

Санкт-Петербургский государственный университет
Кафедра математической теории игр и статистических
решений

Спиридонов Дмитрий Юрьевич

Магистерская диссертация

**Анализ социально-экономических показателей
методами прикладной статистики**

Направление 01.04.02

Прикладная математика и информатика

Магистерская программа

Исследование операций и системный анализ

Научный руководитель,
кандидат физ.-мат. наук,
старший преподаватель
Кумачева С.Ш.

Санкт-Петербург

2017

Оглавление

Введение.....	3
Актуальность.....	4
Постановка задачи	5
Обзор литературы	5
Глава 1. Формирование и обработка данных.....	8
1.1 Формирование выборки	8
1.2 Обработка данных.....	11
Глава 2. Регрессионный анализ	13
2.1 Выбор зависимого фактора.....	13
2.2 Методология построения моделей.....	15
2.3 Построение и верификация регрессионных моделей	17
2.3.1 Реализация первого подхода.....	17
2.3.2 Реализация второго подхода.....	28
2.3.3 Анализ на наличие выбросов	38
2.3.4 Построение и верификация модели по результатам пункта 2.3.3 ...	46
2.3.5 Еще один подход к построению модели	50
2.4 Интерпретация моделей.....	55
2.5 Результаты моделирования.....	57
2.5.1 Модель 1.1	57
2.5.2 Модель 2.3	59
Глава 3. Кластеризация	61
3.1 Результаты кластеризации.....	62
3.2 Сравнение показателей Российской Федерации.....	64
Выводы.....	66
Заключение.....	67
Список литературы	70
Приложение.....	74

Введение

На сегодняшний день на планете существует более 200 государств и зависимых территорий [1]. Они расположены в различных частях света, одни – размером с небольшой город, другие занимают огромные площади, но, несмотря на все различия, большая часть имеет широкие торгово-финансовые взаимосвязи [2]. Изучение национальных экономик государств является важной и неотъемлемой частью современного мира. Ученые стараются выявить зависимости и закономерности, которые не только помогают лучше понять и оценить ситуацию, но и предоставляют возможность выработать решения для достижения важных целей [3].

Социально-экономическое состояние страны характеризуется многими факторами, начиная от общего капитала государства, заканчивая состоянием окружающей среды и тем, как граждане этих государств проводят свой досуг. Сегодня такого рода информация собирается различными организациями практически по всему миру [4, 5]. Международная статистика позволяет сравнивать государства по многим критериям. Она помогает определить страны-ориентиры, страны схожего уровня развития и, в общем, представить то, как живет население Земного шара в его различных частях.

В современном обществе статистика и статистические данные играют крайне важную роль. С одной стороны, лица, принимающие решения на различных уровнях власти и в бизнесе, нуждаются в описательной информации в виде числовых данных. С другой стороны, общественность и средства массовой информации используют ее для отражения современной картины мира и сложившейся ситуации внутри отдельных регионов [6].

Аналитическая статистика в совокупности с макроэкономическим анализом могут ответить на многие вопросы. В частности, движется ли общество в заявленном политиками социально-экономическом курсе? Увеличился или уменьшился уровень безработицы, а также каковы ее

причины? Повысился ли объем выброса CO₂ в сравнении с показателем десятилетней давности? Насколько экономика страны эффективна по отношению к странам-соседям?

Актуальность

На сегодняшний день собрано огромное количество статистических данных по различным сферам жизнедеятельности. Однако большая часть этой информации отражает лишь частную сторону жизни и не позволяет ответить на вопрос о качестве жизни населения в целом. Данный вопрос стал особенно интересен в последние годы ввиду сложных экономической и политической ситуаций во многих странах.

Начиная с 2009 года, Евростат [7] принимает активное участие в развитии данной тематики, разрабатывая новые подходы к оценке прогресса общества и его благосостояния [6]. Этот вопрос обсуждался на заседании ООН в 2012 году [8]. Также большой вклад в развитие области внесли различные международные организации, в частности OECD (*The Organisation for Economic Co-operation and Development*) и Gallup [5, 9]. Ежегодно публикуются отчеты, отражающие реальное положение дел во всем мире или внутри отдельных государств [5]. В некоторых из них дается оценка схожим показателям, например, уровню счастья [6] или уровню благосостояния населения.

В ряде случаев подобного рода показатели формируются, исходя из проводимых опросов. Но возможно ли провести оценку, опираясь только на данные, находящиеся в открытом доступе?

