Национальный Исследовательский Университет «Высшая Школа Экономики»
МЕЖДУНАРОДНЫЙ ИНСТИТУТ ЭКОНОМИКИ И ФИНАНСОВ

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

На тему «Анализ и оптимизация успеваемости студентов».

Студентка 4 курса, 5 группы
Суншева Алина
Сафарбиевна

Научный руководитель
К.э.н., доцент
Черняк Владимир Ильич

Москва, 2013 г.

Благодарности

Моему научному руководителю, Черняк Владимиру Ильичу, за неформальное отношение и искреннюю заинтересованность, моральную поддержку, добрые слова и грамотные наставления, за мой интерес к этой теме и возможность ею заниматься.

Замкову Олегу Олеговичу и Белянину Алексею Владимировичу, за интересные идеи и помощь в подборе литературы.

Моим друзьям и однокурсникам Цой Сергею и Соломатину Ярославу, за совместные обсуждения и споры, которые помогли мне в этой работе.

Моей Маме, за её постоянную помощь и поддержку, без которой эта работа не была бы написана.

Оглавление

Введение.

Целью данной работы является анализ и повышение эффективности работы Международного института экономики и финансов. Разработана методология анализа качества показателей образовательного процесса.

МИЭФ мы рассматриваем как систему, состоящую из множества элементов. Индикатором качества работы этой системы является успеваемость студентов. Исследования этой характеристики работы факультета, а также факторов, на нее влияющих, проводятся в МИЭФ систематически [см., например, Yakovlev S., Zamkov O., (2013), О.О. Замков (2011), Lockshin J., Zamkov O. (2009), А.А. Пересецкий, М.А. Давтян, (2011)].

Эта работа отличается тем, что в ней представлен новый для сферы образования аппарат, основанный на Марковских цепях. Мы доказываем возможность применения этого аппарата к рейтингам студентов МИЭФ. Это позволяет нам делать прогнозы и использовать результаты этих прогнозов в работе факультета.

Аппарат, представленный в работе, дает возможность предвосхищать проблемы, с которыми сталкивается МИЭФ и повышает степень управляемости образовательного процесса.

Второй раздел работы посвящен анализу МИЭФ как системы и того, какое место в жизнедеятельности этой системы занимает успеваемость студентов. В этой части мы формулируем цели и показываем актуальность рассмотрения вопроса успеваемости для всех элементов системы.

В третьем разделе мы изучаем данные студенческих рейтингов с 2003 года. Проводится сравнительный анализ результатов студентов, поступивших в разные годы. Каждый год организационная политика руководства МИЭФ менялась, что могло повлиять на результаты студентов. Здесь мы тестируем гипотезы о том, значимы ли различия между результатами студентов в разные годы. Мы также изучаем динамику изменения результатов.

Четвертый раздел посвящен динамическому моделированию результатов студентов с 2003 года с помощью Марковских процессов. Мы тестируем данные на принадлежность к Марковским цепям, строим прогнозы и тестируем результаты.

В заключении представлены основные выводы и результаты работы.

Обзор литературы.

При написании этой работы я столкнулась с необходимостью ознакомится с теорией системного анализа. Без представления МИЭФ как системы, невозможно оценить действительную роль успеваемости студентов в деятельности факультета.

В учебной и научной литературе представлено множество различных методик проведения системного анализа. В данной работе мы не проводим полноценного системного анализа, а рассматриваем лишь начальные этапы для определения целей и движущей силы развития системы. В учебнике Волковой В.Н. и Денисова А.А. описаны основные методики системного анализа, среди которых методики Э. Квейда, С. Юнга, Ю.И. Черняка. В нашем анализе мы использовали методику последнего.

В третьей части работы мы тестируем гипотезы о значимости различий между результатами студентов, поступивших в соседние годы. Мы используем непараметрические тесты, так как не делаем предположений относительно вида распределений результатов. В статье Tomkins (2007) описаны различные непараметрические тесты и даются рекомендации относительно того, в каких ситуациях лучше применять тот или иной тест. По предложенной в этой статье схеме был выбран тест Колмогорова-Смирнова для двух выборок. Методика тестирования была изучена по работам Darling (1957) и Куликов Е.И. (2008). Примеры применения теста Колмогорова-Смирнова были изучены по статьям Clark (2008) и Kliygi Calli & Weverbergh (2009), которые помогли убедиться в применимости данного теста к нашим исследованиям.

В работе используются Марковские процессы первого порядка. Подробное описание используемой теории представлено в приложении 5, по материалам курсовой работы (Суншева А.) за второй курс. Основными источниками служили учебник Кельберт М.Я., Сухов Ю.М. (2010), Вентцель Е.С., Овчаров Л.А(2011).

В работе тестируются гипотезы о принадлежности рассматриваемых процессов к цепям Маркова. Литературы на тему тестирования Марковских процессов довольно много. Эти работы можно разделить на те, которые показывают тестирование гипотез через функцию максимального правдоподобия [Good (1955), Hoel (1954)] и те, которые используют статистику Хи-квадрат [Billingsley (1961)]. В работе Billingsley (1961) показано, что тест с использованием функции максимального правдоподобия идентичен тесту хи-квадрат. В своей работе мы применяем тест хи-квадрат, так как вычисления в этом случае менее громоздкие.

Различия между полученными прогнозами и фактическими данными мы тестируем с помощью теста Вилкоксона [Куликов Е.И. (2008)].

МИЭФ как система.

В середине девяностых годов, группа энтузиастов определила проблему – «создать в России такую экономическую образовательную программу, которая была бы конкурентоспособна на мировом уровне»[1]. Была проделана огромная работа, в результате которой появился и быстрыми темпами стал развиваться МИЭФ.

Сейчас МИЭФ – это институт:

- «преподаватели которого ведут исследования, чьи результаты получают международное признание, а также реализуют бакалаврскую программу двух дипломов в области экономики (пять специализаций) и магистерскую программу «Финансовая экономика»;
- который служит «точкой роста» НИУ-ВШЭ как исследовательского университета международного уровня;
- который за пятнадцать лет подготовил более 700 выпускников с дипломами двух ведущих вузов России и Англии, что не имеет прецедентов в России;
- где учатся около 650 студентов, работают более 120 преподавателей, шестнадцать из которых имеют степень PhD и семь наняты совместно с факультетом экономики ВШЭ.» [2]

Таким образом, главная цель создания МИЭФ достигнута. И теперь перед МИЭФ встают новые задачи.

МИЭФ - это саморазвивающаяся система. Глобальные цели сегодняшнего МИЭФ как системы сформулированы директором МИЭФ Сергеем Яковлевым в интервью газете «Академическая среда»:

«Среди стратегических задач:

- дальнейшее развитие МИЭФ как «центра академического совершенства»,
- консолидация его исследовательского потенциала для расширения спектра и повышения качества образовательных программ,
- содействие международному признанию НИУ-ВШЭ в мире»

Глобальная цель включает в себя подцели, которые выложены на сайте МИЭФ:

МИЭФ - научно-образовательный центр в сфере экономики, деятельность которого имеет международное признание, и

- Предлагает образовательные программы мирового уровня.
- Привлекает лучших абитуриентов (что подтверждается их дальнейшими успехами в освоении программы) с глобального рынка образования.
- Располагает составом преподавателей, ведущих научные исследования, результаты которых получают международное признание (публикации, выступления на конференциях).

[1] *Яковлев С.М., Академическая Среда, Окна Академического Роста Выпуск, посвященный пятнадцатилетию МИЭФ; № 07(12), 2012*

[2] *Яковлев С.М., Академическая Среда, Окна Академического Роста Выпуск, посвященный пятнадцатилетию МИЭФ; № 07(12), 2012*

- Обеспечивает предпосылки для высокого спроса на выпускников, как со стороны ведущих работодателей, так и со стороны ведущих университетов мира, а не только в своей стране.
- Играет заметную роль в национальном и международном сообществе, активно взаимодействует с внешней средой (бизнесом, государством, академическими кругами)

Миссия МИЭФ – сделать доступным студентам в России самое современное экономическое образование, сопоставимое с образованием ведущих университетов мира.[3]

Одна из основополагающих целей МИЭФ – привлечение лучших абитуриентов (и как следствие – в дальнейшем это лучшие студенты)

Почему большое количество абитуриентов последние годы стремится попасть в МИЭФ?

- Хотят получить качественное образование, признаваемое в мировых кругах, добиться успеха в трудовой и научно-исследовательской деятельности или продолжить образование в других зарубежных вузах.
- Хотят получить два диплома и хорошо оплачиваемую работу.
- Поступают в МИЭФ по настоянию родителей опять же для получения диплома и хорошо оплачиваемой работы.
- Боятся сдавать вступительные испытания (или проваливают их) в другие престижные вузы.

Почему лучший абитуриент не идет в МИЭФ?

- Не знает о нем.
- Не знает английский язык.
- Не хочет учиться (боится трудностей, ищет, где легче).
- Не хочет или не может оплачивать обучение.
- Не считает МИЭФ лучшим факультетом или поступает в другой вуз по настоянию родителей.

Повышение рейтинга вуза (по всем критериям) – единственный действенный рычаг для привлечения лучших абитуриентов. А успеваемость студентов – один из важнейших критериев оценки работы вуза, факультета.

Высокая успеваемость студентов вуза есть результат нескольких факторов:

- высокая образовательная база абитуриентов,
- высокий уровень преподавания
- высокая информационно-техническая оснащенность вуза
- эффективная организация учебного процесса
- мотивация студентов

3 Информация с официального сайта МИЭФ (icef.hse.ru).

В свою очередь высокая успеваемость студентов приводит к повышению спроса на выпускников на рынке труда и в научно-образовательном секторе. Повышается рейтинг вуза, и большой поток абитуриентов стремится стать студентами МИЭФ. Среди большего числа абитуриентов у Вуза появляется возможность выбрать лучших. То есть Вуз может улучшить успеваемость студентов за счет принятия на первый курс абитуриентов с высокой базовой подготовкой.

Из схемы видно, что при постоянстве прочих показателей, высокая успеваемость является двигателем развития системы по спирали. Понятно, что повышать успеваемость до бесконечности нельзя, но есть предел, к которому нужно стремиться. Для того, чтобы поддерживать жизнеспособность, конкурентоспособность МИЭФ нельзя опускаться ниже достигнутого уровня.

В МИЭФ руководство может влиять на успеваемость студентов через переподготовку преподавателей, перераспределение на группы [Yakovlev S., Zamkov O., (2013)] и увеличивая или уменьшая уровень набора на первый курс.

Уровень преподавания на МИЭФ дважды в год оценивается студентами. Кроме этого, на основе результатов внешних экзаменов, проводится эконометрический анализ, выявляющий влияние преподавания на отметки [Yakovlev S., Zamkov O., (2013)]. Преподавателей информируют о полученных результатах, и, при необходимости, проводятся специальные тренинги для лекторов и семинаристов.

Одной из часто тестируемых гипотез относительно влияния распределения студентов на группы является гипотеза о том, что результаты студента зависят от результатов одногруппников. То есть слабый студент в сильной группе покажет результаты лучше, чем в слабой группе. До поступления студент, как правило, имеет сформированные результатами ЕГЭ, родителями и школьными учителями ожидания того, на каком уровне он будет учиться. Распределение на группы с определением самой сильной и самой слабой групп может привести к несоответствию того, что студент от себя ожидал и того, в какую группу он попал. Это, в свою очередь, может влиять на то, как студент относится к учебе и на его показатели успеваемости.

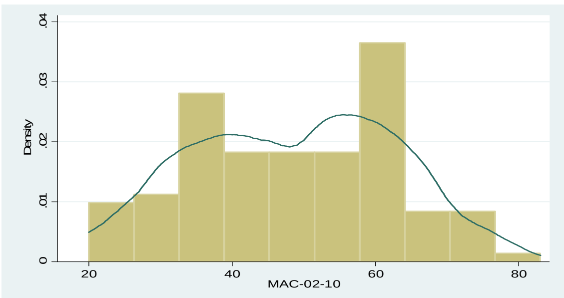
Политика МИЭФ по распределению студентов на группы и перераспределению в середине первого курса каждый год меняется. Исследования, проведенные в МИЭФ [Yakovlev S., Zamkov O., (2013), О.О. Замков (2011), Lockshin J., Zamkov O. (2009)] показали, что от перераспределения в январе выигрывают только сильные студенты, из чьих групп переводят слабых студентов.

Анализ успеваемости.

Для анализа успеваемости студентов мы используем данные итоговых годовых рейтингов студентов с 2003 года, доступные на сайте mief.hse.ru. Описательная статистика распределений оценок по всем предметам за период с 2003 по 2012 годы представлена в приложении 2. Средние показатели успеваемости студентов из года в год менялись по всем предметам. Эти изменения каждый год могли быть вызваны сразу несколькими факторами, например, другим уровнем базовых знаний поступивших абитуриентов, изменениями в преподавательском составе, в программе или методике, новой политикой распределения на группы. Все эти факторы вместе мы назовем политикой организации учебного процесса. Мы исследуем значимость различий эконометрическими методами, на основе чего можно будет сделать вывод, принесли ли результаты изменения в политике организации учебного процесса. Для тестирования различий результатов одного года поступивших от другого мы проводим аналогию с тестированием различий результатов в контрольной и экспериментальной группах. Определим показатели одного года как результаты «экспериментальной» группы, а показатели предшествующего года – как «контрольную» группу.

Для проведения анализа успеваемости были собраны и сгруппированы данные рейтингов МИЭФ по всем предметам студентов поступивших с 2003 по 2011 годы за все годы их обучения. К участию в рейтинге курса определенного года поступления принимаются оценки студентов закончивших 1 курс соответствующего года. В рейтинг 2 и последующих лет обучения включаются только оценки студентов из списка первого года обучения, и исключаются оценки студентов повторно изучающих курс из числа поступивших в более ранние годы.

На рисунке 1 проиллюстрированы распределения итоговых результатов по макроэкономике второго курса студентов, поступивших в 2010 году, по математическому анализу студентов, поступивших в 2011 году. (В приложении 3 представлены графики распределений основных предметов с 2007 по 2011 годы.) Эти распределения не имеют колоколообразную форму. Поскольку распределение, схожее с нормальным, имеют результаты лишь в нескольких годах по некоторым предметам (см. Приложение 3 - графики распределений основных предметов с 2007 по 2011 годы), количество студентов каждый год менялось, а оценки дисперсий значительно отличаются друг от друга (см. Приложение 2), мы не можем использовать стандартные t-тесты для парных выборок для сравнения этих распределений.



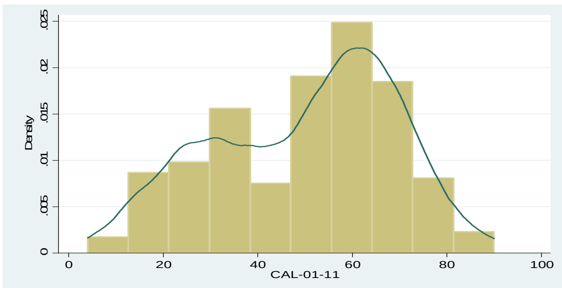


Рисунок 1*: Графики распределения результатов по макроэкономике второго курса для студентов, поступивших в 2010 году, и по математическому анализу для студентов, поступивших в 2011 году. Источник: расчеты автора.*

На рисунке 2 изображены ящичные диаграммы результатов по микроэкономике-1 для студентов, поступивших в 2009 и 2010 годах, и по математическому анализу для студентов, поступивших в 2010 и 2011 годах. На диаграмме видна неоднородность распределений результатов по этим предметам для поступивших в разные годы.

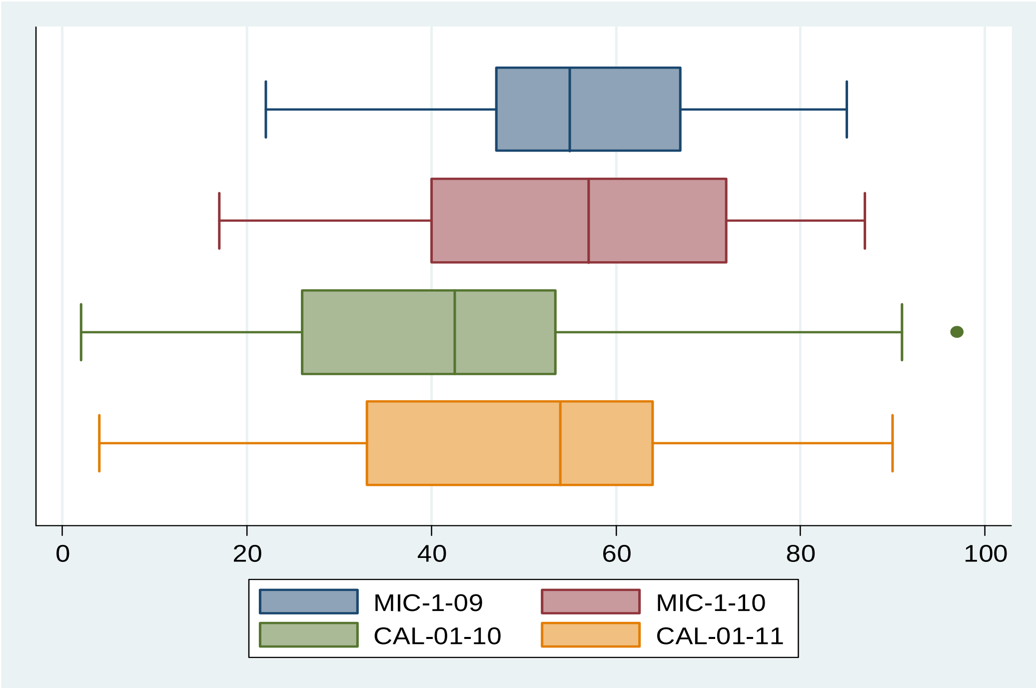


***Рисунок 2**: Диаграммы (box plot) распределения результатов по микроэкономике первого курса для студентов, поступивших в 2009 и 2010 годах, и по математическому анализу для студентов, поступивших в 2010 и 2011 годах. Источник: расчеты автора.*

Поскольку мы не можем использовать стандартные тесты, мы используем непараметрический тест Колмогорова-Смирнова для двух выборок. По сравнению с другими непараметрическими тестами этот тест имеет меньшую силу (power of the test), однако это вызвано тем, что тестируя различия между двумя распределениями, он тестирует все показатели распределения – не только различия в медианах [Куликов Е.И. 2008, стр. 253-257].

