

## **Предсказание успеваемости в вузе на основе результатов ЕГЭ на примере факультета экономики ГУ-ВШЭ**

### **1. Роль ЕГЭ в процедуре приема в вузы**

С 2009 года ЕГЭ стал единой формой итоговой аттестации всех выпускников школ и приема абитуриентов в вузы. Тем самым завершился переход от системы приема в вузы по результатам вступительных экзаменов к единому государственному экзамену. Для востребованных направлений подготовки число претендентов на обучение превышает число доступных бюджетных мест в вузах, поэтому возникает необходимость отбора абитуриентов. Процедура отбора призвана обеспечить качество будущего студента через проецирование критериев отбора на будущие успехи. Эффективность системы отбора в вуз можно оценивать исходя из того, насколько она в состоянии предоставить абитуриентам с относительно лучшими результатами экзаменов большие возможности по обучению в привлекательных для них вузах.

Выбор дисциплин вступительных испытаний и шкал их оценивания играет важнейшую роль в процедуре отбора абитуриентов. Выбор предметов экзаменов определяет набор тестируемых специфических способностей абитуриентов. Размер шкалы оценивания выполняет функцию весового коэффициента, при суммировании оценок большее значение имеет предмет, имеющий больший размер шкалы.

Выбор набора предметов и их весов должен опираться на статистический анализ связей между результатами ЕГЭ и показателями успеваемости в вузе. Эти связи специфичны для направления, по которому ведется обучения, а также зависят от особенностей образовательной программы того или иного вуза. Простая процедура выбора оптимального набора дисциплин и шкал оценивания основана на регрессионном анализе линейной модели, в которой в качестве объясняемой переменной выступает скалярный интегральный показатель вузовской успеваемости (рейтинг студентов), а в роли объясняющих факторов – набор результатов школьных выпускных экзаменов. Коэффициенты при объясняющих переменных (оценках за тот или иной предмет) в данной модели служат основой для нахождения весовых коэффициентов, ассоциируемых с конкретной дисциплиной.

Оценка модели позволяет определить вектор оптимального прогнозирования вузовской успеваемости по школьным дисциплинам. При неоптимальном выборе дисциплин вступительных испытаний и шкал оценок происходит неблагоприятный отбор абитуриентов. В вуз частично попадают учащиеся, демонстрирующие худшую успеваемость, чем некоторые абитуриенты, отсеянные во время приемной кампании.

Стандартизированные тестирования школьников проводятся во многих странах. Например, в США многие вузы учитывают при приеме результаты стандартизированных тестов SAT. В отличие от российского ЕГЭ, SAT имеет более чем столетнюю историю. Исследование эффективности SAT также имеет довольно длинную историю. Еще в 1960 году Fishman, Pasanella [1] исследовали корреляционные связи между результатом SAT и средним баллом за первый год обучения в колледже. Коэффициент корреляции колебался от 0,34 до 0,82, среднее значение составило 0,61. В России данные для подобных исследований появились совсем недавно.

## 2. Результаты ЕГЭ и вузовская успеваемость после первого года обучения

Для эмпирического анализа использовались данные о 158 студентах факультета экономики НФ ГУ-ВШЭ и 157 студентах факультета экономики ГУ-ВШЭ, поступивших в 2009 году по результатам ЕГЭ. Данные о результатах ЕГЭ при поступлении в 2009 г. находились в открытом доступе на официальном сайте ГУ-ВШЭ [2]. На рисунке 1 приведены диаграммы рассеяния результатов ЕГЭ по трем предметам: математике, обществознанию, русскому языку. Баллы по этим дисциплинам учитываются при поступлении на факультет экономики в Нижегородский филиал ГУ-ВШЭ. При поступлении на факультет экономики в ГУ-ВШЭ, Москва, необходимо было предоставить также результат ЕГЭ по иностранному языку. На рисунке 2 приведены диаграммы рассеяния результатов ЕГЭ по четырем предметам.