Постановка задачи

Для ответа на указанные вопросы требуется решить следующие задачи:

- сформировать и реализовать подход к оценке качества жизни населения с использованием находящихся в открытом доступе индексов и показателей, с помощью которых описываются различные сферы жизнедеятельности человека;
- выявить факторы, играющие наиболее важную роль в формировании оценки качества жизни;
- определить положение, которое занимает Российская Федерация среди других государств мира;
- выявить направления, по которым наша страна отстает от ведущих государств.

Обзор литературы

Ознакомление с областью исследования началось с отчета, опубликованного в 2009 году, в результате неудовлетворительных показателей представленной статистики об экономическом и социальном состоянии Франции [10]. Главной целью работы было выявление причин такой ситуации. В ней авторы не только затрагивают важные вопросы и проблемы (такие, как проблемы современных способов измерений и оценок, на которые привыкли опираться аналитики при принятии решений; важность оценки качества жизни; сложность и реальная глубина многих простых, на первый взгляд, процессов), но и предлагают пути их преодоления. В статье описывается многомерность вопроса об уровне жизни населения. Представленная в нем модель легла в основу текущей работы, также на этапе сбора информации были учтены данные в этом отчете со стороны. Отдельно стоит отметить, что именно эта статья оказала существенное влияние на

Европейскую Статистическую Систему (ЕСС), которая в последствие приняла около 50 особых мер [11].

В отчете [12] отражен подход к формированию оценки такого понятия как «уровень счастья», также прилагается подробный анализ этого показателя в течение последних лет. Отчет впервые был опубликован в 2012 году в поддержку собрания ООН. Он охватывает 155 стран и территорий. В нем предполагается, что показатель зависит от таких факторов, как ВВП, уровень социальной поддержки, ожидаемая продолжительность здоровой жизни, свобода принятия решений и других. Регрессионный анализ, проводимый в отчете, основан на принципе использования наибольшей доступной выборки данных по необходимым для проведения исследования показателям. Аналогичным принципом будем руководствоваться в ходе текущего исследования. Одним из значимых результатов, публикуемых в отчете, является тот факт, что вариация уровня счастья в большей степени зависит от психического и физического здоровья, а также персональных взаимоотношений, чем от неравенства в доходах. Подобную зависимость сложнее выявить среди данных по бедным странам.

Важность оценки уровня счастья можно проследить на примере королевства Бутан. В 1970-е годы там впервые было введено определение валового национального счастья (ВНС). В последствие на основе этого концептуального понятия был разработан одноименный индекс, результаты подробного исследования которого описаны в [13]. На основе его интерпретации и анализа были проведены многочисленные реформы, более того, значимость этого показателя с 2008 года закреплена конституционно. Все эти действия помогли королевству стать открытой страной и пережить связанные с этим экономические и социальные потрясения.

В [14] раскрывается тема сложной природы вопроса качества жизни. В ней авторы приводят различные существующие формулировки этого понятия,

особое внимание уделяя тому, что данный показатель должен учитывать как объективные, так и субъективные оценки. Также в работе описывается методология формирования специальных анкет для проведения опросов и опубликованы результаты исследования по Вологодской области.

В документе [15], опубликованном OECD в 2013 году в рамках инициативы «Лучшая жизнь» (*Better Life Initiative*), описываются основные принципы и важность правильного измерения субъективного благосостояния. В ней авторы наиболее полно определяют это понятие, представляют концептуальные основы правильного измерения данных, на которые будем максимально опираться в текущей работе, а также приводят методологию формирования и проведения опросов. В отчете описывается возможное дальнейшее использование результатов исследования, а также наиболее удобные способы анализа полученных данных.

В отчете [16] представлены краткие результаты проведения 5го всемирного форума по статистике, знаниям и политике (*5th OECD World Forum on Statistics, Knowledge and Policy*), в котором приняло участие 1400 человек из 60 стран мира. Подавляющая часть выступлений была посвящена вопросам, связанным с качеством жизни и уровнем благосостояния, отражая важность рассматриваемого в текущей работе вопроса.

Помимо индекса уровня счастья, на сегодняшний день существует показатель мирового уровня счастья (*Planet Happy Index*), который впервые был предложен в 2006 году организацией New Economical Foundation. В отчете [17] описывается процесс формирования данного индекса как совокупности нескольких факторов, приведенных к необходимому виду. Некоторые из предложенных подходов применимы и реализованы в текущем исследовании.