Мы тестируем нулевую гипотезу о том, что распределения итоговых оценок по одному и тому же предмету у разных курсов одинаковы.

$$H_0\text{: различия в распределениях незначительны}$$

Против альтернативной гипотезы:

$$H_a\text{: распределения значительно отличаются друг от друга}$$

Результаты проведенных тестов приведены ниже в Таблице 1.

Таблица 1.: Значения тестовой статистики Колмогорова-Смирнова и P-value. Знак (*) отображает значимость на пятипроцентном уровне, знак (**) – на однопроцентном уровне. Источник: расчеты автор.

Предметы 1 курса				
Предмет	Годы		d - статистика	p-value
Микроэкономика 1	2003	2004	0,0746	0,903
Микроэкономика 1	2004	2005	0,1009	0,548
Микроэкономика 1	2005	2006	0,0951	0,631
Микроэкономика 1	2006	2007	0,2335	0,003(**)
Микроэкономика 1	2007	2008	0,1605	0,058
Микроэкономика 1	2008	2009	0,1004	0,431
Микроэкономика 1	2009	2010	0,1621	0,029(*)
Микроэкономика 1	2010	2011	0,0926	0,357
Макроэкономика 1	2003	2004	0,0797	0,852
Макроэкономика 1	2004	2005	0,1363	0,2
Макроэкономика 1	2005	2006	0,0782	0,848
Макроэкономика 1	2006	2007	0,1857	0,028(*)
Макроэкономика 1	2007	2008	0,1817	0,02(*)
Макроэкономика 1	2008	2009	0,1004	0,431
Макроэкономика 1	2009	2010	0,1253	0,156
Макроэкономика 1	2010	2011	0,0856	0,456
Математический Анализ	2003	2004	0,0697	0,95
Математический Анализ	2004	2005	0,1395	0,191
Математический Анализ	2005	2006	0,1577	0,098
Математический Анализ	2006	2007	0,1269	0,28
Математический Анализ	2007	2008	0,1297	0,19
Математический Анализ	2008	2009	0,2449	0(**)
Математический Анализ	2009	2010	0,0946	0,462
Математический Анализ	2010	2011	0,2856	0(**)
Статистика 1	2003	2004	0,1678	0,072
Статистика 1	2004	2005	0,1721	0,049(*)
Статистика 1	2005	2006	0,2118	0,008(**)
Статистика 1	2006	2007	0,0507	0,998
Статистика 1	2007	2008	0,0738	0,838
Статистика 1	2008	2009	0,0836	0,668
Статистика 1	2009	2010	0,1407	0,08
Статистика 1	2010	2011	0,1402	0,039(*)
Общий рейтинг 1	2003	2004	0,119	0,253
Общий рейтинг 1	2004	2005	0,0935	0,678
Общий рейтинг 1	2005	2006	0,1476	0,149
Общий рейтинг 1	2006	2007	0,1356	0,225
Общий рейтинг 1	2007	2008	0,1254	0,237
Общий рейтинг 1	2008	2009	0,1421	0,1(*)
Общий рейтинг 1	2009	2010	0,1057	0,321
Общий рейтинг 1	2010	2011	0,0627	0,837

Предметы 2 курса				
Предмет	**Годы**		**d - статистика**	**p-value**
Микроэкономика 2	2003	2004	0,1397	0,288
Микроэкономика 2	2004	2005	0,0952	0,699
Микроэкономика 2	2005	2006	0,1592	0,148
Микроэкономика 2	2006	2007	0,1717	0,097
Микроэкономика 2	2007	2008	0,1228	0,335
Микроэкономика 2	2008	2009	0,2608	0(**)
Микроэкономика 2	2009	2010	0,2566	0,001(**)
Макроэкономика 2	2003	2004	0,1332	0,335
Макроэкономика 2	2004	2005	0,1353	0,262
Макроэкономика 2	2005	2006	0,3076	0(**)
Макроэкономика 2	2006	2007	0,2167	0,017(*)
Макроэкономика 2	2007	2008	0,1448	0,168
Макроэкономика 2	2008	2009	0,2506	0,01(*)
Макроэкономика 2	2009	2010	0,115	0,384
Математика для экономистов	2003	2004	0,1054	0,628
Математика для экономистов	2004	2005	0,105	0,575
Математика для экономистов	2005	2006	0,1768	0,081
Математика для экономистов	2006	2007	0,3022	0(**)
Математика для экономистов	2007	2008	0,1159	0,405
Математика для экономистов	2008	2009	0,1524	0,109
Математика для экономистов	2009	2010	0,2024	0,014(*)
Статистика 2	2003	2004	0,1873	0,064
Статистика 2	2004	2005	0,248	0,003(**)
Статистика 2	2005	2006	0,0729	0,951
Статистика 2	2006	2007	0,2311	0,009(**)
Статистика 2	2007	2008	0,0813	0,836
Статистика 2	2008	2009	0,2552	0,001(**)
Статистика 2	2009	2010	0,1062	0,487
Линейная Алгебра	2003	2004	0,1708	0,113
Линейная Алгебра	2004	2005	0,2457	0,004(**)
Линейная Алгебра	2005	2006	0,1511	0,2(*)
Линейная Алгебра	2006	2007	0,4125	0(**)
Линейная Алгебра	2007	2008	0,3105	0(**)
Линейная Алгебра	2008	2009	0,1408	0,169
Линейная Алгебра	2009	2010	0,0766	0,866
Общий рейтинг 2	2003	2004	0,186	0,221
Общий рейтинг 2	2004	2005	0,0963	0,685
Общий рейтинг 2	2005	2006	0,0805	0,899
Общий рейтинг 2	2006	2007	0,213	0,02(*)
Общий рейтинг 2	2007	2008	0,1179	0,383
Общий рейтинг 2	2008	2009	0,1649	0,067
Общий рейтинг 2	2009	2010	0,0531	0,996

Предметы 3 курса				
Предмет	**Годы**		**d - статистика**	**p-value**
Микроэкономика 3	2003	2004	0,1561	0,234
Микроэкономика 3	2004	2005	0,1505	0,268
Микроэкономика 3	2005	2006	0,1398	0,368
Микроэкономика 3	2006	2007	0,183	0,136
Микроэкономика 3	2007	2008	0,1934	0,075
Микроэкономика 3	2008	2009	0,1977	0,036(*)
Макроэкономика 3	2003	2004	0,0911	0,881
Макроэкономика 3	2004	2005	0,1529	0,255
Макроэкономика 3	2005	2006	0,0833	0,93
Макроэкономика 3	2006	2007	0,3448	0(**)
Макроэкономика 3	2007	2008	0,222	0,027(*)
Макроэкономика 3	2008	2009	0,0995	0,686
Эконометрика	2003	2004	0,1246	0,523
Эконометрика	2004	2005	0,1271	0,476
Эконометрика	2005	2006	0,1014	0,771
Эконометрика	2006	2007	0,2243	0,036(*)
Эконометрика	2007	2008	0,1516	0,262
Эконометрика	2008	2009	0,1058	0,612
Бухгалтерский учет	2003	2004	0,139	0,374
Бухгалтерский учет	2004	2005	0,141	0,353
Бухгалтерский учет	2005	2006	0,1304	0,481
Бухгалтерский учет	2006	2007	0,1312	0,506
Бухгалтерский учет	2007	2008	0,167	0,171
Бухгалтерский учет	2008	2009	0,0928	0,769
Общий рейтинг 3	2003	2004	0,1234	0,54
Общий рейтинг 3	2004	2005	0,1474	0,294
Общий рейтинг 3	2005	2006	0,1037	0,75
Общий рейтинг 3	2006	2007	0,1571	0,273
Общий рейтинг 3	2007	2008	0,1626	0,194
Общий рейтинг 3	2008	2009	0,1083	0,578

Предметы 4 курса				
Предмет	**Годы**		**d - статистика**	**p-value**
Корпоративные Финансы	2003	2004	0,1926	0,116
Корпоративные Финансы	2004	2005	0,2864	0,003(**)
Корпоративные Финансы	2005	2006	0,3406	0(**)
Корпоративные Финансы	2006	2007	0,4553	0(**)
Корпоративные Финансы	2007	2008	0,499	0(**)
Международная экономика	2003	2004	0,1857	0,257
Международная экономика	2004	2005	0,4078	0(**)
Международная экономика	2005	2006	0,3523	0,001(**)
Международная экономика	2006	2007	0,1295	0,627
Международная экономика	2007	2008	0,0852	0,946
Общий рейтинг 4	2003	2004	0,0908	0,914
Общий рейтинг 4	2004	2005	0,3051	0,001(**)
Общий рейтинг 4	2005	2006	0,3483	0(**)
Общий рейтинг 4	2006	2007	0,1922	0,128
Общий рейтинг 4	2007	2008	0,2289	0,029(*)

Итак, результаты студентов, поступивших в разные годы, имеют значимые отличия не по всем предметам. Практически все значимые отличия между годами можно назвать «случайными», так как случаи различия в результатах по предметам единичны. Однако, этого нельзя сказать о различиях в оценках между поступившими в 2006 и 2007 году. Тесты показали отличия распределений результатов студентов, поступивших в эти годы по микроэкономике 1, макроэкономике 1, макроэкономике 2, математике для экономистов, статистике 2, линейной алгебре, макроэкономике 3, эконометрике и корпоративным финансам. Также есть значимая разница в итоговых рейтингах.

***Таблица 2**: Описательная статистика результатов студентов, поступивших в 2006 и 2007 годах. Источник: расчеты автора.*

Переменная	Год	Среднее	#	Мин	25%	50%	75%	Макс
Микроэкономика 1	**2006**	60,16514	109	27	50	59	75	87
	2007	54,91964	112	20	45	52	66,5	83
Макроэкономика 2	**2006**	38,75862	87	6	29	39	48	67
	2007	35,3	100	9	24	33	44	75
Математика для экономистов	**2006**	71,91954	87	41	59	73	86	98
	2007	63,32	100	32	51,5	64	75,5	94
Статистика 2	**2006**	58,01149	87	21	42	59	69	96
	2007	53,59	100	20	40	50	62,5	96
Линейная алгебра	**2006**	68,44706	85	22	54	72	82	97
	2007	49,66327	98	14	36	49	64	83
Макроэкономика 3	**2006**	49,26667	75	11	38	48	63	83
	2007	38,87143	70	13	27	34,5	47	85
Эконометрика 3	**2006**	47,67568	74	20	35	45,5	61	82
	2007	40,62857	70	11	28	38	50	79
Корпоративные финансы	**2006**	51,54545	66	21	43	51,5	61	78
	2007	34,75758	66	0	25	33,5	44	71

Из таблицы 2 видно, что по всем предметам, где наблюдается значимая разница, средние показатели выше у студентов, поступивших в 2006 году. При этом по микроэкономике 1, математике для экономистов, линейной алгебре, эконометрике и корпоративным финансам минимумы максимумы 2006 года превышают минимумы и максимумы 2007 года. Обращаем внимание на то, что в оба года поступления с первого на второй курс не перешли 12 человек, но в 2007 году поступивших было больше, а значит, в процентном соотношении «выживших» в 2006 году было больше.

Тесты указывают на значимые различия результатов по корпоративным финансам с 2004 по 2008 годы. Средние показатели (Таблица 3) не проявляют какой-либо устойчивой динамики.

Таблица 3: *Описательная статистика результатов по корпоративным финансам с 2004 по 2008 годы. Источник: расчеты автора.*

Год	Среднее	#	Мин	25%	50%	75%	Макс
2004	53	74	18	40	53	66	92
2005	61,01408	71	22	51	62	73	86
2006	51,54545	66	21	43	51,5	61	78
2007	34,75758	66	0	25	33,5	44	71
2008	56,06173	81	6	45	55	71	95

Из выше изложенного можно сделать вывод, что изменения политики организации учебного процесса в 2007 году имели негативный эффект на результаты практически всех предметов.

Чтобы изучить дальнейшую динамику успеваемости я решила сравнить данные 2006 года со всеми следующими за «провальным» 2007 годами.

Еще раз тестируем нулевую гипотезу о том, что распределения итоговых оценок по одному и тому же предмету у разных курсов одинаковы.

$$H_0\text{: различия в распределениях незначительны}$$

Против альтернативной гипотезы:

$$H_a\text{: распределения значительно отличаются друг от друга}$$

На этот раз мы сравниваем распределения 2006 года против следующих за 2007 годом.

Результаты проведенных тестов приведены ниже в Таблице 4.

Таблица 4.: *Значения тестовой статистики Колмогорова-Смирнова и P-value. Знак (*) отображает значимость на пятипроцентном уровне (отвергаем нулевую гипотезу на 5% уровне значимости), знак (**) – на однопроцентном уровне. Источник: расчеты автора. Расчеты проводились на основе итоговых оценок по предметам.*

Предмет	годы		d-статистика	p-value
Микроэкономика 1	2006	2008	0,1419	0,129
	2006	2009	0,1364	0,174
	2006	2010	0,1556	0,06
	2006	2011	0,1233	0,191
Микроэкономика 2	2006	2008	0,0977	0,674
	2006	2009	0,2366	0,005(**)
	2006	2010	0,3516	0,000(**)
Макроэкономика 2	2006	2008	0,1236	0,371
	2006	2009	0,2612	0,001(**)
	2006	2010	0,0332	0,000(**)
Математика для экономистов	2006	2008	0,3391	0,000(**)
	2006	2009	0,2605	0,002(**)
	2006	2010	0,3465	0,000(**)
Статистика2	2006	2008	0,2577	0,002(**)
	2006	2009	0,1024	0,621
	2006	2010	0,0905	0,769
Линейная алгебра	2006	2008	0,1499	0,177
	2006	2009	0,1666	0,104
	2006	2010	0,1702	0,091
Макроэкономика 3	2006	2008	0,2028	0,048(*)
	2006	2009	0,2454	0,008(**)
Эконометрика	2006	2008	0,1481	0,272
	2006	2009	0,1534	0,224

При анализе значимых результатов тестирования мы видим, что по Микроэкономике2, Макроэкономике2, эта значимость отражает положительный сдвиг показателей. Статистика2, Математика для экономистов, Макроэкономика3 и Эконометрика3 хотя и ниже показателей 2006 года, но имеют динамику к улучшению показателей.

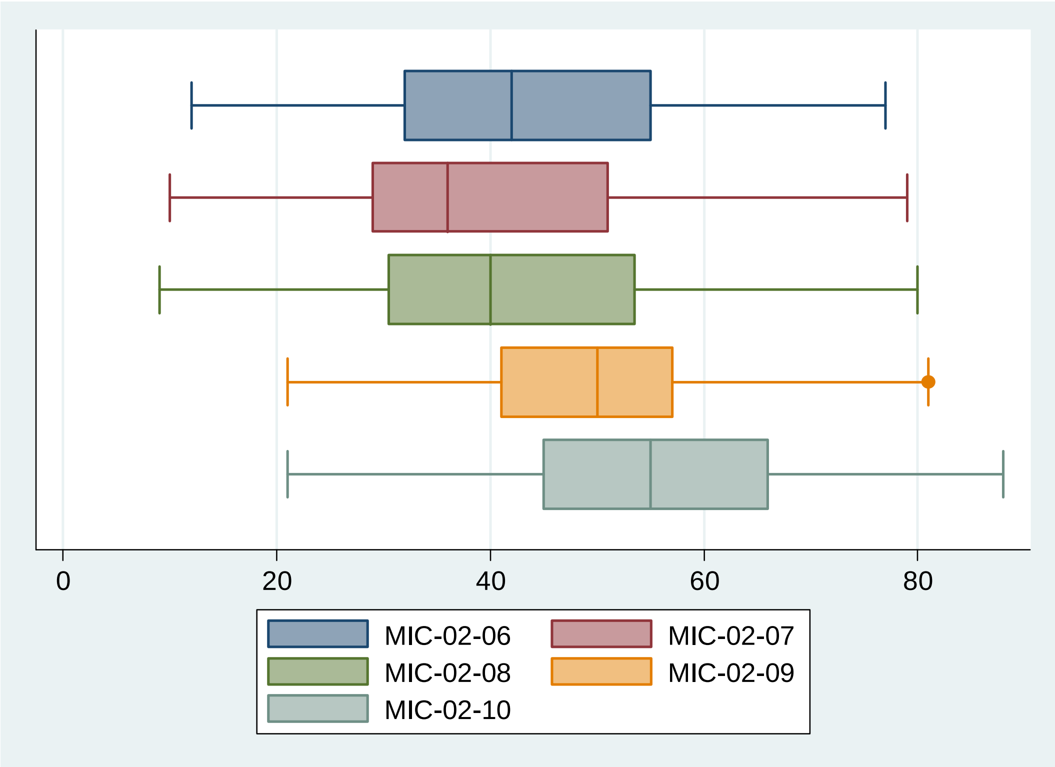


***Рисунок 3**: Диаграммы (box plot) распределения результатов по микроэкономике второго курса для студентов, поступивших в 2006 - 2010 годах. Источник: расчеты автора.*

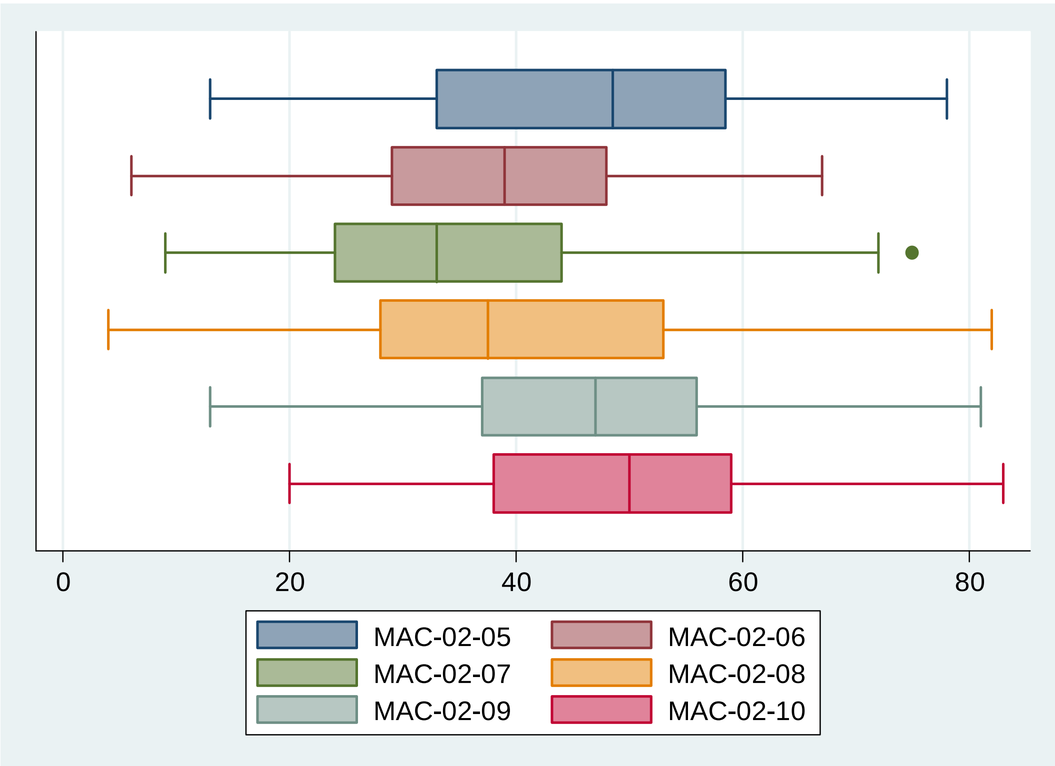


***Рисунок 4:** Диаграммы (box plot) распределения результатов по макроэкономике второго курса для студентов, поступивших в 2005 - 2010 годах. Источник: расчеты автора.*