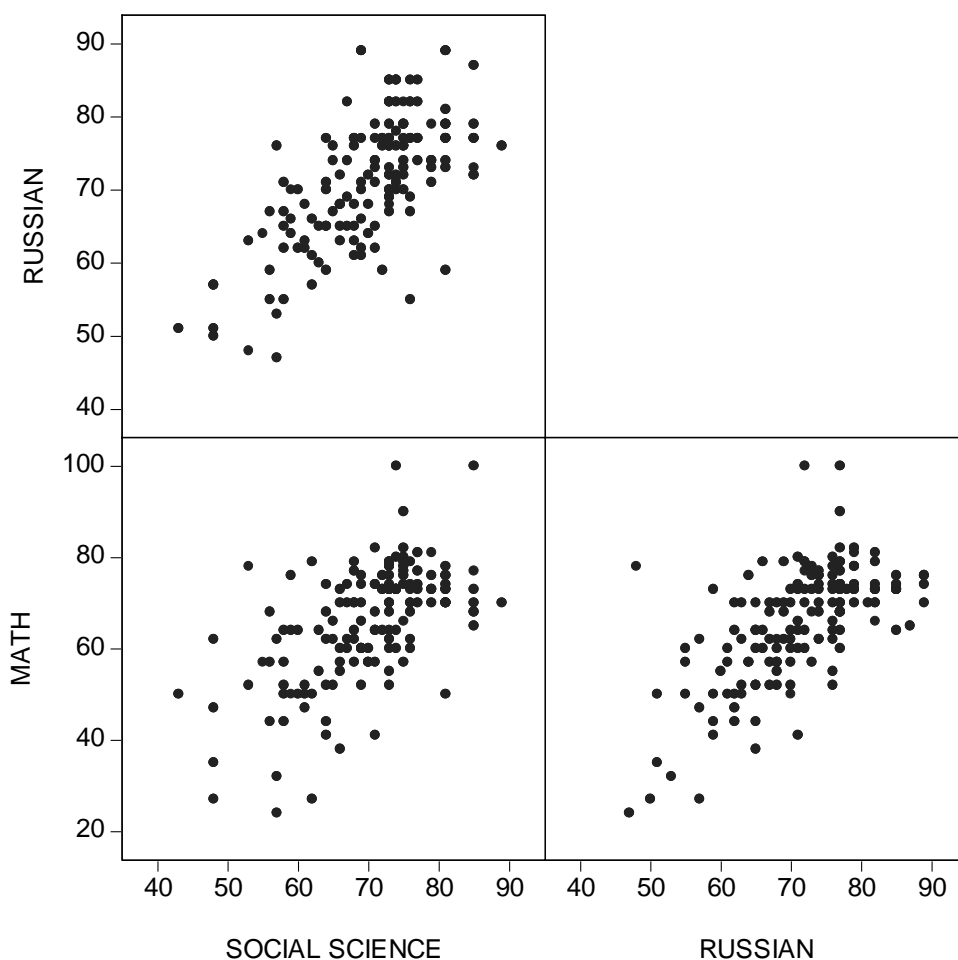


Рис. 1. Диаграммы рассеяния результатов ЕГЭ по трем предметам, Нижний Новгород.