Глава 1. Формирование и обработка данных

1.1 Формирование выборки

Как отмечено в работах [10, 14, 15], качество жизни – это очень сложное понятие, единого определения которого не существует. Ученые до сих пор ведут дискуссии по этому поводу [14], одни наделяют его более материальным смыслом, другие – социальным. Возникает закономерный вопрос о том, как оценить и измерить этот показатель. Возможный ответ представлен в работах [12, 17], однако, в них не раскрываются в полной мере причины использования тех или иных факторов. В работах [10, 15] предлагается иной подход к решению этого вопроса. Он основан на непосредственном анализе данных, относительно различных критериев, и сконцентрирован вокруг формируемых опросов, что не применимо в рамках данного исследования. Однако, в работе [10] описывается модель человеческой жизнедеятельности, которая делит ее на 8 основных сфер. Данная модель получила дальнейшее развитие в работе [6] и легла в основу анализа, проводимого Евростатом в рамках Евросоюза [6]. Она получила название «8 + 1», ввиду добавленного девятого аспекта – общий опыт жизни (*overall experience of life*). Эта модель представлена ниже, она легла в основу при поиске факторов и была принципиально модифицирована в ходе исследования.

Европейская модель человеческой жизнедеятельности «8+1»:

1. *Material living conditions (материальное благосостояние)*
2. *Productive or main activity (основная рабочая деятельность)*
3. *Health (здоровье)*
4. *Education (образование)*
5. *Leisure and social interactions (досуг и общение)*
6. *Economic and physical safety (физическая и экономическая безопасность)*

7. *Governance and basic rights (лояльность к правительству и основные права)*
8. *Natural and living environment (окружающая среда и жилищные условия)*
9. *Overall experience of life (общий опыт жизни)*

При поиске и отборе факторов были максимально учтены принципы и советы, описанные в работах [6, 15]. Поиск производился по общедоступным статистическим базам, полный список которых представлен в приложении. Так была собрана максимально полная и новая информация по 35 показателям. Все данные датируются периодом с 2014 по 2016 год. Количество стран, представленных указанными факторами, варьировалось от 105 до 225. В силу того, что по некоторым показателям в рассмотренный период времени имелось большое количество пропусков, часть из них была по возможности заменена данными за ближайший доступный период. Аналогичный подход применялся в работе [17]. Список индексов:

- UR - *Unemployment rate*. (Уровень безработицы).
- IH - *Intentional homicides per 100000 people*. (Количество убийств на 100000 населения).
- PSaAV - *Political Stability and Absence of Violence/Terrorism*. (Политическая стабильность и отсутствие насилия/терроризма).
- TTR – *Total Tax Revenue*. (Суммарные (итоговые) налоговые поступления).
- VA – *Voice and Accountability*. (Право голоса и прозрачность выборов).
- GE – *Government Effectiveness*. (Эффективность правительства).
- CC – *Control of Corruption*. (Контроль коррупции).
- RQ – *Regulatory Quality*. (Качество управления).
- FiW – *Freedom in the World*. (Свобода в мире).
- HDI – *Human Development Index*. (Индекс развития человеческого потенциала).
- EYoS – *Expected Years of Schooling*. (Ожидаемая длительность обучения (лет)).

- MYoS – *Mean Years of Schooling*. (Средняя длительность обучения (лет)).
- TTRate – *Total Tax Rate*. (Итоговая (общая) налоговая ставка).
- ITRate – *Income Tax Rate*. (Ставка налога на прибыль).
- DCPS – *Domestic Credit to Private Sector*. (Внутренний кредит частному сектору (проценты от ВВП)).
- EF – *Economic Freedom*. (Экономическая свобода).
- WPF – *World Press Freedom*. (Свобода прессы).
- PD – *Public Debt to GDP*. (Государственный долг).
- GD – *Government Debt to GDP*. (Аналогичен PD (см. выше)).
- MF – *index of Moral Freedom*. (Свобода морали).
- Gini – *Gini index*. (Индекс Джини).
- GI – *Gender Inequality index*. (Индекс гендерного неравенства).
- IHDI – *Inequality -adjusted HDI*. (ИРЧП с учетом полового неравенства).
- TR – *Tax Revenue*. (Налоговые поступления).
- GGGI – *Global Gender Gap Index*. (Глобальный индекс гендерного неравенства).
- Palma – *Palma ratio*. (Соотношение Пальма).
- QR – *Quintile Ratio*. (Квантильное соотношение).
- DR – *Death Rate*. (Уровень смертности).
- GDP – *Gross Domestic Product per capita*. (ВВП на душу населения).
- Inf – *Inflation rate*. (Уровень инфляции).
- LEB – *Life Expectancy at Birth*. (Ожидаемая продолжительность жизни).
- Pol - *Pollution rate*. (Уровень загрязнения).
- CinAt - *Cinema Attendance*. (Посещаемость кинотеатров).
- Hexр - *Health Expenditure*. (Стоимость здравоохранения, выраженная в долларах США).
- Happy - *World Happiness*. (Уровень счастья).