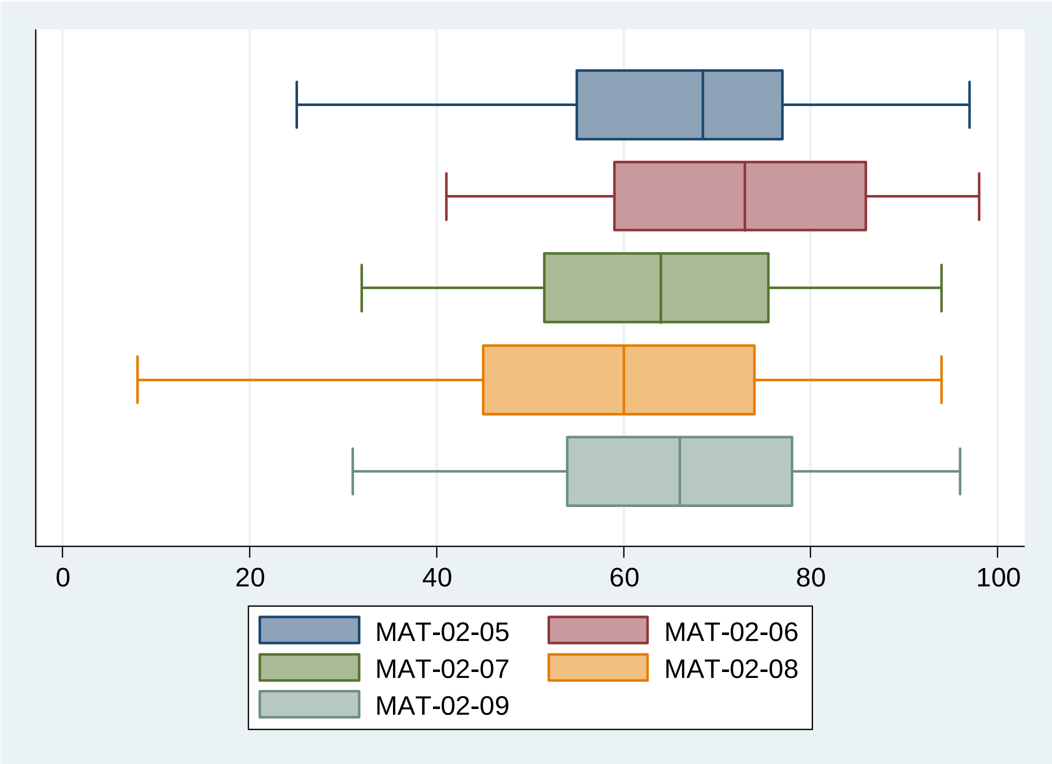


Рисунок 5: *Диаграммы (box plot) распределения результатов по математике для экономистов для студентов, поступивших в 2005 - 2009 годах. Источник: расчеты автора.*

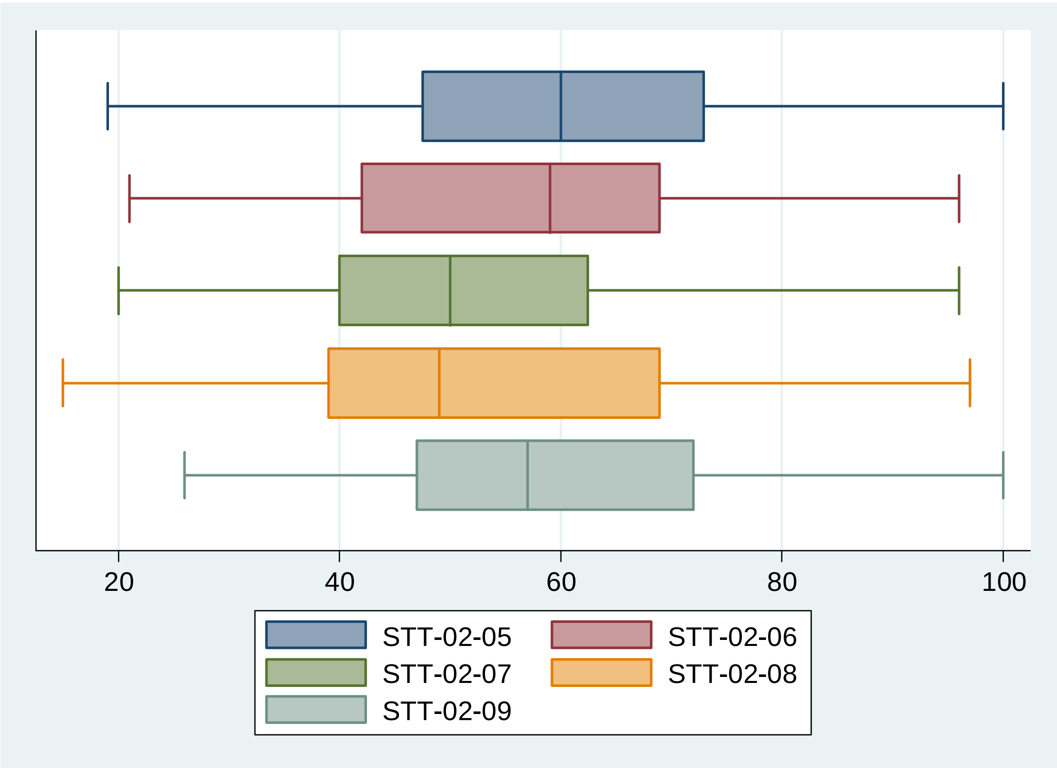


Рисунок 6: *Диаграммы (box plot) распределения результатов по статистике второго курса для студентов, поступивших в 2005 - 2009 годах. Источник: расчеты автора.*

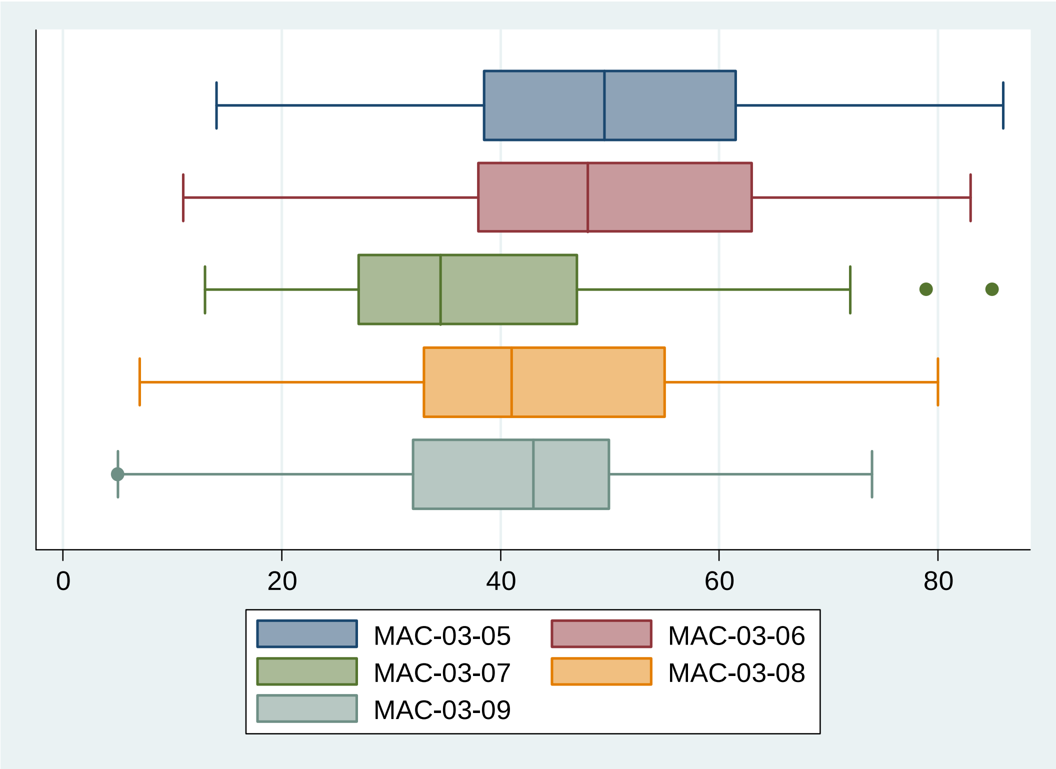


***Рисунок** 7: Диаграммы (box plot) распределения результатов по макроэкономике третьего курса для студентов, поступивших в 2005 - 2009 годах. Источник: расчеты автора.*

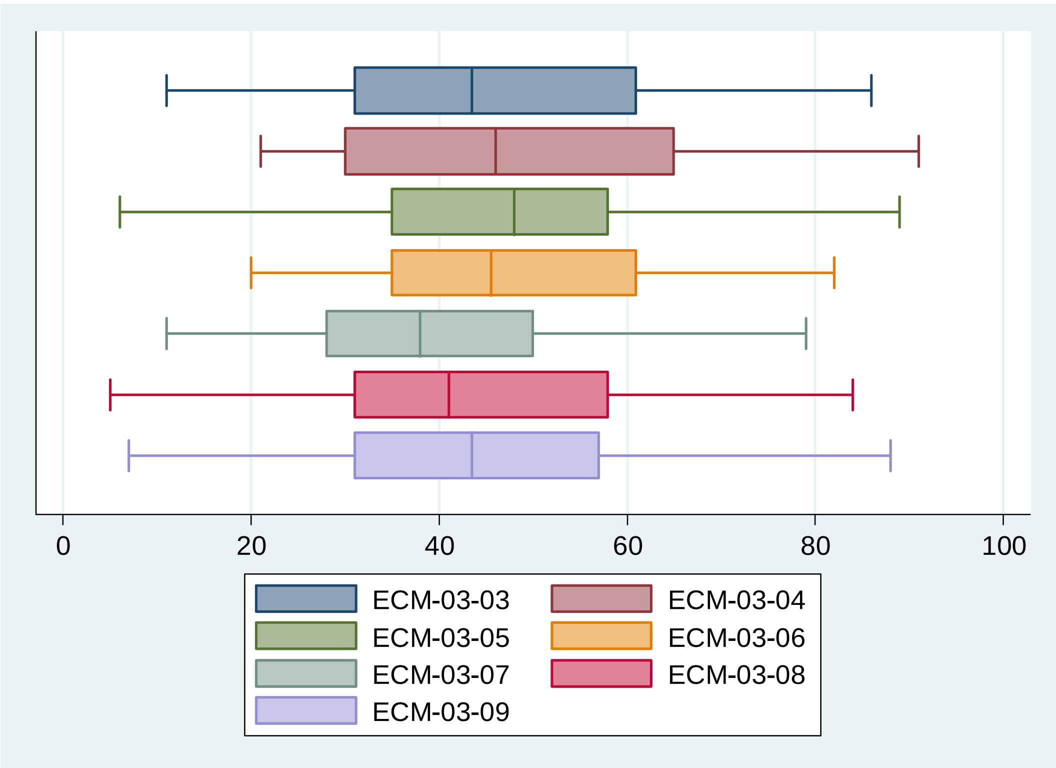


***Рисунок 8:** Диаграммы (box plot) распределения результатов по эконометрике для студентов, поступивших в 2003 - 2009 годах. Источник: расчеты автора.*

Анализ общего рейтинга за 3 и 4 курсы дает основания говорить о положительной динамике показателей, начиная с 2007года.

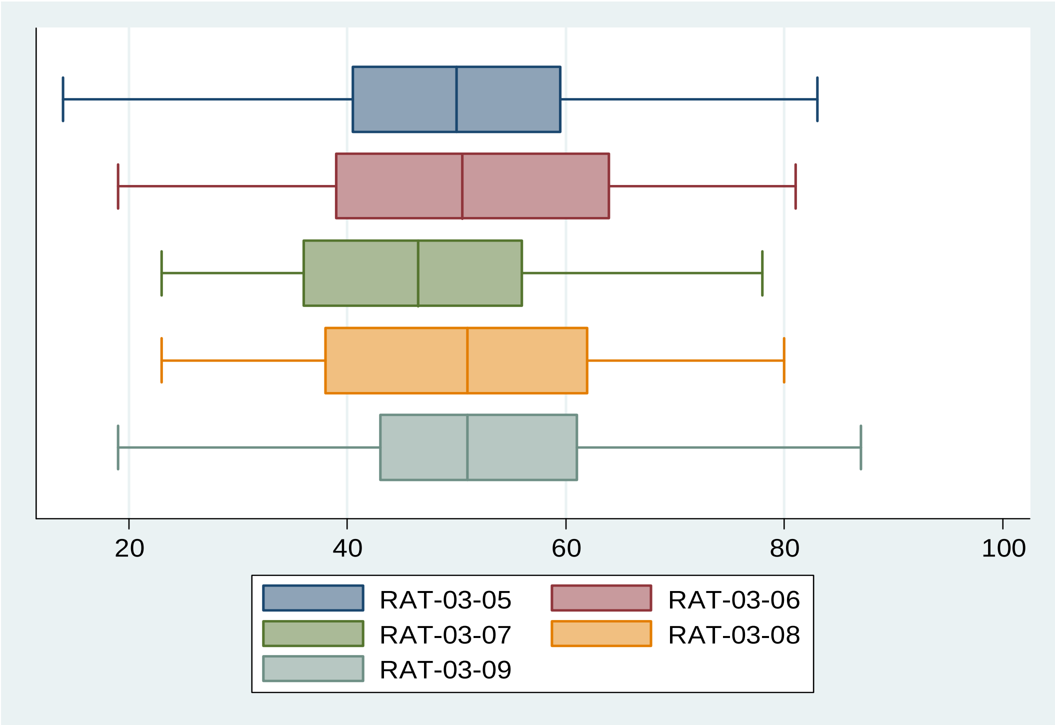


Рисунок 9: *Диаграммы (box plot) распределения итоговых рейтингов по окончании третьего курса студентов, поступивших в 2005 - 2009 годах. Источник: расчеты автора.*

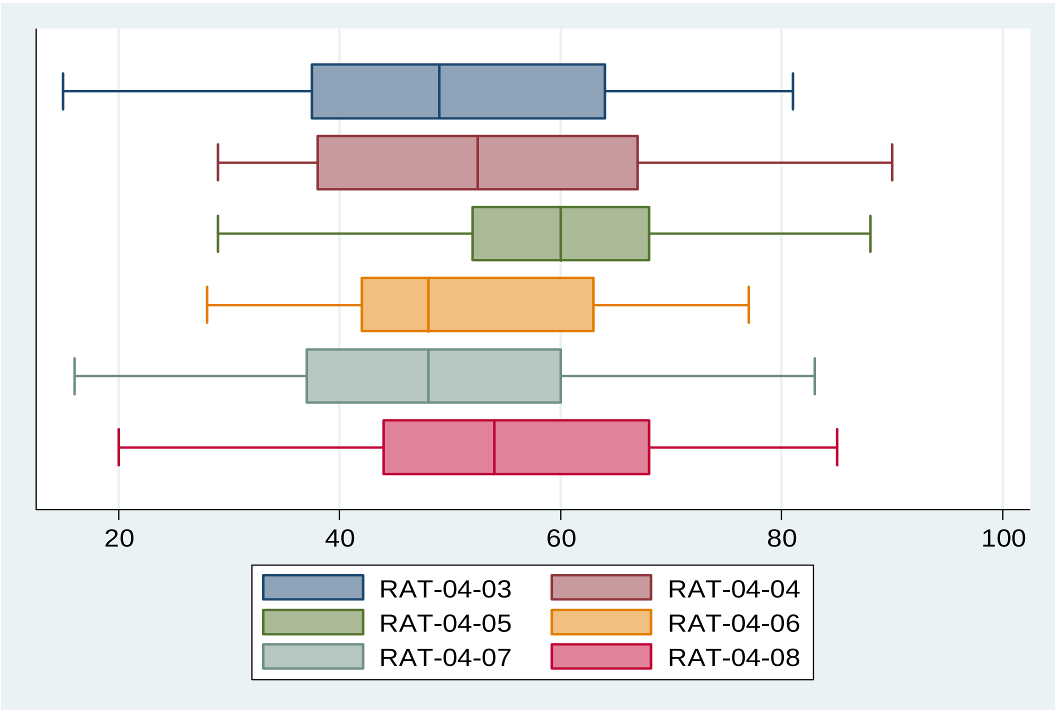


Рисунок 10: *Диаграммы (box plot) распределения итоговых рейтингов по окончании четвертого курса студентов, поступивших в 2003 - 2008 годах. Источник: расчеты автора.*

Итак, успеваемость студентов на протяжении всего изучаемого периода в среднем стабильна. Тесты показали значимые различия между поступившими в 2006 и 2007 годах, причем средние показатели 2007 года ниже. Однако, начиная с 2008 года наблюдается положительная динамика. На этом основании делается вывод о том, что в 2007 году была выбрана неэффективная организационная политика, к которой система постепенно адаптировалась, либо политика была скорректирована. Введение приема абитуриентов только по результатам ЕГЭ в 2008 году не оказало отрицательного влияния на успеваемость, как можно было предположить. По результатам поступивших в последние годы можно сделать вывод о том, что количество студентов не оказывает отрицательного влияния на успеваемость.

Теоретическое моделирование цепи Маркова.[4]

Марковские процессы занимают особое место в теории случайных процессов. С их помощью можно моделировать различные ситуации в динамике.

Теорию Марковских процессов активно используют в самых разных областях науки, от молекулярной физики до лингвистики.