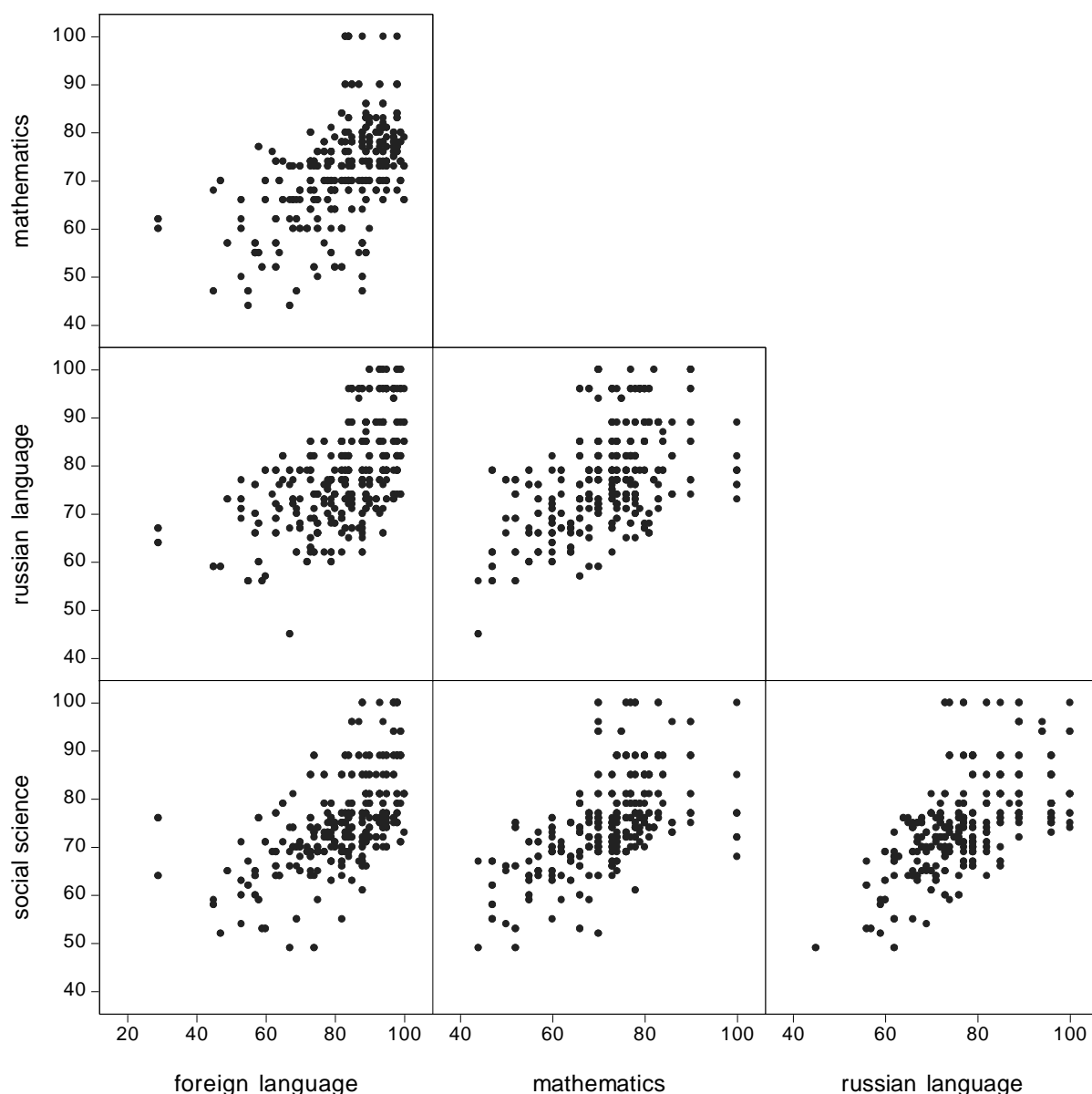


Рис. 2. Диаграммы рассеяния результатов ЕГЭ по четырем предметам, Москва.

В таблицах 1 и 2 представлены ковариационная и корреляционная матрицы результатов поступивших абитуриентов.

В таблице 2 представлены результаты анализа главных компонент для Нижнего Новгорода. Первая главная компонента с приблизительно равной долей предметов может быть интерпретирована как фактор общих способностей абитуриентов. Таким образом общие способности объясняют 75% вариации баллов. Вторая компонента характеризует проявления математических способностей учащихся в противовес «гуманитарным». Третья компонента может рассматриваться как показатель развития чисто лингвистических навыков.

Аналогичный анализ для 4 дисциплин вступительных испытаний в ГУ-ВШЭ, Москва, приведен в таблице 4.