Подробное описание каждого из них представлено в приложении. Если совместить полученный список с описанной выше Европейской моделью «8+1», то наблюдается следующая картина:

- | | |
|--|--|
| 1. <i>Material living conditions:</i> | Gini, PR, QR, TTR, TTRate, ITRate, TR, Palma |
| 2. <i>Productive or main activity:</i> | UR |
| 3. <i>Health:</i> | LEB, DR, Hexр |

4. *Education:* EYoS, MYoS, HDI
5. *Leisure and social interactions:* CinAt
6. *Economic and physical safety:* IH, DCPS, PD, GD, PSaAV
7. *Governance and basic rights:* GE, IHDI, GGGI, GI, VA, CC, RQ, EF, WPF, MF
8. *Natural and living environment:* Pol
9. *Overall experience of life:* Happy

Дополнительно к указанному структурированному списку был добавлен пункт – глобальные показатели (*Global Indexes*), к которому были отнесены общие макроэкономические показатели, в силу их социальной, политической и экономической значимости:

10. *Global indexes:* GDP, FiW, Inf

Итоговая таблица имеет размерность 236x35 и не представлена в работе ввиду громоздкости.

1.2 Обработка данных

Из представленного выше распределения показателей видно, что сферы описываются различным количеством индексов. Более того, ввиду неполноты данных, имеет место следующая закономерность: размерность выборки тем выше, чем меньше индексов рассмотрено. Под размерностью выборки здесь будем понимать количество стран, по которым имеются данные по всем рассмотренным показателям. Уменьшение количества используемых индексов может привести к неполноте представления всех аспектов жизнедеятельности человека. С учетом этой закономерности в работе была применена следующая методология обработки данных.

Обработка проводилась в два этапа. На первом этапе исключались индексы, имеющие наибольшую корреляцию с другими, но при условии, что

оставшиеся будут представлять все сферы жизнедеятельности. На втором, более строгом, этапе исключались показатели, уровень корреляции которых составлял величину, большую 0.6. Пороговое значение соответствует среднему значению шкалы Чеддока [18]. При исключении предпочтение отдавалось тому показателю, который описывал большее количество стран. По итогам второго этапа, условие представления всех сфер жизнедеятельности не должно быть обязательно выполнено. Вообще говоря, формирование этого ограничения продиктовано стремлением провести исследование на максимально очищенном от внутренних зависимостей наборе данных, который описывал бы основные аспекты жизни.

На основе описанного принципа по итогам первого этапа исключены:

VA, GE, RQ, HDI, EYoS, ITRate, WPF, GD, Gini, GI, IHDI, TR, QR, CC.

По итогам второго – дополнительно:

GDP, Hexp, Pol, FiW, Palma, LEB, CinAt, PSaAV.

Корреляционные матрицы и сформированные данные, соответствующие этапам, представлены в приложении. Таким образом получено три набора данных:

1. **Начальные данные** – матрица, включающая все представленные индексы.
2. **Исходные данные** – матрица, полученная после первого этапа исключения индексов.
3. **Отсеянные данные** – матрица, сформированная после второго этапа исключения индексов.

Глава 2. Регрессионный анализ

В настоящей главе будет описана методология построения моделей и представлен регрессионный анализ данных, проведенный на их основе.

2.1 Выбор зависимого фактора

В первую очередь возникает вопрос о выборе зависимого фактора. В качестве такового, очевидно, должен выступать показатель, который максимально полно охватывает понятие качества жизни. На сегодняшний день существует несколько специальных индексов, часть из которых была упомянута ранее:

1. Индекс уровня счастья (*World Happiness*)
2. Международный индекс счастья (*Happy Planet Index*)
3. Индекс качества жизни (*Quality of Life Index*)
4. Индекс удовлетворенностью жизнью (*Satisfaction with Life Index*)

Все они описывают схожие понятия, но с использованием различных методик. Чем же они отличаются, и какой из них наиболее удобен для текущего исследования?

Индекс уровня счастья, как упоминалось ранее, был впервые вычислен в 2012 году. Ежегодно отчеты публикуются на специальном сайте [8]. Данный показатель охватывает примерно 155 стран и формируется, исходя из проводимых опросов. При его расчете предполагается, что искомый фактор зависит от ВВП, уровня социальной поддержки, ожидаемой продолжительности здоровой жизни, свободы принятия решений, щедрости и ощущения уровня коррупции в стране.

Международный индекс счастья впервые был рассчитан в 2006 году и на тот момент охватывал 178 стран мира. Показатель оценивает устойчивое благосостояние (*sustainable wellbeing*) [17]. Он сравнивает, насколько

эффективно жители разных стран используют природные ресурсы для достижения долгой и благополучной жизни. Предполагается, что индекс зависит от таких показателей, как ожидаемая продолжительность жизни, оценка удовлетворенностью жизнью (формируется из опроса), неравенство исходов [17] и экологический след [21]. В 2016 году показатель описывал 140 стран мира.