Существуют разные способы классификации случайных процессов в целом и Марковских процессов в частности. Основное разделение проводится по типам времени и состояний процессов – дискретным или непрерывным. В данной работе мы используем Марковские процессы с дискретным временем и дискретными состояниями, известные как Марковские цепи или цепи Маркова.

Основное отличие Марковских цепей от всех других процессов заключается в том, что вероятность каждого состояния системы в момент времени t+1 зависит только от состояния системы в момент t. У простой цепи Маркова конечное число состояний, и переход из состояния в состояние происходит в определенные моменты времени. Именно такой процесс мы используем для прогнозирования успеваемости студентов.

Итак, цепью Маркова называется Марковский случайный процесс с дискретным временем, в котором все возможные состояния i = 1,2,…n можно заранее перечислить, а переход из состояния в состояние происходит мгновенно (скачком), но только в определенные моменты времени, называемыми шагами процесса.

Случайный механизм, определяющий переход системы из состояния в состояние называется матрицей перехода, состоящей из вероятностей перехода p_{ij} из состояния i в состояние j. Матрица перехода P называется стохастической, или вероятностной.

Таким образом pij – это условная вероятность того, что система будет находиться в состоянии j в следующий момент, при условии что в данный момент она находится в состоянии i. значит все элементы P неотрицательны, но не превышают 1, и сумма элементов в любой строке равна 1:

$$0 \leq p_{ij} \leq 1 \;\; \forall i, j \in I$$

$$\sum_{j \in I} p_{ij} = 1 \;\; \forall i \in I$$

Матрица P, обладающая такими свойствами называется стохастической, т. е. вероятностной.

---

4 Этот раздел составлен по материалам курсовой работы за второй курс (Суншева А.) и является кратким конспектом теории Марковских цепей, составленным по учебникам Вентцель Е.С., Овчаров Л.А., «Теория случайных процессов и её инженерные приложения», Москва, 2011 и Кельберт М.Я., Сухов Ю.М., «Марковские цепи как отправная точка теории случайных процессов и их приложения», Москва, 2010.

Простейший случай имеет вид 2x2 (пространство из 2 состояний). Можно считать, что состояниями являются 0 и 1. Тогда элементы матрицы имеют вид p_{ij}, *i,j*=0,1, а стохастическую матрицу можно представить в виде:

$$\begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$$

где $0 \leq \alpha, \beta \leq 1$.

В частности при $\alpha = \beta = 0$ получаем единичную матрицу, а при $\alpha = \beta = 1$ – антидиагональную матрицу:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Система с единично матрицей остается в начальном состоянии навсегда, а в антидиагональном случае она меняет состояние в каждый момент времени, переходя из 0 в 1 и обратно.

С другой стороны при $\alpha = \beta = \frac{1}{2}$ мы получаем матрицу

$$\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

В этом случае система может либо остаться в том же состоянии, либо поменять его с вероятностью ½.

Мы приводим формальное определение цепи Маркова:

«Говорят, что последовательность случайных величин X_n со значениями в конечном или счетном множестве I образует цепь Маркова с дискретным временем с начальным распределением λ и матрицей перехода P, если $\forall i_0, \ldots, i_n \in I$ совместное распределение $P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n)$ равно $\lambda_{i_0} p_{i_0 i_1} \ldots p_{i_{n-1} i_n}$» [Кельберт М.Я., Сухов Ю.М., 2010, стр.15].

Одним из важнейших результатов теории Марковских цепей является то, что вероятность того, что система окажется в состоянии j, при условии, что сейчас она находится в состоянии i, можно рассчитать, зная исходное распределение λ:

$$P\ (X_k = i, X_{n+k} = j) = (\lambda P^k)_i (P^n)_{ij} \qquad (1)$$

и

$$P(X_{k+n} = j | k = i) = \frac{P(X_k = i, X_{k+n} = j)}{P(X_k = i)} = \frac{(\lambda P^k)_i (P^n)_{ij}}{(\lambda P^k)_i} = (P^n)_{ij} \qquad (2)$$ [5]

5 Подробный вывод формул (1) и (2) см. в Приложении 6.

На основании рейтинга студенты разделяются на n групп. Обозначим балл студента по рейтингу за X, Рейтинг $\mathcal{R}$ – множество действительных чисел от 0 до 100 - делится на n-1 непересекающихся подмножеств Δ_i таких, что $\bigcup_i \Delta_i = \mathcal{R}$. В k-ю группу входят все студенты прошедшие аттестацию с баллом рейтинга $X \epsilon \Delta_k$, в n-ю группу относим студентов не прошедших аттестацию или покинувших курс по разным причинам. В каждой группе у нас A_k студентов, k=1,2,…,n.

Под состоянием системы мы будем понимать определенное распределение студентов по группам, в зависимости от рейтинга. Тогда наша система в начальный момент времени t=0 выглядит в виде столбца:

$$S^0=\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{bmatrix}$$

Таким образом наша система состоит из распределения студентов по n группам, и изменение состояния системы в моменты времени t=1,2,3 – связано с переходом студентов из одной группы в другую.

Понятно, что такая система может принимать конечное число состояний. Система может менять свои состояния в определенные моменты времени t. Предполагается, что состояние системы в момент времени t зависит только от состояния системы в момент времени t-1 (при t=0 состояние системы по итогам первого курса, t=1 – состояние системы по итогам второго курса и т.д. Считаем, что состояние системы в текущем году зависит только от её состояния в предыдущий год), то есть мы определяем данный процесс изменения системы как цепь Маркова.

Составляем матрицу перехода для следующего курса:

$$A^1=\begin{bmatrix} a_{11}^1 & \cdots & a_{1n}^1 \\ \vdots & \ddots & \vdots \\ a_{n1}^1 & \cdots & a_{nn}^1 \end{bmatrix},$$

где a_{ij} - это количество человек, которые были в i-й группе для t=0 и попали в j-ю группу в момент t=1(т.е. после второго курса). Соответственно система примет следующее значение:

$$S^1=\begin{bmatrix} A_1^1 = \sum_i a_{i1}^1 \\ A_2^1 = \sum_i a_{i2}^1 \\ \vdots \\ A_n^1 = \sum_i a_{in}^1 \end{bmatrix}$$

Аналогично для третьего курса:

$$A^2=\begin{bmatrix} a_{11}^2 & \cdots & a_{1n}^2 \\ \vdots & \ddots & \vdots \\ a_{n1}^2 & \cdots & a_{nn}^2 \end{bmatrix}$$

где a_{ij} - это количество человек, которые были в i-й группе для t=0 и попали в j-ю группу в момент t=2 (после третьего курса). И система примет вид

$$S^2=\begin{bmatrix} A_1^2 = \sum_i a_{i1}^2 \\ A_2^2 = \sum_i a_{i2}^2 \\ \vdots \\ A_n^2 = \sum_i a_{in}^2 \end{bmatrix}$$

Находим вероятностную (стохастическую) матрицу перехода для второго курса:

$$P=\begin{bmatrix} a_{11}^1/A_1 & a_{12}^1/A_1 & \cdots & a_{1n}^1/A_1 \\ a_{21}^2/A_2 & a_{22}^1/A_2 & \cdots & a_{2n}^1/A_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}^1/A_n & a_{n2}^1/A_n & \cdots & a_{nn}^1/A_n \end{bmatrix}$$

Возведением матрицы P в квадрат мы получим матрицу P^2 прогнозированных вероятностей, каждый элемент p_{ij} которой будет показывать вероятность перехода из состояния i (t=0) в состояние j(t=2).

Полученную матрицу можно перевести в реальные значения, помножив каждый элемент на соответствующее ему количество студентов в группе (в момент времени t=0).

В результате имеем матрицу, каждый элемент которой показывает прогнозируемое количество студентов перешедших из i-й группы в j-ю.

Мы можем прогнозировать состояние системы в период t=3, перемножив стохастическую матрицу P2 и вектор исходного распределения. Аналогично, можно спрогнозировать состояние в момент времени t=4, перемножив стохастическую матрицу P3 и вектор исходного распределения.

Моделирование успеваемости как Марковской цепи.

Мы моделируем успеваемость студентов на основе итоговых годовых рейтингов студентов с 2003 по 2012 годы, доступных на сайте mief.hse.ru. Для всех курсов, поступивших с 2003 года, есть 4 итоговых рейтинга за четыре года обучения на бакалавриате. Рейтинги студентов мы разделяем на пять групп, которые являются состояниями в нашей цепи Маркова. Это значит, что мы определяем переход студента из одной группы в другую по результатам каждого учебного года, как систему, изменения состояний которой описываются цепью Маркова.

Разделение на группы происходит следующим образом:

1 группа – рейтинг выше 80 баллов;

2 группа – рейтинг выше 60, но меньше 80 баллов;

3 группа – рейтинг выше 40, но меньше 60 баллов;

4 группа – рейтинг выше 20, но меньше 40 баллов;

5 группа – рейтинг меньше 20 баллов.

Мы обозначаем состояния S_i, i = 1,2,…,5. Граф перехода системы из состояния в состояние представлен на рисунке 3. Стрелки обозначают возможные переходы из состояния в состояние.

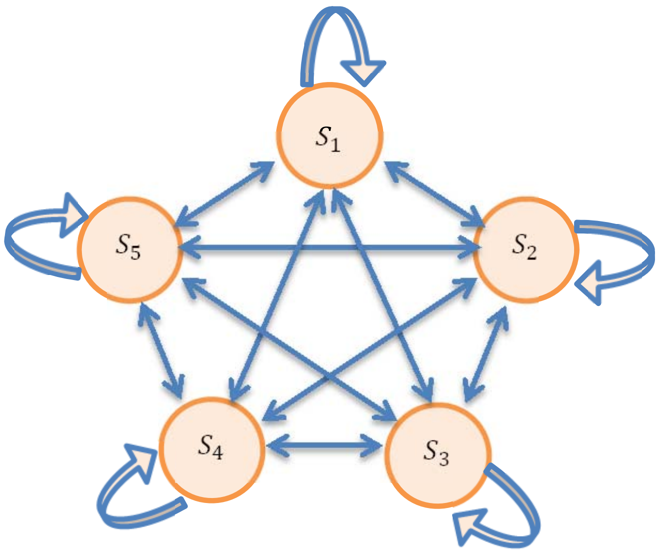


***Рисунок 3**: Граф состояний системы перехода студентов из группы в группу.*

Рассмотрим систему на примере студентов, поступивших в 2006 году. Вектор исходных состояний показывает количество студентов, оказавшихся в том или ином состоянии по итогам первого года обучения. Мы обозначаем вектор исходных состояний S_0.

Для 2006 года поступления, $S_0 = \begin{bmatrix} 2 \\ 37 \\ 48 \\ 16 \\ 8 \end{bmatrix}$, что означает, что в 1 группе было два студента, во второй – 37 студентов и так далее. Зная рейтинг студентов после второго года обучения, мы можем составить матрицу перехода, показывающую, сколько студентов перешло из одной группы в другую. Матрицу перехода мы обозначаем A_1:

$$A_1 = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 25 & 12 & 0 & 0 \\ 0 & 1 & 25 & 12 & 10 \\ 0 & 0 & 2 & 3 & 11 \\ 0 & 0 & 0 & 3 & 5 \end{bmatrix}$$

Каждый член матрицы A_1 отображает сколько человек перешли из группы i в группу j, где i исчисляет строки, а j – столбцы. Используя исходный вектор состояний S_0 и матрицу перехода A_1, мы можем вычислить стохастическую матрицу перехода (см. Приложение 5), P_1:

$$P_1 = \begin{bmatrix} 0{,}5 & 0{,}5 & 0 & 0 & 0 \\ 0 & 0.676 & 0.324 & 0 & 0 \\ 0 & 0{,}021 & 0{,}521 & 0{,}25 & 0{,}208 \\ 0 & 0 & 0{,}125 & 0{,}188 & 0{,}687 \\ 0 & 0 & 0 & 0{,}375 & 0{,}625 \end{bmatrix}$$

Стохастическая матрица перехода показывает вероятности перехода из состояния в состояние. Например, значение в ячейке, где *i*=2, *j*=1, равно 0,056, что значит, что после второго курса вероятность перейти из второй группы в первую равна 5,6%. Стохастические матрицы перехода были просчитаны для всех доступных лет – с 2003 по 20011.

Мы обращаем внимание, что значения в матрице P_1 являются оцененными вероятностями. Оценки считаются через функцию максимального правдоподобия [Anderson and Goodman, 1957].

Тестирование гипотезы о принадлежности к цепи Маркова первого порядка.

Мы тестируем процесс перехода из одной группы в другую на принадлежность к Марковским процессам. За нулевую гипотезу мы принимаем «случайные» вероятности [Billingsley (1961)] в стохастической матрице. То есть переходы из состояния в состояние являются процессом без памяти, и переход в следующее состояние не зависит от того, в каком состоянии система находится сейчас. Альтернативной гипотезой является то, что система является Марковской цепью первого порядка, то есть состояние в следующем периоде зависит от состояния в настоящий момент времени, но не зависит от того, в каком состоянии система была до этого. Поскольку мы предполагаем, что наш процесс, является именно цепью Маркова первого порядка, такая альтернативная гипотеза нас устраивает. В тестах используем тестовую статистику из Billingsley (1961) и Hoel (1954).

Итак, нулевая гипотеза: H_0: процесс является цепью Маркова без памяти

Альтернативная гипотеза: H_{a}: процесс является цепью Маркова первого порядка

Тестовой статистикой является Хи-квадрат статистика:

$$\sum_{ij} \frac{\left(f_{ij} - \frac{f_{i\cdot} \cdot f_{\cdot j}}{n}\right)^2}{\left(\frac{f_{i\cdot} \cdot f_{\cdot j}}{n}\right)}$$

где f_{ij} – значения матрицы перехода (не стохастической);

$f_{i\cdot}$ – суммы значений по строкам матрицы перехода;

$f_{\cdot j}$ – суммы значений по столбцам матрицы перехода;

n – сумма всех значений матрицы перехода.

При нулевой гипотезе, данная статистика распределена по Хи – квадрат распределению с $(k-1)^2$ степенями свободы, где k – количество состояний в системе. [Подробный вывод см. Dawson and Good (1957)].

Мы тестируем гипотезу для всех лет с 2003 по 2010. Мы считаем тестовую статистику и сравниваем с критическим значением с $(5-1)^2 = 16$ степенями свободы: $\chi^2(16) \approx 31.9$ для уровня значимости 1%.

В таблице 5 представлены тестовые статистики для каждого года:

Таблица 5*: Тестовые статистики для гипотезы процессов без памяти для 2003-2010 годов. Источник: расчеты автора.*

Год	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
Статистика	141,34	120,59	172,98	148,36	91,64	160,38	213,31	220,02

Таким образом, мы отвергаем нулевую гипотезу в пользу альтернативной. То есть используемый нами процесс можно определить как цепь Маркова первого порядка.

Прогнозирование и анализ результатов.

Итак, мы с уверенностью можем сказать, что изучаемый процесс принадлежит к цепям Маркова первого порядка. Это позволяет нам использовать аппарат данных процессов для прогнозирования. Мы имеем вектор исходного распределения и стохастическую матрицу. Возведя матрицу P_1в квадрат, мы получаем стохастическую матрицу P_2 (Кельберт М.Я, Сухов Ю.М., 2010, стр. 12-15):

$$P_2 = \begin{bmatrix} 0{,}25 & 0{,}588 & 0{,}162 & 0 & 0 \\ 0 & 0{,}463 & 0{,}388 & 0{,}081 & 0{,}068 \\ 0 & 0{,}025 & 0{,}309 & 0{,}255 & 0{,}411 \\ 0 & 0{,}003 & 0{,}089 & 0{,}324 & 0{,}584 \\ 0 & 0 & 0{,}047 & 0{,}305 & 0{,}648 \end{bmatrix}$$

Эта матрица показывает вероятности перехода из группы на первом курсе в группу на третьем курсе. То есть значение в ячейке, где *i*=2, *j*=1, равное 0,089, показывает, что вероятность перейти из второй группы по итогам первого курса в первую по итогам третьего курса составляет 8,9%.

Перемножив матрицу P_2и исходный вектор распределений S_0, мы получаем прогноз распределений студентов по группам по итогам третьего курса, S_2^{exp}:

$$S_2^{exp} = \begin{bmatrix} 1 \\ 20 \\ 31 \\ 22 \\ 37 \end{bmatrix}$$

Это значит, что прогнозируемое количество студентов в первой группе, то есть тех, чей рейтинг превышает 80 баллов, составляет 1 человек. 20 человек – во второй группе, и т.д.

Аналогично, мы можем прогнозировать вектор распределений для четвертого курса. Для этого мы возводим стохастическую матрицу перехода в третью степень и умножаем ее на исходный вектор распределений. В результате получаем вектор S_3^{exp}:

$$S_3^{exp} = \begin{bmatrix} 0 \\ 14 \\ 26 \\ 26 \\ 45 \end{bmatrix}$$

То есть прогнозируемое количество студентов в первой группе по итогам четвертого курса составляет 0 человек, а в пятой – 45 человек.

Для поступивших в 2006 году известны рейтинги после каждого года обучения, а значит мы можем сравнить прогнозы, основанные на цепях Маркова, с реальными данными. Вектор фактических значений S_2 для третьего года обучения и S_3 для четвертого года обучения:

$$S_2 = \begin{bmatrix} 1 \\ 21 \\ 32 \\ 19 \\ 38 \end{bmatrix} \quad S_3 = \begin{bmatrix} 0 \\ 21 \\ 35 \\ 10 \\ 45 \end{bmatrix}$$

Вычисляем векторы разниц: $D_2 = S_2 - S_2^{exp}$ и $D_3 = S_3 - S_3^{exp}$:

$$D_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ -3 \\ 1 \end{bmatrix} \quad D_3 = \begin{bmatrix} 0 \\ 7 \\ 9 \\ -16 \\ 0 \end{bmatrix}$$

Итак, мы можем прогнозировать распределения студентов на год вперед и на два года вперед.