Таблица 1. Корреляционная матрицы баллов ЕГЭ по трем предметам, Нижний Новгород.

	Математика	Русский язык	Обществознание
Математика	1.000000		
Русский язык	0.635016	1.000000	
Обществознание	0.587211	0.659993	1.000000

Таблица 2. Корреляционная матрицы баллов ЕГЭ по 4 предметам, Москва.

	Математика	Русский язык	Обществознание	Иностр. язык
Математика	1.000000			
Русский язык	0.297166	1.000000		
Обществознание	0.410145	0.448863	1.000000	
Иностр. язык	0.417826	0.503463	0.488760	1.000000

Таблица 3. Главные компоненты для совокупности результатов трех экзаменов, Нижний Новгород.

Собственные числа			
	Значение	Относительная доля суммарной дисперсии	Накопленная относительная доля суммарной дисперсии
1	2.255292	0.7518	0.7518
2	0.415177	0.1384	0.8902
3	0.329531	0.1098	1.0000
Собственные вектора			
Переменная	PC 1	PC 2	PC 3
Математика	0.567232	0.770544	0.290707
Русский язык	0.589380	-0.133260	-0.796789
Обществознание	0.575221	-0.623301	0.529732

Таблица 4. Главные компоненты для совокупности результатов 4 экзаменов, Москва.

Собственные числа				
	Значение	Относительная доля суммарной дисперсии	Накопленная относительная доля суммарной дисперсии	
1	2.289860	0.5725	0.5725	
2	0.711260	0.1778	0.7503	
3	0.528237	0.1321	0.8823	
4	0.470643	0.1177	1.0000	
Собственные вектора				
Переменная	PC 1	PC 2	PC 3	PC 4
Математика	0.451244	0.801319	0.277142	0.278315
Русский язык	0.492716	-0.583063	0.320235	0.560997
Обществознание	0.517516	-0.005391	-0.855197	0.028044
Иностр. язык	0.534591	-0.133777	0.298797	-0.779126

Основной интерес представляет связь результатов ЕГЭ 2009 года и рейтинга студентов после первого года обучения. По регламенту ГУ-ВШЭ рейтинг формируется как взвешенная числом кредитов сумма итоговых оценок, полученных студентом по учебным дисциплинам первого курса. Максимально возможное значение рейтинга – 600. Данные о рейтинге обновляются на сайте два раза в год. На рисунке 3 показаны диаграммы рассеяния для Нижегородского филиала ГУ-ВШЭ, характеризующие связи между рейтингом и баллами ЕГЭ по трем предметам. Нулевые значения рейтинга означают, что студент выбыл. Диаграммы для ГУ-ВШЭ, Москва, приведены на рисунке 4.

Очевидно, рейтинг по итогам первого года в вузе положительно коррелирует с баллами, полученными на ЕГЭ в школе. Для более детального исследования связи использовался регрессионный анализ.