Индекс качества жизни был вычислен в 2013 году. Он разработан компанией Economist Intelligence Unit, которая использовала результаты опроса на тему «удовлетворенностью жизнью» для вычисления весовых коэффициентов, используемых в многофакторной регрессионной модели [19]. Эта модель включает девять регрессоров, которые представляют следующие сферы жизни: материальное благосостояние, здоровье, политическую стабильность и защиту, семейную жизнь, общественную жизнь, климат и географию, гарантии работы, политические свободы и гендерное неравенство. В итоговой таблице было представлено только 80 государств, ввиду нехватки данных по остальным.

Индекс удовлетворенностью жизнью предложен социальным психологом университета Лестера Адрианом Уайтом [20]. Он учитывает не только результаты опроса, но и социальное и экономическое развитие страны. Последний раз индекс был вычислен в 2016 году, он предоставляет данные по 178 странам мира.

Сравним вышеуказанные показатели. Сразу можно исключить из последующего рассмотрения индекс качества жизни, так как наиболее новые результаты датируются лишь 2013 годом, и описывают только 80 государств, что противоречит принципам актуальности и полноты нынешнего исследования. Второй показатель формируется исходя из небольшого числа агрегированных факторов, которые лишь косвенно представляют основные сферы жизнедеятельности человека. Среди оставшихся показателей более

предпочтительным является индекс уровня счастья. Несмотря на меньшую итоговую выборку, в открытом доступе имеется более полная информация о его составляющих и способе вычисления, в отличие от четвертого показателя.

Естественным является возникновение вопроса о том, можно ли воспринимать «уровень счастья» как показатель «качества жизни»? Ведь уровень развития страны не всегда отражает уровень счастья населения [22]. Однако, авторы [12] учитывали не только субъективную удовлетворенность жизнью, но и такие социально-экономические показатели, как ВВП, уровень социальной поддержки и др. Также, по итогам проведенного кластерного анализа, который подробно описан в главе 3, выявлена закономерность: чем выше социально-экономические показатели государства, тем больший рейтинг она получает. Таким образом, рассмотренный показатель можно считать пригодным для использования в качестве зависимого фактора.

2.2 Методология построения моделей

Ввиду того, что в работе рассматривается большое количество индексов, построение линейных многофакторных регрессионных моделей проводится методом последовательного включения факторов в модель [23]. Критерием отбора служит прирост скорректированного коэффициент детерминации R_{adj}^2 . Он является более показательным, в сравнении с коэффициентом детерминации R^2 , так как имеет возможность уменьшаться при добавлении малозначимых факторов [24].

Выбор порогового значения прироста R_{adj}^2 , вообще говоря, остается за исследователем. В текущей работе оно установлено на уровне 0.04, так как увеличение этого значения больше, чем на 0.01, приводит к тому, что объясняемая переменная описывается лишь одним фактором, не зависимо от того, по какому из наборов данных была построена модель. Это является неприемлемым результатом с точки зрения интерпретируемости. При

уменьшении порогового значения в модель включается большее число факторов, однако часть коэффициентов теряет свою значимость, вследствие чего требуется построение новой регрессии.

Ввиду того, что представленные в текущей работе индексы содержат данные по различающимся наборам стран, имеет место зависимость размера пригодной для использования выборки от набора рассмотренных показателей. Поэтому в работе будет реализовано два подхода при построении моделей:

1. Из используемой выборки изначально удаляются все наблюдения, содержащие пропуск хотя бы по одному из рассмотренных показателей. Далее строится регрессионная модель, согласно описанному выше методу.

2. На первом этапе исключение наблюдений проводится по ходу построения регрессии. Другими словами, на каждом шаге построения модели мы будем брать максимально возможное число стран для данного, рассмотренного именно на этом шаге набора индексов. Это приводит к тому, что величина R_{adj}^2 вычисляется по выборкам разной размерности и не может адекватно служить мерой сравнения качества моделей. Таким образом мы получаем эвристическое решение, которое может не являться наилучшим в своем классе [24]. На втором этапе мы пойдем в «обратную» сторону, а именно: зафиксированный, полученный после первого этапа список стран дополняется показателями, которые имеют данные по нему. Так формируется статическая выборка, по которой будет строиться итоговая модель методом последовательного включения факторов.