Мы провели расчеты матриц перехода и стохастических матриц перехода для всех курсов, поступивших с 2003 года, и сделали прогноз, который сравнили с доступными данными. Результаты всех расчетов представлены в таблице 6:

***Таблица 6**: Результаты прогнозов и фактические данные с 2003 по 2009 годы. Источник: расчеты автора.*

2003	$S_2^{exp} = \begin{bmatrix} 4 \\ 17 \\ 5 \\ 3 \\ 45 \end{bmatrix}$	$S_2 = \begin{bmatrix} 1 \\ 17 \\ 31 \\ 15 \\ 39 \end{bmatrix}$	$S_3^{exp} = \begin{bmatrix} 5 \\ 12 \\ 30 \\ 2 \\ 54 \end{bmatrix}$	$S_3 = \begin{bmatrix} 2 \\ 19 \\ 24 \\ 17 \\ 41 \end{bmatrix}$
2004	$S_2^{exp} = \begin{bmatrix} 8 \\ 24 \\ 28 \\ 25 \\ 29 \end{bmatrix}$	$S_2 = \begin{bmatrix} 3 \\ 22 \\ 24 \\ 29 \\ 35 \end{bmatrix}$	$S_3^{exp} = \begin{bmatrix} 9 \\ 16 \\ 20 \\ 23 \\ 45 \end{bmatrix}$	$S_3 = \begin{bmatrix} 3 \\ 28 \\ 22 \\ 21 \\ 39 \end{bmatrix}$
2005	$S_2^{exp} = \begin{bmatrix} 7 \\ 25 \\ 29 \\ 37 \\ 19 \end{bmatrix}$	$S_2 = \begin{bmatrix} 1 \\ 18 \\ 41 \\ 17 \\ 41 \end{bmatrix}$	$S_3^{exp} = \begin{bmatrix} 8 \\ 17 \\ 22 \\ 40 \\ 30 \end{bmatrix}$	$S_3 = \begin{bmatrix} 4 \\ 33 \\ 27 \\ 7 \\ 47 \end{bmatrix}$
2006	$\mathrm{S}_2^{\mathrm{exp}} = \begin{bmatrix} 1 \\ 20 \\ 31 \\ 23 \\ 37 \end{bmatrix}$	$\mathrm{S}_2 = \begin{bmatrix} 1 \\ 21 \\ 32 \\ 19 \\ 38 \end{bmatrix}$	$\mathrm{S}_3^{\mathrm{exp}} = \begin{bmatrix} 0 \\ 14 \\ 26 \\ 26 \\ 45 \end{bmatrix}$	$\mathrm{S}_3 = \begin{bmatrix} 0 \\ 21 \\ 35 \\ 10 \\ 45 \end{bmatrix}$
2007	$\mathrm{S}_2^{\mathrm{exp}} = \begin{bmatrix} 2 \\ 11 \\ 43 \\ 38 \\ 25 \end{bmatrix}$	$\mathrm{S}_2 = \begin{bmatrix} 0 \\ 13 \\ 31 \\ 26 \\ 48 \end{bmatrix}$	$\mathrm{S}_3^{\mathrm{exp}} = \begin{bmatrix} 2 \\ 7 \\ 36 \\ 43 \\ 30 \end{bmatrix}$	$\mathrm{S}_3 = \begin{bmatrix} 1 \\ 15 \\ 30 \\ 18 \\ 54 \end{bmatrix}$

2008	$S_2^{exp} = \begin{bmatrix} 3 \\ 22 \\ 50 \\ 31 \\ 39 \end{bmatrix}$	$S_2 = \begin{bmatrix} 0 \\ 24 \\ 41 \\ 26 \\ 54 \end{bmatrix}$	$S_3^{exp} = \begin{bmatrix} 2 \\ 15 \\ 45 \\ 36 \\ 46 \end{bmatrix}$	$S_3 = \begin{bmatrix} 4 \\ 27 \\ 40 \\ 9 \\ 65 \end{bmatrix}$
2009	$S_2^{exp} = \begin{bmatrix} 4 \\ 41 \\ 49 \\ 9 \\ 30 \end{bmatrix}$	$S_2 = \begin{bmatrix} 2 \\ 25 \\ 51 \\ 15 \\ 40 \end{bmatrix}$	$S_3^{exp} = \begin{bmatrix} 5 \\ 40 \\ 43 \\ 7 \\ 38 \end{bmatrix}$	

Тестирование гипотезы о значимости различий между фактическими и спрогнозированными значениями.

Итак, у нас есть результаты прогнозов, сделанных с помощью цепей Маркова и фактические данные. Мы заинтересованы в том, чтобы понять, значимы ли различия, то есть можно ли доверять прогнозам, сделанным на основе оцененных стохастических матриц. Поскольку мы не делаем никаких предположений относительно распределения студентов на группы, то не можем проводить стандартных t-тестов. Проводим непараметрический тест Вилкоксона [Куликов 2008, стр. 263-265].

Итак, нулевая гипотеза:

$$H_0: \text{спрогнозированные и фактические оценки отличаются не начимо}$$

Альтернативная гипотеза:

$$H_{\text{a}}: \text{спрогнозированные и фактические оценки значимо отличны друг от друга}$$

Проводим отдельные тесты для каждого года поступления. Результаты представлены в таблице 7:

***Таблица** 7: P-value для теста Вилкоксона. Источник: расчеты автора.*

Год	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
P-value	0.4716	0.9303	0.7454	0.5149	0.1094	0.3124	0.4375

Мы делаем вывод о том, что спрогнозированные распределения не значимо отличаются от фактических для всех рассматриваемых лет поступлений.

Этот результат позволяет нам использовать весь аппарат цепей Маркова для прогнозирования рейтингов. Эта методика дает возможность представлять изменения, которые только произойдут, а значит, и влиять на эти изменения.

Корректировка результатов.

Несмотря на незначимые различия в спрогнозированных и фактических распределениях, подтвержденных тестами, мы хотели приблизить результаты прогнозов к реальным данным. Для этого мы попытались найти матрицу, умножение на которую стохастической матрицы, давало бы новую скорректированную матрицу, дающую более приближенные к фактическим результаты.

Новая матрица должна быть стохастической, то есть сумма значений по строкам должна быть равна единице. Конструирование такой матрицы не составляет проблемы. Мы считаем стохастическую матрицу, исходя из реальных данных, показывающую вероятности перехода из группы после 1го года обучения в группу после 3го года обучения. Эту матрицу обозначаем R. Нужно найти такую матрицу X, что

$$R = X \cdot P_2$$

Где P_2 – это рассчитанная стохастическая матрица перехода на третий курс.

$$X = R \cdot P_2^{-1}$$

Однако, мы хотим построить аналог матрицы X для всех лет, то есть умножение на такую матрицу стохастической матрицы любого года должно давать скорректированную матрицу, улучшающую прогноз. Но нет такой матрицы X, умножение любой стохастической матрицы на которую дает опять-таки стохастическую матрицу.

Мы попытались решить проблему корректирования результатов, подобрав средние коэффициенты для спрогнозированных векторов. Однако, расчеты показали, что данный подход лишь ухудшает прогноз. Поэтому мы останавливаемся на том, что делаем прогноз с помощью простых цепей Маркова.

Заключение.

Эта работа представляет новую методологию оценки и прогнозирования качества образования в контексте системного анализа МИЭФ. Эта методология основана на аппарате Марковских цепей, в подобном ключе впервые применимом к анализу образования.

Была сделана попытка представить МИЭФ в виде системы и проанализировать один из важных показателей жизнедеятельности этой системы – успеваемость студентов. Мы делаем ряд выводов, на основе проделанной работы.

Мы показываем, что систему оценки знаний студентов, как характеристики работы МИЭФ можно представить в виде Марковского процесса первой степени. Это позволяет нам прогнозировать динамику успеваемости, а ряд статистических тестов подтверждает достоверность полученных прогнозов.

Новая методика позволяет предвидеть изменения в показателях студентов, а значит, дает возможность влить на всю систему. Малейшее изменение организационной политики в настоящем, дает значительные изменения результатов в будущем. Поэтому процесс изменения показателей студентов, представленный в виде цепи Маркова, может служить методом повышения эффективности факультета.

Разработанный аппарат применим к анализу деятельности организации, прогнозированию рейтингов банков или предсказанию курсов валют.

Библиография.

1. *Баруча-Рид А.Т.*. Элементы теории Марковских процессов и их приложения. М.: НАУКА, 1969
2. *Вентцель Е.С., Овчаров Л.А.*. Теория случайных процессов и ее инженерные приложения. 5-е изд. М.: КНОРУС, 2010.
3. *Волкова В.Н., Денисов А.А.,* Теория систем и системный анализ. М.: Издательство Юрайт, 2013.
4. *Гнеденко Б.В., Коваленко И.Н.*. Введение в теорию массового обслуживания. М.: ЛКИ, 2011.
5. *Гнеденко Б.В.*. Очерк истории теории вероятностей. Курс теории вероятностей. 6-е изд. М.: НАУКА, 1988.
6. *Замков О.О., 2011.* Эконометрический анализ факторов академических достижений студентов в МИЭФ НИУ ВШЭ. Национальный Исследовательский Университет «Высшая Школа Экономики», стр.384-391.
7. *Качала В.В.* Теория систем и системный анализ. М.: Академия, 2013.
8. *Кельберт М. Я., Сухов Ю. М.*. Вероятность и статистика в примерах и задачах. Т. 2: Марковские цепи как отправная точка теории случайных процессов и их приложения. М.: МЦНМО, 2010.
9. *Колмогоров А.Н.*. Основные понятия теории вероятностей. М. 1974.
10. *Кремер Н. Ш.*. Теория вероятностей и математическая статистика. 3-е изд. М.: ЮНИТИ-ДАНА, 2009.
11. *Куликов Е.И.* Прикладной статистический анализ. М.: Горячая линия – Телеком, 2008.
12. *Марков А.А.*. Избранные труды. Ленинград: Академия Наук СССР, 1951.
13. *Пересецкий А.А., Давтян М.А.,* 2011. Эффективность ЕГЭ и олимпиад как инструмента отбора абитуриентов. Прикладная Эконометрика, №3(23), стр.41-56.
14. *Саати Т. Л.*. Элементы теории массового обслуживания и ее приложения. 3-е изд. М.: ЛИБРОКОМ, 2010.
15. *Сарымсаков Т.А.*. Основы теории процессов Маркова. М.: ГОСУДАРСТВЕННОЕ ИЗДАТЕЛЬСТВО ТЕХНИКО-ТЕОРЕТИЧЕСКОЙ ЛИТЕРАТУРЫ, 1954.
16. *Стратонович Р.Л.*. Условные Марковские процессы и их применение к теории массового обслуживания. М.: МГУ,1965.
17. *Сулицкий В.Н.*. Деловая статистика и вероятностные методы в управлении и бизнесе. М.: ДЕЛО, 2009
18. *Anderson T.W. and Goodman, Leo A.,* 1957. Statistical Inference about Markov Chains. The Annals of Mathematical Statistics, Vol. 28, No.1, pp. 89-110.
19. *Androushchak Gregory, Poldin Oleg, Yudkevich Maria,* 2012. Peer Effects in Exogenously Formed University Student Groups. Higher School of Economics, WP BRP 03/EDU/2012.
20. *Billingsley, Patrick,* 1961. Statistical Methods in Markov Chains. The Annals of Mathematical Statistics, Vol. 32, No. 1, pp. 12-40.
21. *Clark William A.V.,* 2008, Reexamining the Moving to Opportunity Study and Its Contribution to Changing the Distribution of Poverty and Ethnic Concentration. Demography, Vol.45, No.3, pp. 515-535.

22. *Darling D.A.,* 1957. The Kolmogorov – Smirnov, Cramer – von – Mises Tests.The Annals of Mathematical Statistics, Vol.28, No.4, pp.823-838.
23. *Dryden, M. M.,* 1969. Share price movements: A Markovian approach. The Journal of Finance, 24, pp(49-60).
24. *Good I.J,* 1955. The Likelihood Ratio Test for Markoff Chains. Biometrika, Vol. 42, No.3/4, pp. 531-533.
25. *Gregory A.W.,* 1989. A Nonparametric Test for Autoregressive Conditional Heteroscedasticity: A Markov-Chain Approach. Journal of Business & Economic Statistics, 9, pp(107-115)
26. *Hoel, Paul G,* 1954. A Test for Markoff Chains. Biometrika, Vol.41, No3/4, pp. 430-433.
27. *Kiser Terry L., McCready Thomas A., Schwertman Niel C.,* 2005. Can the Commitee Meet? A Markov Chain Analysis. Mathematics Magazine, Vol. 78, No.1, pp.57-63.
28. *Kliygi Calli M. and Weverbergh M.,* 2009. Forecasting Newspaper Demand with Censored Regression. The Journal of Operational Research Society, Vol.60, No.7, pp.944-951.
29. *Lockshin J.L., Zamkov O.O.,* 2009. Predictors of Academic Achievement and their Possible Applications. Policy Futures in Education, Vol. 7. № 4, pp.399-409.
30. *McQueen G., Thorley S.,* 1991. Are Stock Returns Predictable? A Test Using Markov Chains. The Journal of Finance, 46, pp(239-263).
31. *Tomkins C,* 2006. An Introduction to Non-parametric Statistics for Health Scientists. University of Alberta Health Sciences Journal, Vol. 3, No.1, pp. 20-26.
32. *Turner, C. M., Startz R., Nelson C. R.,* 1989. A Markov model of heteroscedasticity, risk, and learning in the stock market. The Journal of Financial Economics, 25, pp(3-22).
33. *Yakovlev Sergey, Zamkov Oleg*, 2013.Factors and Models of Academic Success of Double Degree Programme in Economics. Report at 2nd International Scientific Conference “Economic and Social Development”, 4-5 April, 2013 Paris, France.

Приложение 1: Обозначения переменных

Обозначение	Расшифровка
cal0103	Calculus - Математический анализ 1 курс, поступление 2003 года
cal0104	Calculus - Математический анализ 1 курс, поступление 2004 года
cal0105	Calculus - Математический анализ 1 курс, поступление 2005 года
cal0106	Calculus - Математический анализ 1 курс, поступление 2006 года
cal0107	Calculus - Математический анализ 1 курс, поступление 2007 года
cal0108	Calculus - Математический анализ 1 курс, поступление 2008 года
cal0109	Calculus - Математический анализ 1 курс, поступление 2009 года
cal0110	Calculus - Математический анализ 1 курс, поступление 2010 года
cal0111	Calculus - Математический анализ 1 курс, поступление 2012 года
mac0103	Macroeconomics 1 - Макроэкономика 1 курс, поступление 2003 года
mac0104	Macroeconomics 1 - Макроэкономика 1 курс, поступление 2004 года
mac0105	Macroeconomics 1 - Макроэкономика 1 курс, поступление 2005 года
mac0106	Macroeconomics 1 - Макроэкономика 1 курс, поступление 2006 года
mac0107	Macroeconomics 1 - Макроэкономика 1 курс, поступление 2007 года
mac0108	Macroeconomics 1 - Макроэкономика 1 курс, поступление 2008 года
mac0109	Macroeconomics 1 - Макроэкономика 1 курс, поступление 2009 года
mac0110	Macroeconomics 1 - Макроэкономика 1 курс, поступление 2010 года
mac0111	Macroeconomics 1 - Макроэкономика 1 курс, поступление 2011 года
mic103	Microeconomics 1 - Микроэкономика 1 курс, поступление 2003 года
mic104	Microeconomics 1 - Микроэкономика 1 курс, поступление 2004 года
mic105	Microeconomics 1 - Микроэкономика 1 курс, поступление 2005 года
mic106	Microeconomics 1 - Микроэкономика 1 курс, поступление 2006 года
mic107	Microeconomics 1 - Микроэкономика 1 курс, поступление 2007 года
mic108	Microeconomics 1 - Микроэкономика 1 курс, поступление 2008 года
mic109	Microeconomics 1 - Микроэкономика 1 курс, поступление 2009 года
mic110	Microeconomics 1 - Микроэкономика 1 курс, поступление 2010 года
mic111	Microeconomics 1 - Микроэкономика 1 курс, поступление 2011 года
stt0103	Statistics 1 - Статистика 1 курс, поступление 2003 года
stt0104	Statistics 1 - Статистика 1 курс, поступление 2004 года
stt0105	Statistics 1 - Статистика 1 курс, поступление 2005 года
stt0106	Statistics 1 - Статистика 1 курс, поступление 2006 года
stt0107	Statistics 1 - Статистика 1 курс, поступление 2007 года
stt0108	Statistics 1 - Статистика 1 курс, поступление 2008 года
stt0109	Statistics 1 - Статистика 1 курс, поступление 2009 года
stt0110	Statistics 1 - Статистика 1 курс, поступление 2010 года
stt0111	Statistics 1 - Статистика 1 курс, поступление 2011 года
lal0203	Linear Algebra - Линейная алгебра 2 курс, поступление 2003 года
lal0204	Linear Algebra - Линейная алгебра 2 курс, поступление 2004 года
lal0205	Linear Algebra - Линейная алгебра 2 курс, поступление 2005 года
lal0206	Linear Algebra - Линейная алгебра 2 курс, поступление 2006 года
lal0207	Linear Algebra - Линейная алгебра 2 курс, поступление 2007 года
lal0208	Linear Algebra - Линейная алгебра 2 курс, поступление 2008 года
lal0209	Linear Algebra - Линейная алгебра 2 курс, поступление 2009 года
lal0210	Linear Algebra - Линейная алгебра 2 курс, поступление 2010 года
mac0203	Macroeconomics 2 - Макроэкономика 2 курс, поступление 2003 года

