Значи- мость	Стандарти- зованные коэффици- циенты	Коэффициенты классификации		Структурные коэффициенты
					0	1	
ОЦЕНКА	0,853	176,513	0,000	0,756	12,756	14,180	0,888
ОБРАЗОВАНИЕ_М	0,972	29,012	0,000	0,073	0,678	0,729	0,360
ОБРАЗОВАНИЕ_О	0,957	45,881	0,000	0,277	0,731	0,921	0,453
ДОХОД	0,999	1,327	0,250	0,034	3,217	3,249	0,077
КНИГИ	0,985	15,534	0,000	0,026	0,581	0,601	0,263
СЕМЬЯ	1,000	0,035	0,851	0,067	5,246	5,408	0,013
ШКОЛА	0,991	8,790	0,003	0,117	1,136	1,377	0,198
СТАРАТЕЛЬНОСТЬ	0,991	9,024	0,003	0,037	1,655	1,690	0,201
ДРУЗЬЯ	0,953	50,450	0,000	0,195	-4,451	-4,010	0,475
ЗДОРОВЬЕ	0,998	2,398	0,122	0,083	3,608	3,708	0,103
ПОЛ	0,985	16,060	0,000	-0,131	3,723	3,455	-0,268

Примечание.

$N = 1023$; 704 (1) и 319 (0)

Собственное (характеристическое) значение = 0,219

Каноническая корреляция = 0,424

Лямбда Уилкса = 0,82; $\chi^2 = 201,42$; значимость = 0,00

M-критерий Бокса = 156,84; Значимость = 0,00

74,9% образцов верно классифицированы

Заклочение

При применении логистической модели и дискриминантного анализа для определения принадлежности индивида к первой или второй группе (продолжать или нет обучение), нами получены очень схожие результаты. В данном конкретном случае, когда перед нами стоит задача только классификации, разница между двумя методами оказалась несущественной, – оба продемонстрировали хорошие возможности для предсказания.

Следует отметить, что методы дополняют друг друга, благодаря чему можно уточнить насколько результаты их применения робастны, т.е. если мы найдем, что одинаковые независимые переменные являются ключевыми и в ЛР, и в ДА, тогда можно в большей степени доверять результатам для принятия решения. Кроме того, каждый метод лучше выполняет конкретные задачи. По мнению Д.Н. Гухараты, в логистической регрессии степень согласия вторична, а в первую очередь «важен ожидаемый знак коэффициентов и их статистическая и практическая значимость» [23, р. 606]. Поэтому, для изучения статистической зависимости и проверки предсказаний теории при дихотомических зависимых переменных, целесообразно использовать ЛР. При наличии необходимых предпосылок, в задачах классификации можно в равной степени использовать метод ЛР или ДА. Когда нас интересует, как и какие переменные дискриминируют зависимую переменную, лучше использовать ДА.

Согласно результатам предыдущих исследований, проведенных в России, успеваемость – главный предиктор образовательных траекторий, впрочем образование и доход родителей также оказывают на образовательные траектории определенное влияние. С помощью логистической регрессии и дискриминантного анализа также было установлено, что в Ярославской области успеваемость (ОЦЕНКА) представляет собой важный предиктор, но образование отцов (ОБРАЗОВАНИЕ_О) и индикатор планов на обучение

друзей (ДРУЗЬЯ) также играют важную роль, переменная ДРУЗЬЯ в особенности верно предсказывает, сколько и какие школьники не планируют продолжать обучение.

И наоборот, уровень образования матерей (ОБРАЗОВАНИЕ_М), материальное положение семьи (ДОХОД), количество книг дома (КНИГИ), структура семьи (СЕМЬЯ) и насколько старательно ребенок выполняет свои школьные «обязанности» (СТАРАТЕЛЬ-НОСТЬ) не объясняют, или не связаны напрямую с его планами на будущее, в частности, с его образовательной траекторией. Кроме того, ДОХОД и СЕМЬЯ даже не вносят вклад в дискриминантную функцию. Новые исследования позволят проверить эти результаты с помощью других, более надежных показателей.

С помощью логистической регрессии и дискриминантного анализа были построены модели, которые помогают правильно предсказывать планы на продолжение образования примерно 7 из 10 респондентов. Обладая информацией об успеваемости школьника, планах обучения его друзей и уровне образования его отца, можно узнать, собирается ли он продолжать свое обучение.

ЛИТЕРАТУРА

1. Кузьмина Ю.В., Тюменева Ю.А. Читательская грамотность 15-летних школьников: значимость семейных, индивидуальных и школьных характеристик: по данным российской выборки PISA-2009 // *Вопр. образования*. 2011. № 3. С. 164–191.
2. *Tovar-García E.D.* Approaches to Study Educational and Occupational Trajectories and Outcomes // *Papeles de Trabajo sobre Cultura, Educación y Desarrollo Humano*. 2012. No. 8 (2). P. 1–19.
3. *DiMaggio P.* Cultural Capital and School Success: The Impact of Status Culture Participation on the Grades of U.S. High School Students // *American Sociological Review*. 1982. No. 47 (2). P. 189–201
4. *Coleman J.S.* Social Capital in the Creation of Human Capital // *American Journal of Sociology*. 1988. No. 94. P. 95–120.
5. *Hanushek E.* Assessing the Effects of School Resources on Student Performance: An Update // *Educational Evaluation and Policy Analysis*. 1997. No. 19 (2). P. 141–164.