Описанные подходы будут применены к сформированным в Главе 1 наборам данных. Все вычисления проведены в пакете R [25]. Помимо использования стандартных библиотек, в процессе решения задач, поставленных в работе,

был написан ряд специальных функций. В частности, реализован второй подход к формированию модели. Более того, были модифицированы методы `gqtest()` и `bptest()` из пакета «`lmtest`» с целью автоматизации процесса анализа регрессионных моделей. Коды с подробным описанием функций представлен в приложении.

2.3 Построение и верификация регрессионных моделей

2.3.1 Реализация первого подхода

Напомним, что в пункте 1.2 Главы 1 было сформировано три набора данных:

1. **Начальные данные**
2. **Исходные данные**
3. **Отсеянные данные**

В текущем пункте реализуем первый подход и проведем анализ полученных моделей.

1. **Линейная многофакторная регрессия по начальным данным.**

Исходная выборка имеет размерность 231x35. После исключения наблюдений, содержащих хотя бы один пропуск, размерность составила 68x35. Методом последовательного включения факторов в модель был получен следующий результат:

Таблица. Порядок включения факторов в модель 1.1

Фактор	включен: GDP	включен: UR	включен: MF	включен: DR	остановка
UR	0.0892	0.6541	использован	использован	использован
IN	0.0155	0.6108	0.6547	0.7042	0.7532
PSaAV	0.3482	0.6055	0.6562	0.7032	0.7553
TTR	0.2183	0.6043	0.6544	0.7034	0.7558
VA	0.3843	0.6113	0.6668	0.7033	0.7528
GE	0.5347	0.6100	0.6617	0.7086	0.7622

CC	0.5063	0.6123	0.6596	0.7086	0.7563
RQ	0.4642	0.6112	0.6638	0.7067	0.7585
FiW	0.2653	0.6060	0.6610	0.7030	0.7524
HDI	0.5091	0.6075	0.6654	0.7117	0.7785
EYoS	0.4127	0.6021	0.6569	0.7081	0.7656
MYoS	0.2327	0.6057	0.6510	0.7083	0.7568
TTRate	-0.0138	0.6105	0.6526	0.7030	0.7531
ITRate	0.2038	0.6046	0.6538	0.7072	0.7527
DCPS	0.2222	0.6017	0.6499	0.7049	0.7526
EF	0.3613	0.6139	0.6615	0.7093	0.7549
WPF	0.3081	0.6117	0.6682	0.7042	0.7565
PD	-0.0047	0.6123	0.6508	0.7048	0.7546
GD	-0.0083	0.6206	0.6547	0.7085	0.7596
MF	0.3601	0.6508	0.7076	использован	использован
Gini	0.0184	0.6334	0.6946	0.7321	0.7558
GI	0.3598	0.6016	0.6528	0.7030	0.7796
IHDI	0.4325	0.6016	0.6497	0.7030	0.7726
TR	0.0237	0.6018	0.6503	0.7049	0.7542
GGGI	0.3198	0.6251	0.6686	0.7088	0.7661
Palma	-0.0013	0.6268	0.6845	0.7184	0.7527
QR	0.0125	0.6169	0.6761	0.7156	0.7525
DR	-0.0070	0.6437	0.6751	0.7563	использован
GDP	0.6076	использован	использован	использован	использован
Inf	0.0856	0.6042	0.6577	0.7098	0.7528
LEB	0.4816	0.6261	0.6871	0.7402	0.7757
Pol	0.4100	0.6070	0.6543	0.7034	0.7730
CinAt	0.5134	0.6245	0.6615	0.7069	0.7526
Hexp	0.5480	0.6044	0.6489	0.7030	0.7547

Модель 1.1:

$$QL = 4.8 + 3.699e^{-05} \cdot GDP - 3.587e^{-02} \cdot UR + 2.156e^{-02} \cdot MF - 8.324e^{-02} \cdot DR.$$

Таким образом, мы получили зависимость оценки качества жизни (здесь QL) от следующих факторов: ВВП, уровень безработицы, индекс моральных свобод и уровень смертности. Интерпретация результата будет представлена далее в пункте 2.4. Здесь же проведем верификацию полученной модели.

Как известно, модель и данные должны удовлетворять основным постулатам регрессионного анализа [23, 24]. Проверим выполнение этих условий.

1) *Доверительные интервалы для коэффициентов:*

Доверительные интервалы для коэффициентов модели 1.1			
Коэффициент	2.5%	97.5%	
Своб. Член	4.280	5.320	значим
GDP	2.808e-05	4.590e-05	значим
UR	-5.794e-02	-1.381e-02	значим
MF	1.244e-02	3.068e-02	значим
DR	-1.280e-01	-3.844e-02	значим

Ни один из представленных в таблице интервалов не содержит ноль, что говорит о статистической значимости всех коэффициентов уравнения.