Обозначение	Расшифровка
mac0204	Macroeconomics 2 - Макроэкономика 2 курс, поступление 2004 года
mac0205	Macroeconomics 2 - Макроэкономика 2 курс, поступление 2005 года
mac0206	Macroeconomics 2 - Макроэкономика 2 курс, поступление 2006 года
mac0207	Macroeconomics 2 - Макроэкономика 2 курс, поступление 2007 года
mac0208	Macroeconomics 2 - Макроэкономика 2 курс, поступление 2008 года
mac0209	Macroeconomics 2 - Макроэкономика 2 курс, поступление 2009 года
mac0210	Macroeconomics 2 - Макроэкономика 2 курс, поступление 2010 года
mat0203	Maths for Economists - Матэк 2 курс, поступление 2003 года
mat0204	Maths for Economists – Матэк 2 курс, поступление 2004 года
mat0205	Maths for Economists - Матэк 2 курс, поступление 2005 года
mat0206	Maths for Economists - Матэк 2 курс, поступление 2006 года
mat0207	Maths for Economists - Матэк 2 курс, поступление 2007 года
mat0208	Maths for Economists - Матэк 2 курс, поступление 2008 года
mat0209	Maths for Economists - Матэк 2 курс, поступление 2009 года
mat0210	Maths for Economists - Матэк 2 курс, поступление 2010 года
mor0203	Методы оптимальных решений 2 курс, поступление 2003 года
mor0204	Методы оптимальных решений 2 курс, поступление 2004 года
mor0205	Методы оптимальных решений 2 курс, поступление 2005 года
mor0206	Методы оптимальных решений 2 курс, поступление 2006 года
mor0207	Методы оптимальных решений 2 курс, поступление 2007 года
mor0208	Методы оптимальных решений 2 курс, поступление 2008 года
stt0203	Statistics 2 - Статистика 2 курс, поступление 2003 года
stt0204	Statistics 2 - Статистика 2 курс, поступление 2004 года
stt0205	Statistics 2 - Статистика 2 курс, поступление 2005 года
stt0206	Statistics 2 - Статистика 2 курс, поступление 2006 года
stt0207	Statistics 2 - Статистика 2 курс, поступление 2007 года
stt0208	Statistics 2 - Статистика 2 курс, поступление 2008 года
stt0209	Statistics 2 - Статистика 2 курс, поступление 2009 года
stt0210	Statistics 2 - Статистика 2 курс, поступление 2010 года
mic0203	Microeconomics 2 - Микроэкономика 2 курс, поступление 2003 года
mic0204	Microeconomics 2 - Микроэкономика 2 курс, поступление 2004 года
mic0205	Microeconomics 2 - Микроэкономика 2 курс, поступление 2005 года
mic0206	Microeconomics 2 - Микроэкономика 2 курс, поступление 2006 года
mic0207	Microeconomics 2 - Микроэкономика 2 курс, поступление 2007 года
mic0208	Microeconomics 2 - Микроэкономика 2 курс, поступление 2008 года
mic0209	Microeconomics 2 - Микроэкономика 2 курс, поступление 2009 года
mic0210	Microeconomics 2 - Микроэкономика 2 курс, поступление 2010 года
acc0303	Accounting - Бухгалтерский учет 3курс, поступление 2003 года
acc0304	Accounting - Бухгалтерский учет 3курс, поступление 2004 года
acc0305	Accounting - Бухгалтерский учет 3курс, поступление 2005 года
acc0306	Accounting - Бухгалтерский учет 3курс, поступление 2006 года
acc0307	Accounting - Бухгалтерский учет 3курс, поступление 2007 года
acc0308	Accounting - Бухгалтерский учет 3курс, поступление 2008 года
acc0309	Accounting - Бухгалтерский учет 3курс, поступление 2009 года
ecm0303	Econometrics - Эконометрика 3 курс, поступление 2003 года
ecm0304	Econometrics - Эконометрика 3 курс, поступление 2004 года
ecm0305	Econometrics - Эконометрика 3 курс, поступление 2005 года

Обозначение	Расшифровка
ecm0306	Econometrics - Эконометрика 3 курс, поступление 2006 года
ecm0307	Econometrics - Эконометрика 3 курс, поступление 2007 года
ecm0308	Econometrics - Эконометрика 3 курс, поступление 2008 года
ecm0309	Econometrics - Эконометрика 3 курс, поступление 2009 года
mic0303	Microeconomics 3 - Микроэкономика 3 курс, поступление 2003 года
mic0304	Microeconomics 3 - Микроэкономика 3 курс, поступление 2004 года
mic0305	Microeconomics 3 - Микроэкономика 3 курс, поступление 2005 года
mic0306	Microeconomics 3 - Микроэкономика 3 курс, поступление 2006 года
mic0307	Microeconomics 3 - Микроэкономика 3 курс, поступление 2007 года
mic0308	Microeconomics 3 - Микроэкономика 3 курс, поступление 2008 года
mic0309	Microeconomics 3 - Микроэкономика 3 курс, поступление 2009 года
mac0303	Macroeconomics 3 - Макроэкономика 3 курс, поступление 2003 года
mac0304	Macroeconomics 3 - Макроэкономика 3 курс, поступление 2004 года
mac0305	Macroeconomics 3 - Макроэкономика 3 курс, поступление 2005 года
mac0306	Macroeconomics 3 - Макроэкономика 3 курс, поступление 2006 года
mac0307	Macroeconomics 3 - Макроэкономика 3 курс, поступление 2007 года
mac0308	Macroeconomics 3 - Макроэкономика 3 курс, поступление 2008 года
mac0309	Macroeconomics 3 - Макроэкономика 3 курс, поступление 2009 года
cor043	Corporate Finance - Корпоративные финансы 4 курс, поступление 2003 года
cor044	Corporate Finance - Корпоративные финансы 4 курс, поступление 2004 года
cor045	Corporate Finance - Корпоративные финансы 4 курс, поступление 2005 года
cor046	Corporate Finance - Корпоративные финансы 4 курс, поступление 2006 года
cor047	Corporate Finance - Корпоративные финансы 4 курс, поступление 2007 года
cor048	Corporate Finance - Корпоративные финансы 4 курс, поступление 2009 года
int0403	International Economics - Международная экономика, поступление 2003 года
int0404	International Economics - Международная экономика, поступление 2004 года
int0405	International Economics - Международная экономика, поступление 2005 года
int0406	International Economics - Международная экономика, поступление 2006 года
int0407	International Economics - Международная экономика, поступление 2007 года
int0408	International Economics - Международная экономика, поступление 2008 года
rat0103	Итоговый рейтинг (МИЭФ) 1 курс, поступление 2003 года
rat0104	Итоговый рейтинг (МИЭФ) 1 курс, поступление 2004 года
rat0105	Итоговый рейтинг (МИЭФ) 1 курс, поступление 2005 года
rat0106	Итоговый рейтинг (МИЭФ) 1 курс, поступление 2006 года
rat0107	Итоговый рейтинг (МИЭФ) 1 курс, поступление 2007 года
rat0108	Итоговый рейтинг (МИЭФ) 1 курс, поступление 2008 года
rat0109	Итоговый рейтинг (МИЭФ) 1 курс, поступление 2009 года
rat0110	Итоговый рейтинг (МИЭФ) 1 курс, поступление 2010 года
rat0111	Итоговый рейтинг (МИЭФ) 1 курс, поступление 2011 года
rat0203	Итоговый рейтинг (МИЭФ) 2 курс, поступление 2003 года
rat0204	Итоговый рейтинг (МИЭФ) 2 курс, поступление 2004 года
rat0205	Итоговый рейтинг (МИЭФ) 2 курс, поступление 2005 года
rat0206	Итоговый рейтинг (МИЭФ) 2 курс, поступление 2006 года
rat0207	Итоговый рейтинг (МИЭФ) 2 курс, поступление 2007 года
rat0208	Итоговый рейтинг (МИЭФ) 2 курс, поступление 2008 года
rat0209	Итоговый рейтинг (МИЭФ) 2 курс, поступление 2009 года
rat0210	Итоговый рейтинг (МИЭФ) 2 курс, поступление 2010 года

Обозначение	Расшифровка
rat0303	Итоговый рейтинг (МИЭФ) 3 курс, поступление 2003 года
rat0304	Итоговый рейтинг (МИЭФ) 3 курс, поступление 2004 года
rat0305	Итоговый рейтинг (МИЭФ) 3 курс, поступление 2005 года
rat0306	Итоговый рейтинг (МИЭФ) 3 курс, поступление 2006 года
rat0307	Итоговый рейтинг (МИЭФ) 3 курс, поступление 2007 года
rat0308	Итоговый рейтинг (МИЭФ) 3 курс, поступление 2008 года
rat0309	Итоговый рейтинг (МИЭФ) 3 курс, поступление 2009 года
rat0403	Итоговый рейтинг (МИЭФ) 4 курс, поступление 2003 года
rat0404	Итоговый рейтинг (МИЭФ) 4 курс, поступление 2004 года
rat0405	Итоговый рейтинг (МИЭФ) 4 курс, поступление 2005 года
rat0406	Итоговый рейтинг (МИЭФ) 4 курс, поступление 2006 года
rat0407	Итоговый рейтинг (МИЭФ) 4 курс, поступление 2007 года
rat0408	Итоговый рейтинг (МИЭФ) 4 курс, поступление 2008 года

Приложение 2: Описательные статистики переменных; источник: расчеты автора.

Переменная	Количество наблюдений	Среднее	Стандартное отклонение	Мин. Значение	Макс. Значение
acc0303	77	51,53247	13,24027	19	82
acc0304	78	53,9359	15,64404	26	90
acc0305	78	49,5	14,90642	13	88
acc0306	71	52,57746	16,80277	16	83
acc0307	70	48,57143	16,15287	19	81
acc0308	91	52,62637	16,60498	17	85
acc0309	96	53,0625	16,39949	11	91
cal0103	101	50,16832	15,27813	9	76
cal0104	105	49,82857	13,72985	14	73
cal0105	115	49,01739	12,13736	15	79
cal0106	107	46,57944	15,54667	6	75
cal0107	115	45,61739	14,80492	15	83
cal0108	144	49,25	16,09956	13	89
cal0109	132	40,51515	17,3867	3	84
cal0110	172	40,45349	18,615	2	97
cal0111	201	49,87562	19,19243	4	90
cor043	64	44,60938	17,42887	2	83
cor044	74	53	16,71108	18	92
cor045	71	61,01408	15,52831	22	86
cor046	66	51,54545	13,9749	21	78
cor047	63	36,4127	13,5297	13	71
cor048	81	56,06173	18,03285	16	95
ecm0303	76	45,60526	18,58464	11	86
ecm0304	77	48,61039	19,3283	21	91
ecm0305	81	47,24691	18,48752	6	89
ecm0306	74	47,67568	15,84956	20	82
ecm0307	70	40,62857	16,07786	11	79
ecm0308	91	45,79121	18,94818	5	84
ecm0309	96	43,95833	17,53968	7	88

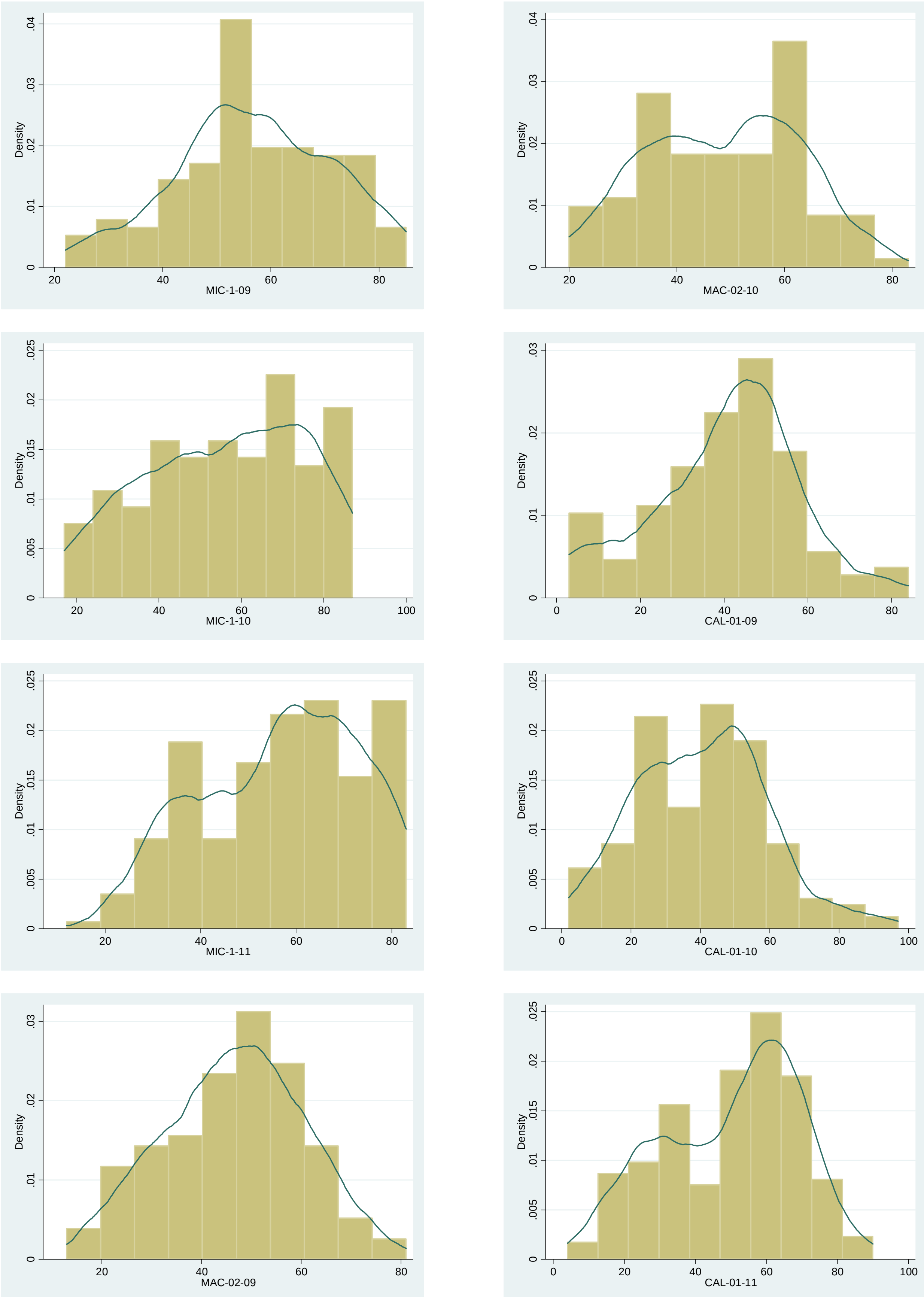
int0403	59	53,77966	16,41915	12	83
int0404	47	50,40426	13,95437	25	77
int0405	66	62,06061	14,29349	25	90
int0406	56	54	15,0478	27	83
int0407	64	55,875	14,85432	27	84
int0408	76	53,52632	14,56798	0	81
lal0203	83	59,77108	16,32096	21	95
lal0204	96	60,16667	21,9025	10	95
lal0205	98	71,10204	17,52376	27	99
lal0206	85	68,44706	17,67127	22	97
lal0207	98	49,66327	16,91516	14	83
lal0208	115	63,64348	18,42109	21	97
lal0209	112	66,47321	14,58122	30	99
lal0210	113	65,30088	13,64269	37	94
mac0103	103	50,39806	17,44366	12	83
mac0104	112	51,64286	15,97174	11	82
mac0105	114	47,64035	16,28242	9	77
mac0106	111	47,0991	16,92602	15	79
mac0107	117	41,30769	15,23294	2	76
mac0108	145	47,13793	17,74574	12	86
mac0109	133	47,07519	16,01402	11	82
mac0110	171	43,92982	18,06901	11	81
mac0111	202	46,01485	17,33385	4	83
mac0203	82	46,59756	16,03227	6	76
mac0204	101	48,46535	18,69629	12	86
mac0205	100	46,14	16,72053	13	78
mac0206	87	38,75862	12,80803	6	67
mac0207	100	35,3	15,30333	9	75
mac0208	116	40,56897	17,0392	4	82
mac0209	113	46,52212	13,95983	13	81
mac0210	113	48,65487	14,10089	20	83
mac0303	73	46,13699	16,49774	3	89

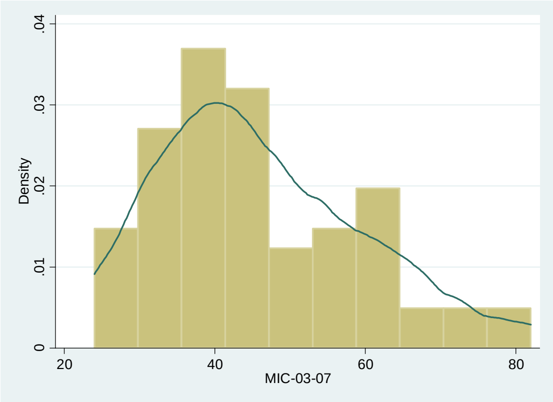
mac0304	78	46,9359	15,47293	19	84
mac0305	80	49,575	15,00782	14	86
mac0306	75	49,26667	16,68575	11	83
mac0307	70	38,87143	15,56093	13	85
mac0308	91	43,13187	15,35521	7	80
mac0309	97	41,78351	14,9131	5	74
mat0203	84	62,46429	18,59424	4	95
mat0204	101	65,42574	16,92829	20	98
mat0205	100	65,54	16,7895	25	97
mat0206	87	71,91954	15,45678	41	98
mat0207	100	63,32	15,47685	32	94
mat0208	116	60,06897	17,86165	8	94
mat0209	113	65,45133	15,59246	31	96
mat0210	112	61,25893	12,82364	38	88
mic0203	80	49,1	15,96642	11	87
mic0204	101	49,46535	20,59639	11	92
mic0205	100	48,04	17,98704	14	89
mic0206	87	43,98851	15,22047	12	77
mic0207	100	39,61	14,98679	10	79
mic0208	116	42,69828	16,94144	9	80
mic0209	113	48,72566	12,06252	21	81
mic0210	113	56,16814	14,46603	21	88
mic0303	80	41,775	12,54256	15	78
mic0304	78	43,97436	15,63959	16	81
mic0305	81	43,35802	12,53326	14	78
mic0306	74	42,17568	14,84934	7	71
mic0307	70	46,04286	13,5523	24	82
mic0308	91	42,97802	17,44578	14	83
mic0309	97	48,15464	15,61072	20	87
mic103	103	55,63107	15,99264	15	85
mic104	111	57,05405	15,446	20	83
mic105	117	58,22222	16,26245	16	88

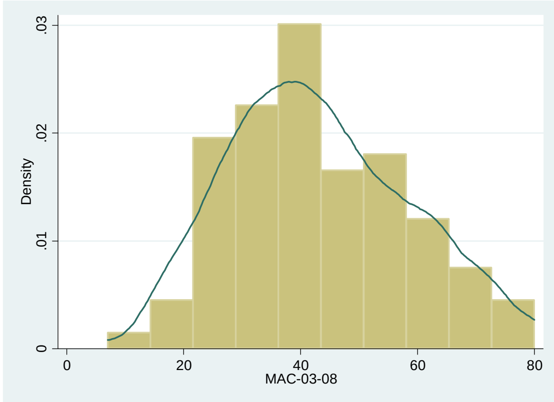
mic106	109	60,16514	14,95457	27	87
mic107	112	54,91964	14,03672	20	83
mic108	145	57,51724	15,30618	17	86
mic109	133	56,57895	14,347	22	85
mic110	171	55,95322	19,26283	17	87
mic111	202	56,63366	16,34986	12	83
mor0203	83	50,38554	20,10329	8	88
mor0204	101	59,33663	26,97824	1	99
mor0205	99	57,92929	20,98019	2	93
mor0206	87	59,73563	22,68052	6	97
mor0207	100	56,94	22,44896	5	99
mor0208	113	52,89381	23,17687	7	93
rat0103	78	53,97436	12,84672	25	83
rat0104	105	58,45714	13,51924	26	84
rat0105	112	57,97321	12,62663	22	84
rat0106	105	54,91429	14,43827	13	84
rat0107	111	51,88288	13,41352	19	80
rat0108	140	55,04286	14,12069	20	86
rat0109	133	51,91729	14,24145	13	82
rat0110	172	51,04651	15,50809	11	83
rat0111	195	49,87179	14,82402	19	86
rat0203	81	53,62963	12,90004	25	83
rat0204	101	54,08911	17,60233	16	87
rat0205	100	55,27	15,19427	24	88
rat0206	86	53,60465	14,222	24	82
rat0207	100	48,93	13,50814	26	80
rat0208	116	52,66379	14,82885	18	85
rat0209	113	55,69912	12,78942	30	85
rat0210	113	56,02655	11,57619	32	84
rat0303	74	47,72973	14,13564	20	82
rat0304	78	49,79487	16,67673	22	84
rat0305	80	49,6875	14,31702	14	83