6. *Farkas G.* Cognitive Skills and Noncognitive Traits and Behaviors in Stratification Processes // *Annual Review of Sociology*. 2003. No. 29. P. 541–562.

7. *Lucas S.R.* Effectively Maintained Inequality: Education Transitions, Track Mobility, and Social Background Effects // *American Journal of Sociology*. 2001. No. 106. P. 1642–1690.

8. *Raftery A.E., Hout M.* Maximally Maintained Inequality: Expansion, Reform, and Opportunity in Irish Education, 1921–75 // *Sociology of Education*. 1993. No. 66 (1). P. 41–62.

9. *Rothman S.* The Changing Influence of Socioeconomic Status on Student Achievement: Recent Evidence from Australia // *LSAY Conference Papers*. Paper Presented at the Annual Meeting of the American Educational Research Association. Chicago, 21–25 April, 2003.

10. *Schildberg-Hoerisch H.* Does Parental Employment Affect Children's Educational Attainment? // *Economics of Education Review*. 2011. No. 30 (6). P. 1456–1467.

11. *Breen R., Jonsson J.O.* Inequality of Opportunity in Comparative Perspective: Recent Research on Educational Attainment and Social Mobility // *Annual Review of Sociology*. 2005. No. 31. P. 223–243.

12. *OECD.* PISA 2009 Results: Overcoming Social Background – Equity in Learning Opportunities and Outcomes. Vol. II. OECD, 2010.

13. *Dostie B., Jayaraman R.* Determinants of School Enrollment in Indian villages // *Economic Development and Cultural Change*. 2006. No. 54. P. 405–421.

14. *Glick P., Sahn D.* Schooling of Girls and Boys in a West African Country: The Effects of Parental Education, Income, and Household Structure // *Economics of Education Review*. 2000. No. 19 (1). P. 63–87.

15. *Mani S., Hoddinott J., Strauss J.* Determinants of Schooling Outcomes: Empirical Evidence from Rural Ethiopia // *Fordham University, Department of Economics, Fordham Economics Discussion Paper Series*. 2009.

16. *Зарицкий Т.* Культурный капитал и доступность высшего образования // *Вестн. обществ. мнения*. 2006. № 2 (82). С. 47–61.

17. *Мкртчян Г.М., Шакин Е.Ю.* Факторы социальной стратификации молодежи в сфере образования, на рынке труда и потребления // *Мониторинг обществ. мнения*. 2004. № 3 (71). С. 95–105.

18. *Очкина А.В.* Культурный капитал семьи как фактор социального поведения // *Мир России*. 2010. № 1. С. 67–88.

19. *Петренко Е.С., Галицкая Е.Г.* Ресурсный потенциал семьи и образовательные траектории детей и взрослых // *Вопр. образования*. 2007. № 3. С. 240–254.

20. *Чередниченко Г.А.* Образовательные и профессиональные траектории выпускников средних школы // *Социол. исслед.* 2010. № 7. С. 88–95.

21. *Кузьмина Ю.В., Тюменева Ю.А., Попов Д.С.* Современные образовательные траектории школьников и студентов // *Социол. исслед.* 2012. № 2. С. 135–142.

22. *Roshchina Y.* Accessibility of Professional Education in Russia // Berlin, ESCIRRU Working Paper No. 13. 2010. P. 1–48.
23. *Gujarati D.N.* Basic Econometrics, 4th ed. USA: McGraw Hill, 2004.
24. *Hastie T., Tibshirani R., Friedman J.* The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. N.Y.: Springer, 2008.
25. *Fisher R.A.* The Use of Multiple Measurements in Taxonomic Problems // Annals of Eugenics. 1936. No. 7(2). P. 179–188.
26. Электронный учебник по статистике. М.: StatSoft, 2012. URL: <http://www.statsoft.ru/home/textbook/default.htm> (дата обращения: 22.08.2013).
27. *Fan X., Wang L.* Comparing Linear Discriminant Function with Logistic Regression for the Two-group Classification Problem // Journal of Experimental Education. 1999. No. 67. P. 265–286.
28. *Lei P., Koehly L.M.* Linear Discriminant Analysis Versus Logistic Regression: A comparison of Classification Errors in the Two-group Case // Journal of Experimental Education. 2003. No. 72. P. 25–49.
29. *Rausch J.R., Kelley K.* A Comparison of Linear and Mixture Models for Discriminant Analysis under Nonnormality // Behavior Research Methods. 2009. No. 41. P. 85–98.
30. *Holden J.E., Finch W.H., Kelley K.* A Comparison of Two-Group Classification Methods // Educational and Psychological Measurement. 2011. No. 71(5). P. 870–901.