2) *Нормальность распределения остатков:*

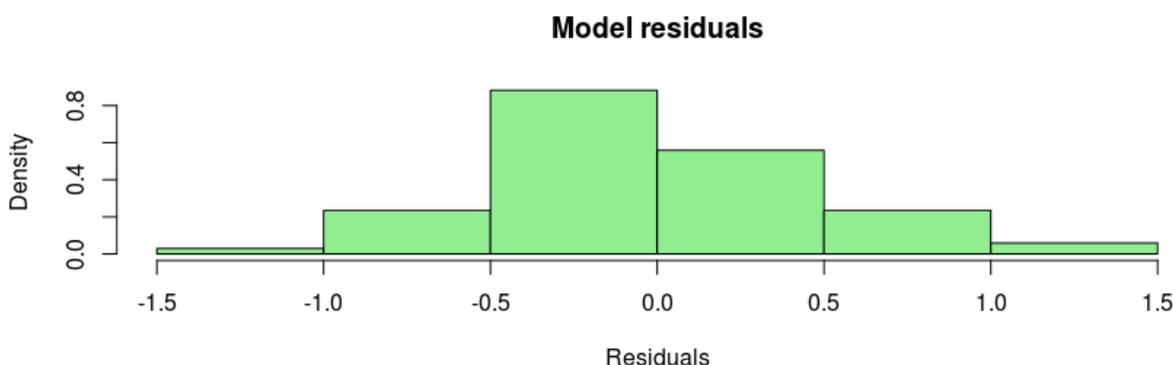


Рисунок 1. Гистограмма частот остатков модели 1.1

По гистограмме частот (см. Рисунок 1) и графику квантилей (см. Рисунок 2) можно сделать предположение о том, что остатки распределены нормально. Однако для более строгой проверки проведем несколько тестов.

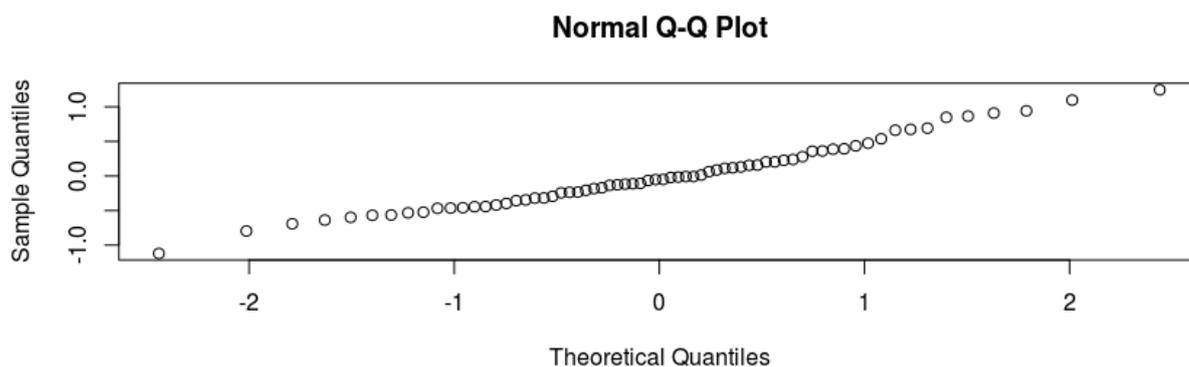


Рисунок 2. График квантилей модели 1.1

Сформулируем нулевую гипотезу: анализируемая выборка происходит из нормально распределенной генеральной совокупности. Уровень значимости примем равным 0.05. Нулевую гипотезу будем считать отвергнутой, если значение p-value будет меньше установленного уровня - 0.05, и принятой в ином случае.

Тест Шапиро-Уилка: $W = 0.978$, $p\text{-value} = 0.263$.

Тест Андерсона-Дарлинга: $A = 0.53$, $p\text{-value} = 0.17$

Тест Колмогорова-Смирнова: $D = 0.081$, $p\text{-value} = 0.337$

Критерий хи-квадрат Пирсона: $P = 10.294$, $p\text{-value} = 0.245$

Нулевая гипотеза принимается по итогам всех тестов. Подробную информацию по каждому из них можно найти в трудах [26, 27, 28, 29, 30].

3) Гомоскедастичность:

Нулевая гипотеза: дисперсии равны. Уровень значимости – 0.05. Как и ранее, нулевая гипотеза принимается, если величина p-value больше установленного уровня значимости.