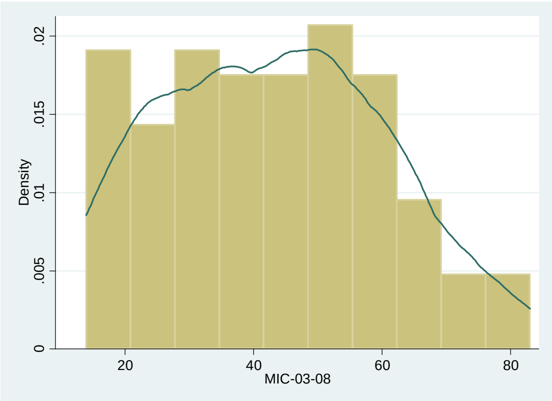
rat0306	74	51,5	14,70276	19	81
rat0307	70	46,37143	14,09509	23	78
rat0308	91	50,78022	14,75639	23	80
rat0309	97	52,01031	14,62411	19	87
rat0403	64	50,51563	16,07916	15	81
rat0404	74	53,5	16,36003	29	90
rat0405	71	59,69014	13,20994	29	88
rat0406	66	52,13636	12,97202	28	77
rat0407	67	47,86567	14,51947	16	83
rat0408	81	55,67901	15,52967	20	85
stt0103	103	43,34951	17,43093	5	76
stt0104	113	48,69912	17,78186	13	82
stt0105	117	53,13675	15,22866	19	84
stt0106	111	46,85586	15,80727	13	80
stt0107	118	45,5678	15,93012	12	74
stt0108	145	44,43448	15,43032	14	78
stt0109	133	44,28571	16,00926	6	78
stt0110	172	40,22674	15,89736	7	79
stt0111	206	36,21359	15,53825	4	74
stt0203	80	57,95	16,68616	22	94
stt0204	101	54,31683	22,52951	13	100
stt0205	100	59,67	17,94214	19	100
stt0206	87	58,01149	17,161	21	96
stt0207	100	53,59	17,39795	20	96
stt0208	115	53,01739	19,2882	15	97
stt0209	113	59,67257	16,74304	26	100
stt0210	113	56,51327	15,51101	26	89

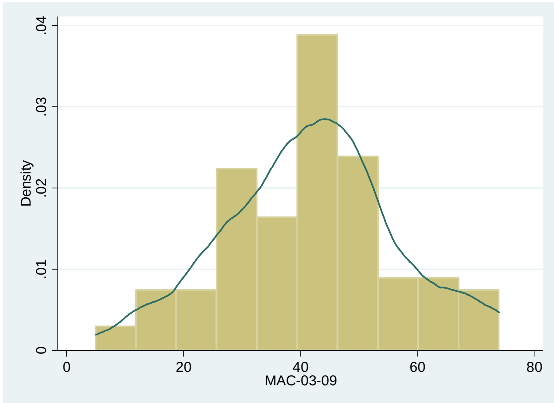
Приложение 3: Графики распределений результатов студентов по предметам; источник: расчеты автора