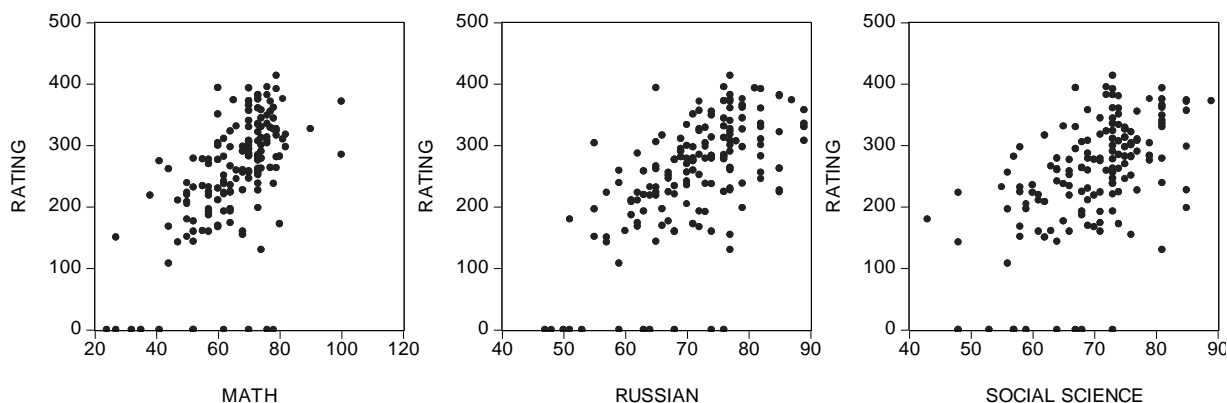


Рис. 3. Зависимость рейтинга после 1 курса от баллов ЕГЭ по дисциплинам вступительных испытаний, Нижний Новгород.