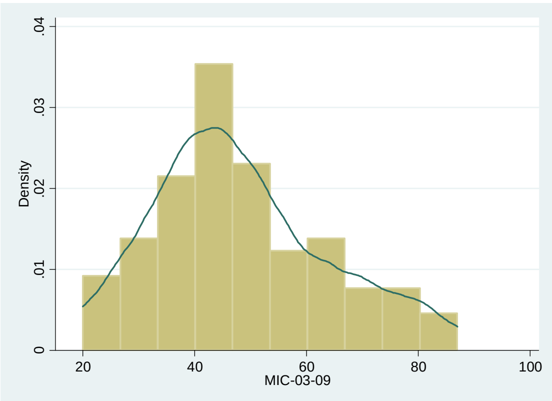


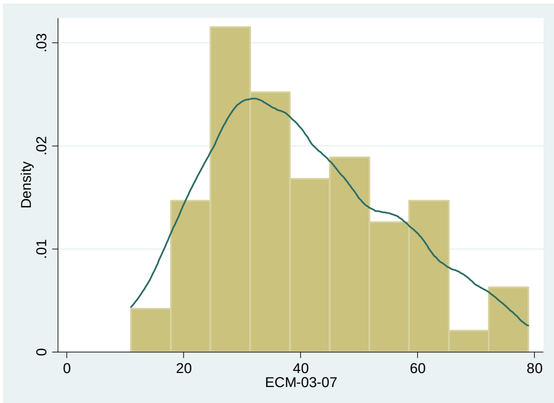

Density
0
.01
.02
.03
.04
20
40
60
80
MIC-03-07

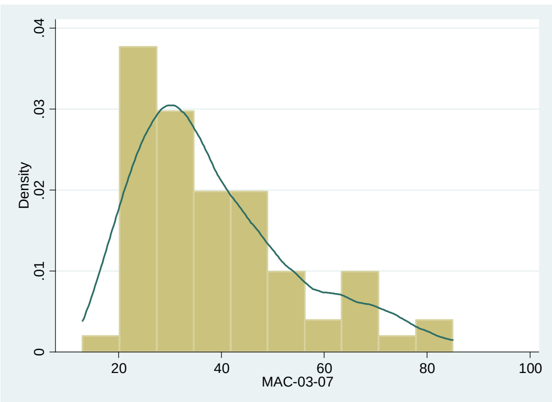

Density
0
.01
.02
.03
0
20
40
60
80
MAC-03-08

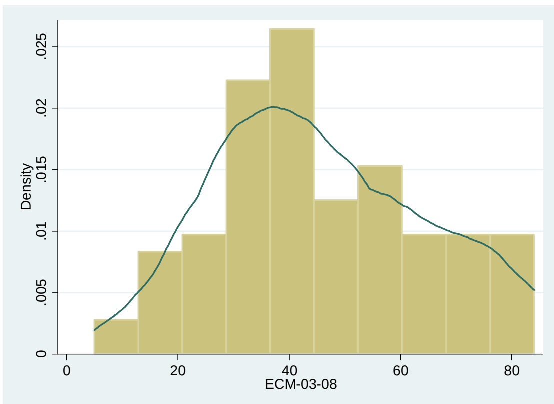

Density
0
.005
.01
.015
.02
20
40
60
80
MIC-03-08

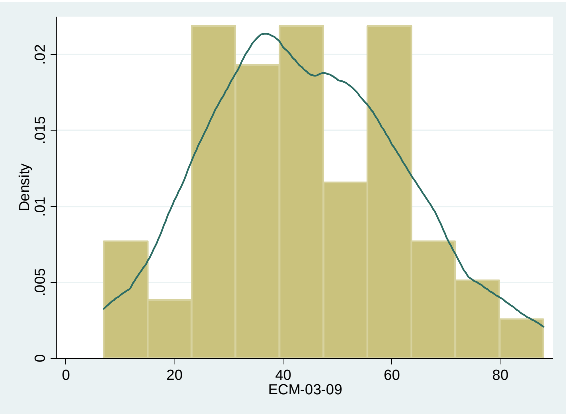

Density
0
.01
.02
.03
.04
0
20
40
60
80
MAC-03-09

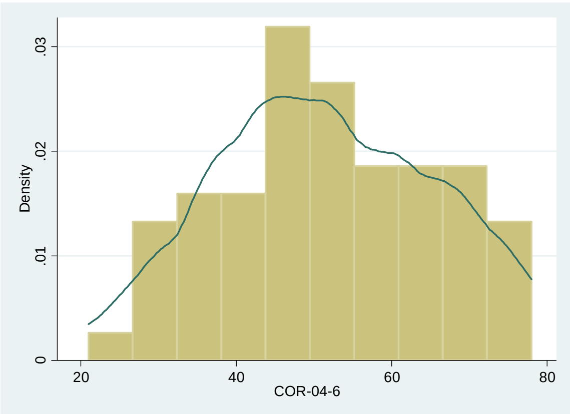

Density
0
.01
.02
.03
.04
20
40
60
80
100
MIC-03-09

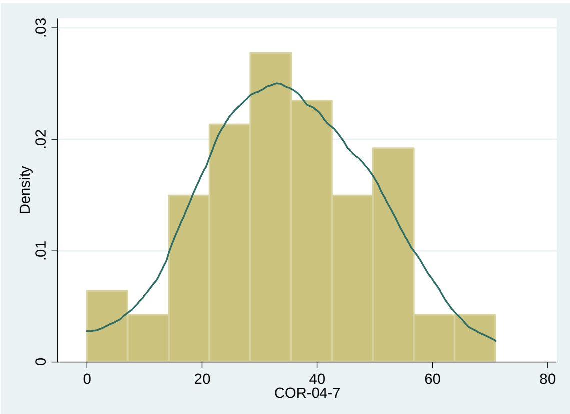

Density
0
.01
.02
.03
0
20
40
60
80
ECM-03-07

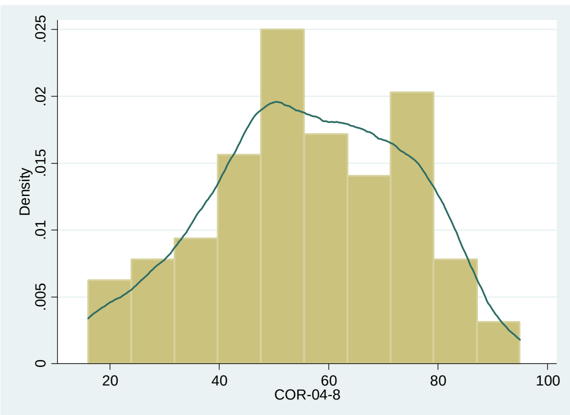

Density
0
.01
.02
.03
.04
20
40
60
80
100
MAC-03-07

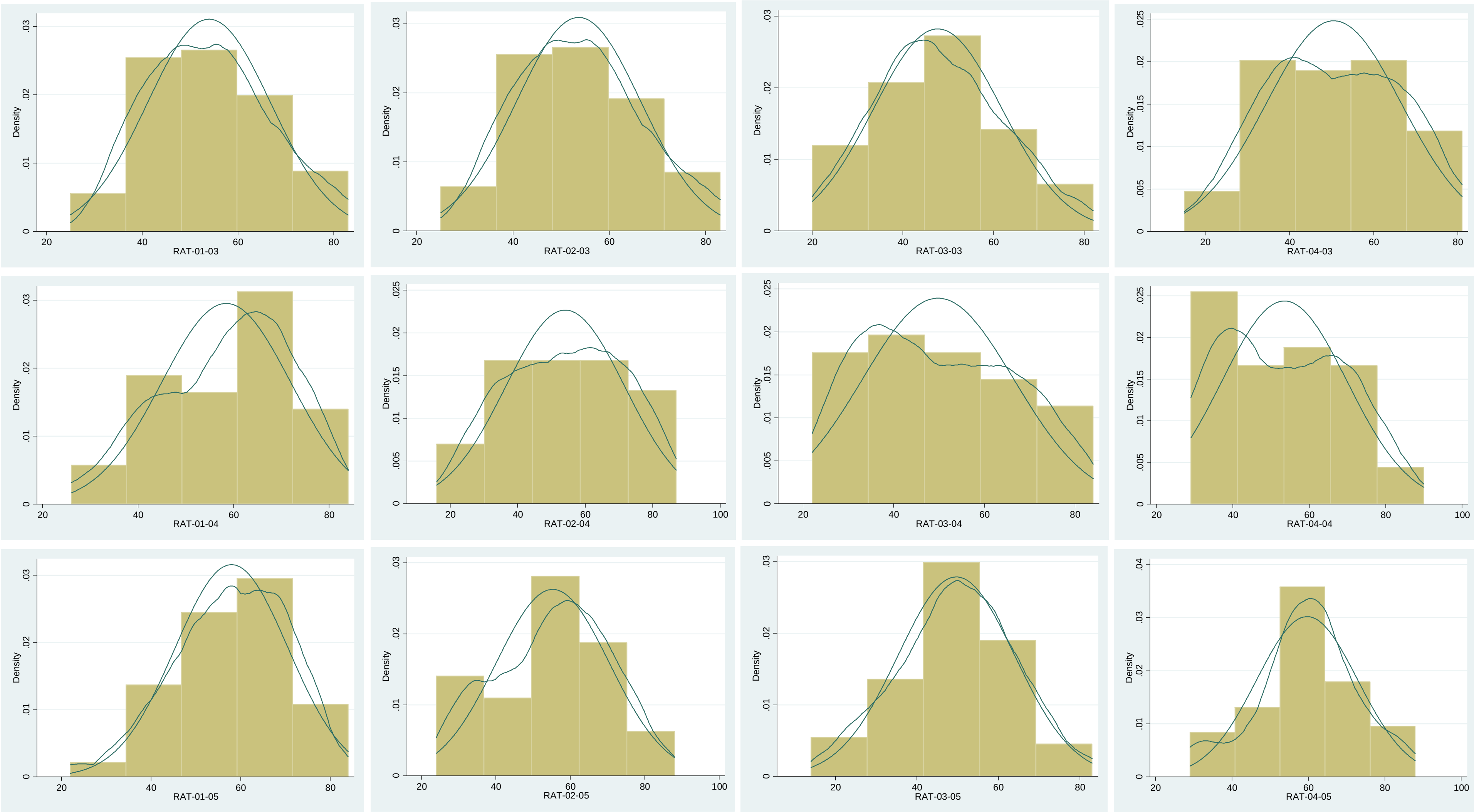

Density
0
.005
.01
.015
.02
.025
0
20
40
60
80
ECM-03-08

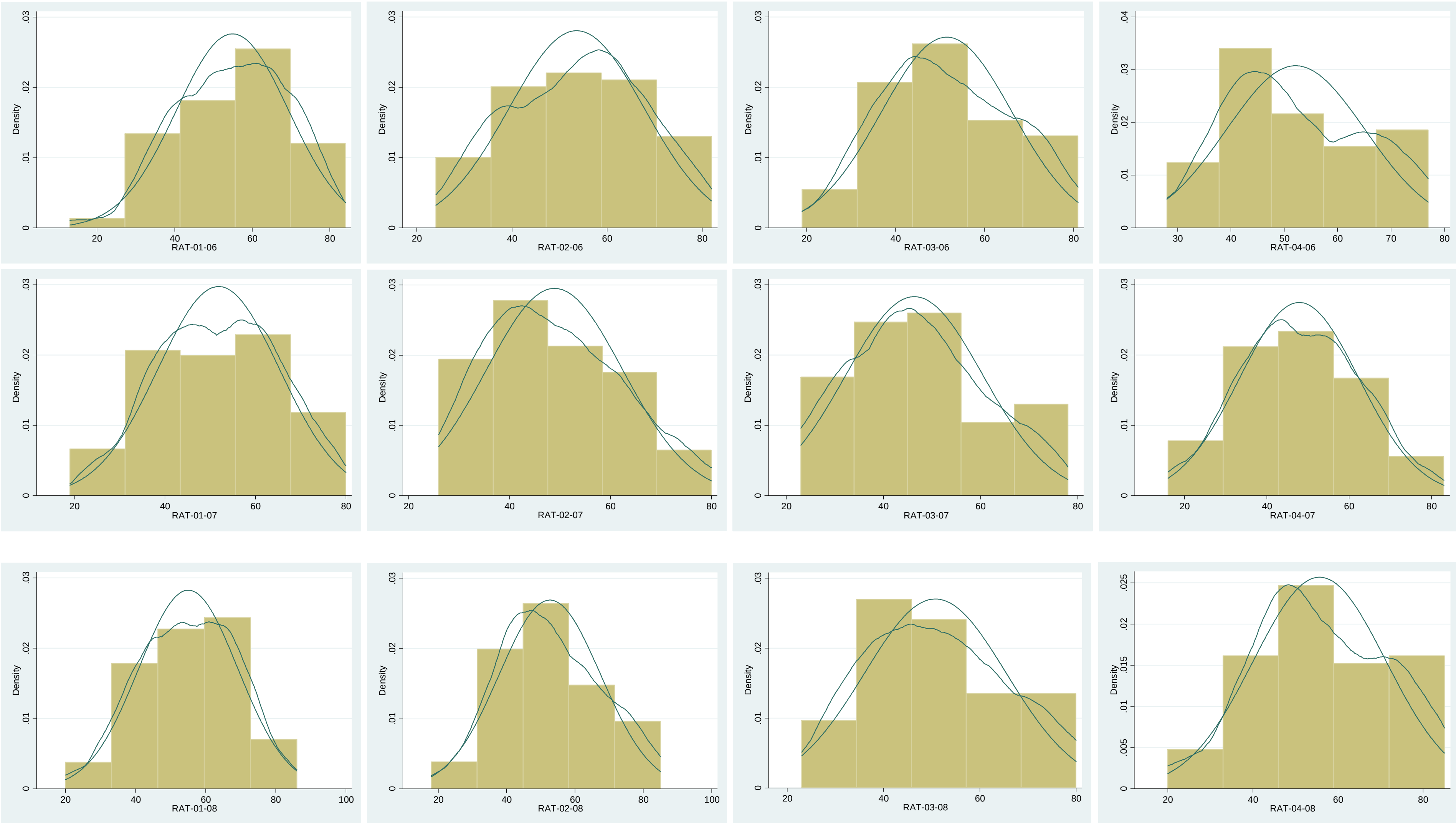

Density
0
.005
.01
.015
.02
0
20
40
60
80
ECM-03-09

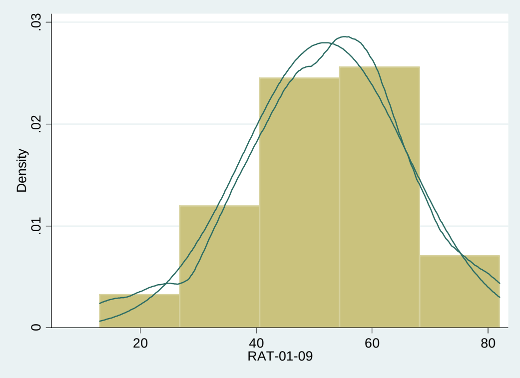

Density
0
.01
.02
.03
20
40
60
80
COR-04-6

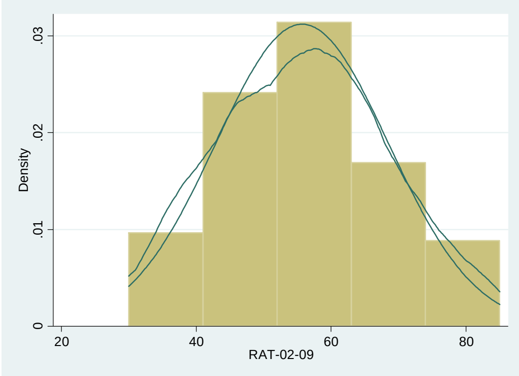

Density
0
.01
.02
.03
0
20
40
60
80
COR-04-7

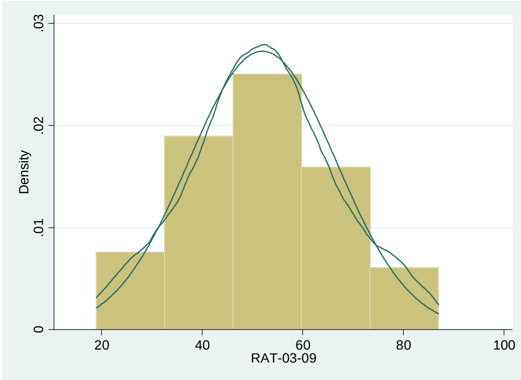

Density
0
.005
.01
.015
.02
.025
20
40
60
80
100
COR-04-8

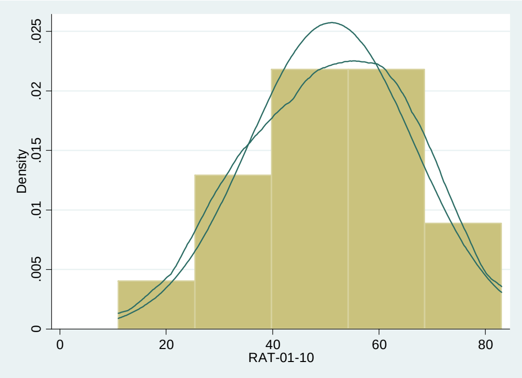

Density
RAT-01-03
Density
RAT-02-03
Density
RAT-03-03
Density
RAT-04-03
Density
RAT-01-04
Density
RAT-02-04
Density
RAT-03-04
Density
RAT-04-04
Density
RAT-01-05
Density
RAT-02-05
Density
RAT-03-05
Density
RAT-04-05

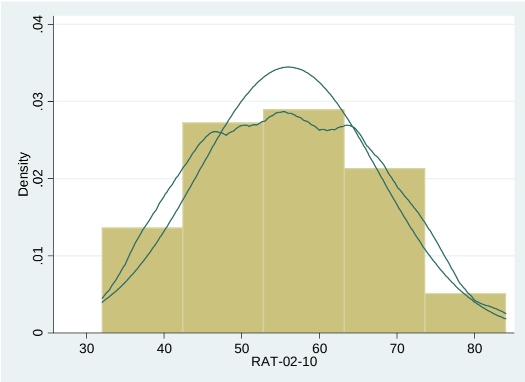

Density
RAT-01-06
Density
RAT-02-06
Density
RAT-03-06
Density
RAT-04-06
Density
RAT-01-07
Density
RAT-02-07
Density
RAT-03-07
Density
RAT-04-07
Density
RAT-01-08
Density
RAT-02-08
Density
RAT-03-08
Density
RAT-04-08

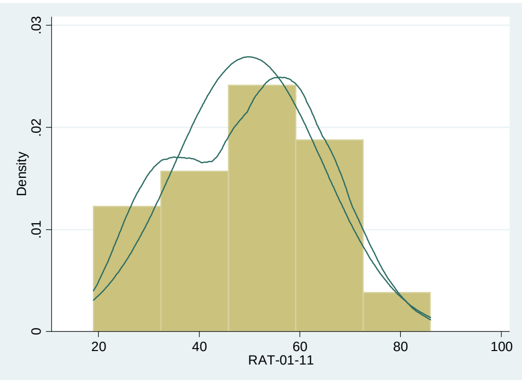

Density
0
.01
.02
.03
20
40
60
80
RAT-01-09


Density
0
.01
.02
.03
20
40
60
80
RAT-02-09


Density
0
.01
.02
.03
20
40
60
80
100
RAT-03-09


Density
0
.005
.01
.015
.02
.025
0
20
40
60
80
RAT-01-10


Density
0
.01
.02
.03
.04
30
40
50
60
70
80
RAT-02-10


Density
0
.01
.02
.03
20
40
60
80
100
RAT-01-11

Приложение 4: Теория случайных процессов и Марковских цепей. Источник: курсовая работа за второй курс Суншева А.

Теория случайных процессов – это наука, изучающая закономерности случайных явлений в динамике их развития. Понятие случайного процесса появилось в начале прошлого века и связано с именами А. Н. Колмогорова, А.Я. Хинчина, Е.Е. Слуцкого, Н. Винера и других.

Это понятие в наши дни является одним из центральных не только в теории вероятностей, но также в естествознании, инженерном деле, экономике, организации производства, теории связи. Теория случайных процессов принадлежит к категории наиболее быстро развивающихся математических дисциплин. Несомненно, это обстоятельство в значительной мере определяется ее глубокими связями с практикой.

XX век не мог удовлетвориться наследием, которое было получено им от прошлого. Действительно, в то время как физика, биолога, инженера интересовал процесс, т.е. изменение изучаемого явления во времени, теория вероятностей предлагала им в качестве математического аппарата лишь средства, изучавшие стационарные состояния. Для исследования изменения во времени теории вероятностей конца XIX начала XX века не имела ни разработанных частных схем, ни тем более общих приемов. Изучение броуновского движения в физике подвело математику к порогу создания теории случайных процессов. В исследованиях датского ученого А.К.Эрланга была начата новая важная область поисков, связанная с изучением загрузки телефонных сетей. Работы Эрланга оказали влияние на развитие не только чисто телефонных задач, но и на теорию случайных процессов: процессов гибели и размножения.[6]

Многие физические явления описываются теорией случайных процессов. К примеру, движение молекулы, в случайные моменты времени сталкивающейся с другими молекулами и при этом меняющей направление и скорость, является случайным процессом. Существует статистическая теория диффузии, основанная на теории случайных процессов и описывающая процессы диффузии и их характеристики. Радиоактивный распад молекул, напряжение в электросети, население города, полет космической ракеты, плотность воды в океане, направление ветра, уровень воды в реке – все это примеры случайных процессов.

В природе не существует неслучайных процессов, однако, есть факторы, влиянием которых в контексте конкретной задачи можно пренебречь. К примеру, решая задачу о составлении расписания самолетов, мы можем предположить, что траектории полета прямолинейны, а скорость полета равномерна. Однако это допущение нельзя делать, если мы сталкиваемся с задачей конструирования автопилота для управления полетом самолета.[7]

Случайная функция некоторой системы S – это случайные переходы системы из состояния в состояние. Случайную функцию называют случайным процессом в том

[6] *Гнеденко Б.В. Очерк по истории теории вероятностей.*
[7] *Вентцель Е.С. Овчаров Л.А. Теория случайных процессов и ее инженерные приложения.*

случае, если мы рассматриваем реализацию функции по времени. То есть случайный процесс – это случайная функция, аргументом которой является время – t.

Если мы зафиксируем время ($t=t_0$) , то случайное состояние системы превращается в некоторую случайную величину - одно из возможных состояний, в котором может оказаться система в момент времени t_0.

Случайный процесс может быть описан одной и более переменными. К примеру, напряжение в электросети описывается одной случайной переменной (напряжение) зависящей от времени, движение частицы – двумя переменными (в двухмерном пространстве – координатами X и Y), зависящими от времени. Примером многомерного случайного процесса может служить полет ракеты: здесь случайными переменными будут не только координаты в пространстве, но и углы наклона ракеты, скорость, запас топлива.[8]

Несмотря на то, что теория случайных процессов – сравнительно новая ветвь в теории вероятностей, существует большое количество отечественной и зарубежной литературы, посвященной этому разделу. Многие издания в незначительной степени отличаются друг от друга в определении случайного процесса, и в большинстве источников дается следующее определение случайного процесса:

«Случайным процессом X(t) называется процесс, значение которого при любом значении аргумента t является случайной величиной» [9]

Случайный процесс можно записать в виде функции двух аргументов: некоторого элементарного события ω и времени t.

$$X(t) = \varphi(t, \omega), \omega \in \Omega, t \in T, X(t) \in I$$

где $\omega -$ элементарное событие, $\Omega -$ пространство элементарных событий, T – *область значений аргумента t функции* X(t), I – *множество возможных значений случайного процесса X(t).*

Если опыт, в ходе которого протекает случайный процесс, уже произведен, то есть уже произошло элементарное событие $\omega \in \Omega$, случайный процесс перестает быть случайным и его зависимость от времени приняла вполне определенный вид – это уже обычная неслучайная функция аргумента *t*.

Реализацией случайного процесса X(t) называется неслучайная функция x(t) в которую превращается случайный процесс в результате опыта. Реализацию процесса можно записать как функцию от времени t при фиксированном элементарном событии ω.

$$x(t) = \varphi(t, \omega_0), t \in T$$

Если произведено более одного опыта, то мы получим несколько реализаций одного и того же процесса – семейство реализаций. На основе семейства реализаций мы можем получить характеристики случайного процесса: математическое ожидание и дисперсию.[10]

[8] *Вентцель Е.С. Овчаров Л.А. Теория случайных процессов и ее инженерные приложения.*
[9] *Кремер Н.Ш. Теория вероятностей и математическая статистика.*

Самыми простыми классификациями случайных процессов являются классификации по времени и по состояниям.

Случайный процесс называется процессом с дискретным временем, если система может менять свои состояния в фиксированные моменты, число которых конечно и счётно. Случайный процесс с непрерывным временем – это процесс, в котором переходы системы из одного состояния в другое могут происходить в любой момент времени.

Случайный процесс называется процессом с дискретными состояниями, если в любой момент времени множество его состояний конечно, и процессом с непрерывными состояниями – если в любой момент времени множество его состояний бесконечно или несчетно. Другими словами, если сечение процесса в любой момент времени является дискретной случайной величиной, то мы имеем дело с процессом с дискретными состояниями; в противном случае – с процессом с непрерывными состояниями.

Таким образом, мы можем разделить все процессы на четыре класса:

1. Процессы с дискретными состояниями и дискретным временем;
2. Процессы с дискретными состояниями и непрерывным временем;
3. Процессы с непрерывными состояниями и дискретным временем;
4. Процессы с непрерывными состояниями и непрерывным временем.[11]

Примером процесса с дискретным временем и дискретными состояниями может служить число билетов лотереи выигравших до момента t из общего числа билетов m. Процесс с дискретными состояниями и непрерывным временем: количество узлов прибора (из n узлов), отказавших до момента времени t. Процесс с непрерывными состояниями и дискретным временем: температура воздуха, измеряемая в определенные моменты времени. И наконец, процесс с непрерывными состояниями и непрерывным временем: напряжение в электросети.

Особое место в теории случайных процессов занимают Марковские процессы.

Случайный процесс, протекающий в системе I с дискретными состояниями $i_1, i_2, \ldots, i_i, \ldots$, называется марковским, или случайным процессом без последействия, если для любого момента времени t_0 вероятность каждого из состояний системы в будущем (при $t>t_0$) зависит только от ее состояния в настоящем($t=t_0$) и не зависит от того, когда и как она пришла в это состояние, то есть не зависит от ее поведения в прошлом(при $t<t_0$). (Будущее зависит от прошлого через настоящее).

Марковские процессы делятся на процессы с дискретным и с непрерывным временем. В некоторых источниках под цепями Маркова понимают только Марковские процессы с дискретным временем, однако есть и авторы (Кельберт, Сухов), которые все Марковские процессы называют цепями Маркова. В данной работе под цепями Маркова будем понимать Марковские процессы с дискретным временем.

Итак, цепью Маркова называется Марковский случайный процесс с дискретным временем, в котором его возможные состояния $i_1, i_2, \ldots$ можно заранее перечислить, а

---

[10] *Основные характеристики случайного процесса в данной работе не рассматриваются.*
[11] *Вентцель Е.С. Овчаров Л.А. Теория случайных процессов и ее инженерные приложения.*

переход из состояния в состояние происходит мгновенно (скачком), но только в определенные моменты времени (t_0, t_1,…), называемые шагами процесса.

Случайный механизм, вызывающий изменение состояния, описывается матрицей перехода P с элементами p_{ij} , где $i,j\in I$. Элемент p_{ij} равен вероятности, с которой система перейдет из состояния i в состояние j за единицу времени. Таким образом p_{ij} – это условная вероятность того, что система будет находиться в состоянии j в следующий момент, при условии что в данный момент она находится в состоянии i. значит все элементы P неотрицательны, но не превышают 1, и сумма элементов в любой строке равна 1:

$$0\leq p_{ij} \leq 1 \ \ \forall i,j \in I$$

$$\sum_{j\in I} p_{ij} = 1 \ \ \forall i \in I$$

Матрица P, обладающая такими свойствами называется стохастической, т. е. вероятностной.

Простейший случай имеет вид 2x2 (пространство из 2 состояний). Можно считать, что состояниями являются 0 и 1.

Тогда элементы матрицы имеют вид p_{ij}, i,j=0,1, а стохастическую матрицу можно представить в виде:

$$\begin{bmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{bmatrix}$$

где $0 \leq \alpha, \beta \leq 1$.

В частности при $\alpha = \beta = 0$ получаем единичную матрицу, а при $\alpha = \beta = 1$ – антидиагональную матрицу:

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Система с единично матрицей остается в начальном состоянии навсегда, а в антидиагональном случае она меняет состояние в каждый момент времени, переходя из 0 в 1 и обратно.

С другой стороны при $\alpha = \beta = \frac{1}{2}$ мы получаем матрицу

$$\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

В этом случае система может либо остаться в том же состоянии, либо поменять его с вероятностью ½.

Пусть X_n- состояние системы в момент *n*. Правила задающие марковскую цепь с начальным распределением λ и матрицей перехода P таковы:

1. X_0имеет распределение λ: $P(X_0 = i) = \lambda_i \ \forall i \in I$;
2. Более общим образом, $\forall n$ и $i_0, \dots, i_n \in I$ вероятность $P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n)$ того, что система находится в состоянии $i_0, i_1, \dots, i_n$ в моменты времени 0, 1, … , *n* записывается как произведение

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i_n}$$

1 – это частный случай 2 при *n*=0.

Для условной вероятности $P(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_n = i_n)$ того, что состояние в момент n+1 есть j, при условии что заданы состояния $i_0, \dots, i_{n-1}$ и

$i_n = i$ в моменты времени $0, \dots, n-1, n$:

$$P(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_n = i) =$$

$$\frac{P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i, X_{n+1} = j)}{P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i)} = \frac{\lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i} p_{ij}}{\lambda_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i}} = p_{ij}$$

Таким образом, при условии, что $X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i$, X_{n+1} имеет распределение $p_{ij}, j \in I$. В частности, условное распределение X_{n+1} не зависит от $i_0, \dots, i_{n-1}$, т. е. зависит только от состояния *i* в последний предшествующий момент *n*.

Эта формула иллюстрирует свойство ограниченной памяти цепи Маркова.

Теперь нас интересует вероятность $P(X_n = j)$того, что в момент *n* наша система находится в состоянии *j*. Для *n*=1:

$$P(X_n = j) = \sum_{i \in I} P(X_0 = i, X_1 = j),$$

где *i* – все возможные начальные состояния.

Тогда, $\sum_{i \in I} P(X_0 = i, X_1 = j) = \sum_{i \in I} \lambda_i p_{ij} = (\lambda P)_j$ для n=1.

И для общих значений n:

$$P(X_n = j) = \sum_{i_0, \dots, i_{n-1}} P(X_0 = i_0, X_1 = i_1, \dots, X_n = j) = \sum_{i_0, \dots, i_{n-1}} \lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} j} = (\lambda P^n)_j$$

где P^n – n-я степень матрицы P. Таким образом стохастический вектор, описывающий распределение случайной величины X_n, можно получить помножив матрицу P^n к начальному стохастическому вектору λ.

Теперь аналогично:

$$P(X_n = i, X_{n+1} = j) = \\ = \sum_{i_0,\dots,i_{n-1}} P(X_0 = i_0, X_1 = i_1, \dots, X_n = i, X_{n+1} = j) = \sum_{i_0,\dots,i_{n-1}} \lambda_{i_0} p_{i_0 i_1} \dots p_{i_{n-1} i} p_{ij} \\ = (\lambda P^n)_j p_{ij}$$

и отсюда следует, что

$$P(X_{n+1} = j | X_n = i) = \frac{P(X_n = i, X_{n+1} = j)}{P(X_n = i)} = \frac{(\lambda P^n)_j p_{ij}}{(\lambda P^n)_j} = p_{ij}.$$

То есть элемент p_{ij} равен условной вероятности того, что в следующий момент состояние будет j, если в данный момент оно есть i.

$$P(X_0 = i, X_n = j) = \\ = \sum_{i_1,\dots,i_{n-1}} P(X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = j) \\ = \sum_{i_i,\dots,i_{n-1}} \lambda_i \; p_{i \; i_1} \dots p_{i_{n-1} j} p_{ij} = \lambda_i (P^n)_{ij}$$

и

$$P(X_n = j | X_0 = i) = \frac{P(X_0 = i, X_n = j)}{P(X_0 = i)} = \frac{\lambda_i (P^n)_{ij}}{\lambda_i} = (P^n)_{ij}.$$

Значит, элемент $(P^n)_{ij}$ матрицы P^n дает вероятность перехода за n шагов из состояния i в состояние j.

В общем случае $P\,(X_k = i, X_{n+k} = j) = (\lambda P^k)_i (P^n)_{ij}$

и

$$P(X_{k+n} = j | k = i) = \frac{P(X_k = i, X_{k+n} = j)}{P(X_k = i)} = \frac{(\lambda P^k)_i (P^n)_{ij}}{(\lambda P^k)_i} = (P^n)_{ij}$$