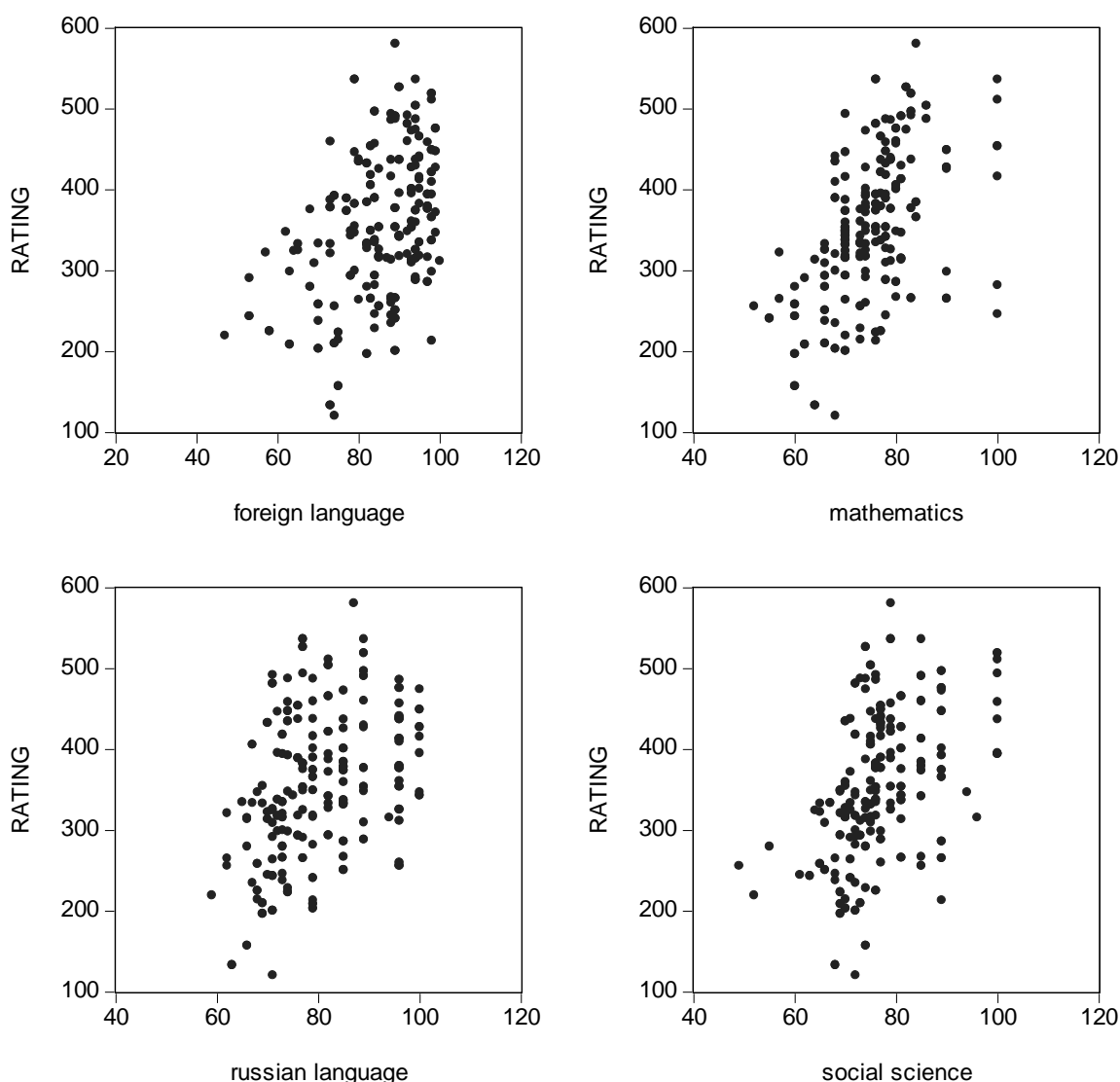


Рис. 4. Зависимость рейтинга после 1 курса от баллов ЕГЭ по дисциплинам вступительных испытаний, Москва.

В таблице 5 даны оценки регрессии для Нижнего Новгорода, в качестве зависимой переменной выступает рейтинг, в качестве объясняющих переменных – баллы ЕГЭ. Все коэффициенты оказались значимы на 1% уровне. Весовые коэффициенты при ЕГЭ по трем дисциплинам близки друг к другу, что свидетельствует о преобладающей роли общих способностей как фактора хорошей успеваемости по дисциплинам первого курса. Тем не менее, обращает на себя внимание то, что наибольшее значение имеет коэффициент при баллах ЕГЭ по русскому языку, самый маленький коэффициент – при математике. Регрессия объясняет 48% вариации рейтинга.

Аналогичные расчеты для факультета экономики в Москве приведены в таблице 6. При включении баллов за все четыре вступительные испытания, коэффициент при математике значим на 1% уровне, при обществознании и русском языке – на 5% уровне, при иностранном языке – на 10 % уровне. Также приведены регрессионные оценки без

иностранного языка, в этом случае коэффициенты при предметах значимы на 1% уровне. Коэффициент детерминации в Москве 36%.

Таблица 5. Регрессионные оценки связи рейтинга и ЕГЭ, Нижний Новгород.

Переменная	Коэффициент	t-статистика
Константа	-312.93	-4.19
Математика	2.74	2.68
Русский язык	3.58	2.92
Обществознание	3.52	2.96
$R^2=0.478$		

Таблица 6. Регрессионные оценки связи рейтинга и ЕГЭ, Москва.

Переменная	Коэффициент	t-статистика	Коэффициент	t-статистика*
Константа	-256.11	-4.45	-245.42	-4.16
Математика	3.41	3.80	3.67	4.08
Русский язык	1.62	2.56	1.98	3.32
Обществознание	1.85	2.43	2.15	2.91
Иностр. язык	0.97	1.78		
$R^2=0.384$		$R^2=0.375$		

\* ковариационная матрица рассчитана с учетом гетероскедастичности остатков (White)

### 3. Заключение

В работе исследовались статистические связи результатов ЕГЭ по отдельным предметам – дисциплинам вступительных испытаний – для студентов, закончивших в 2010 году первый курс на факультетах экономики ГУ-ВШЭ в Нижнем Новгороде и Москве. Методами регрессионного анализа исследовались зависимости рейтингов студентов после первого года с результатами ЕГЭ по математике, обществознанию, русскому языку и иностранному языку.

Баллы ЕГЭ являются неплохими предсказателями рейтинга по итогам первого года, коэффициент детерминации оказался около 48% в Нижнем Новгороде и 38% в Москве. Относительное влияние баллов ЕГЭ по различным дисциплинам на рейтинг различается. В Нижнем Новгороде наибольший вклад дает русский язык, затем с минимальным отрывом обществознание, затем математика. В Москве явно лидирует математика, за ней следуют обществознание и русский, иностранный язык оказался значим лишь на 10% уровне.

### Литература

1. Fishman J.A., Pasanella A.K. College admission selection studies // Review of Educational Research, 30(4), 1960. pp. 298–310
2. Официальный сайт ГУ-ВШЭ. – [www.hse.ru](http://www.hse.ru)