Тест Гольдфельда-Квандта:

Зависимость от первого фактора (GDP): $GQ = 0.414$, $p\text{-value} = 0.990$

Зависимость от второго фактора (DR): $GQ = 0.806$, $p\text{-value} = 0.717$

Зависимость от третьего фактора (MF): $GQ = 0.625$, $p\text{-value} = 0.894$

Зависимость от четвертого фактора (UR): $GQ = 0.754$, $p\text{-value} = 0.774$

Тест Бройша-Пагана: $BP = 5.212$, $p\text{-value} = 0.266$

Тест Уайта: $W = 12.559$, $p\text{-value} = 0.562$

Тест Гольдфельда-Квандта [24, 26] говорит о том, что отсутствует зависимость дисперсий от рассмотренных факторов по отдельности. По итогам тестов Бройша-Пагана [24, 31] и Уайта [24, 27] нулевая гипотеза также принимается.

4) *Наличие систематических ошибок:*

Далее вычислим математическое ожидание остатков. В нашем случае $E = 1.618e-17$. Значение крайне близко к нулю, что подтверждает отсутствие систематических ошибок в остатках.

5) *Мультиколлинеарность:*

Вычислим определитель матрицы $X^T X$, где X - матрица, содержащая данные включенных в модель факторов. $\det(X^T X) = 1.696e + 22$, что значимо отлично от нуля и говорит об отсутствии мультиколлинеарности.

6) *Независимость остатков:*

Нулевая гипотеза: данные являются случайными. Уровень значимости – 0.05.

Тест Дарбина-Уотсона: $DW = 2.307$, $dL = 1.516$, $dU = 1.7$

Тест Льюнга-Бокса: $X = 1.801$, $p\text{-value} = 0.180$

Тест Бройша-Годфри: $LM = 1.802$, $p\text{-value} = 0.179$

В тесте Дарбина-Уотсона статистика принадлежит интервалу от $(4 - dU; 4 - dL)$, поэтому мы не можем что-либо сказать о принятии или отрицании нулевой гипотезы [23, 24]. *p*-value тестов Льюнга-Бокса [24] и Бройша-Годфри [32] больше 0.05, что говорит о независимости остатков.

Таким образом, модель удовлетворяет всем постулатам регрессионного анализа и является пригодной для дальнейшего использования. Средняя ошибка аппроксимации составляет 6,69%, скорректированный коэффициент детерминации – 0.76.

2. Линейная многофакторная регрессия по исходным данным.

Размерность исходной выборки после исключения пропусков составляет 72x21. Проведем действия, аналогичные тем, что описаны в подпункте 1 ранее. Ниже представлена таблица порядка включения факторов в модель.

Таблица. Порядок включения факторов в модель 1.2

Фактор	включен: GDP	включен: MF	включен: DR	остановка
UR	0.0687	0.6494	0.7057	0.7567
IH	0.0136	0.6112	0.6479	0.7185
PSaAV	0.3349	0.6054	0.6444	0.7201
TTR	0.2102	0.6037	0.6472	0.7204
FiW	0.2689	0.6068	0.6464	0.7194
MYoS	0.2458	0.6051	0.6510	0.7252
TTRate	-0.0067	0.6203	0.6521	0.7205
DCPS	0.1780	0.6021	0.6445	0.7184
EF	0.3129	0.6063	0.6469	0.7187
PD	-0.0021	0.6125	0.6562	0.7278
MF	0.3485	0.6496	использован	использован
GGI	0.3133	0.6252	0.6527	0.7362
Palma	-0.0033	0.6285	0.6568	0.7183
DR	-0.0068	0.6437	0.7223	использован
GDP	0.6075	использован	использован	использован
Inf	0.0454	0.6018	0.6444	0.7209
LEB	0.4826	0.6270	0.6670	0.7305
Pol	0.4067	0.6070	0.6447	0.7434
CinAt	0.5152	0.6248	0.6591	0.7210
Hexp	0.5336	0.6030	0.6445	0.7196

Модель 1.2:

$$QL = 4.614 + 3.938e^{-05} \cdot GDP - 2.095e^{-02} \cdot MF - 1.004e^{-01} \cdot DR.$$

В отличие от предыдущего случая, оценка качества жизни (QL) не зависит от уровня безработицы, однако повысилась весомость уровня смертности.

Ввиду аналогичности проводимого анализа, и с целью удобства представления результатов, в текущем подпункте будут представлены только графики и матрица доверительных интервалов для коэффициентов, результаты всех тестов будут описаны позднее в агрегированной таблице.

1) Доверительные интервалы для коэффициентов:

Доверительные интервалы для коэффициентов модели 1.2			
Коэффициент	2.5%	97.5%	
Своб. Член	4.112	5.115	значим
GDP	3.031e-05	4.844e-05	значим
MF	1.172e-02	3.018e-02	значим
DR	-1.462e-01	-5.452e-02	значим

Все коэффициенты статистически значимы.

2) Нормальность распределения остатков:

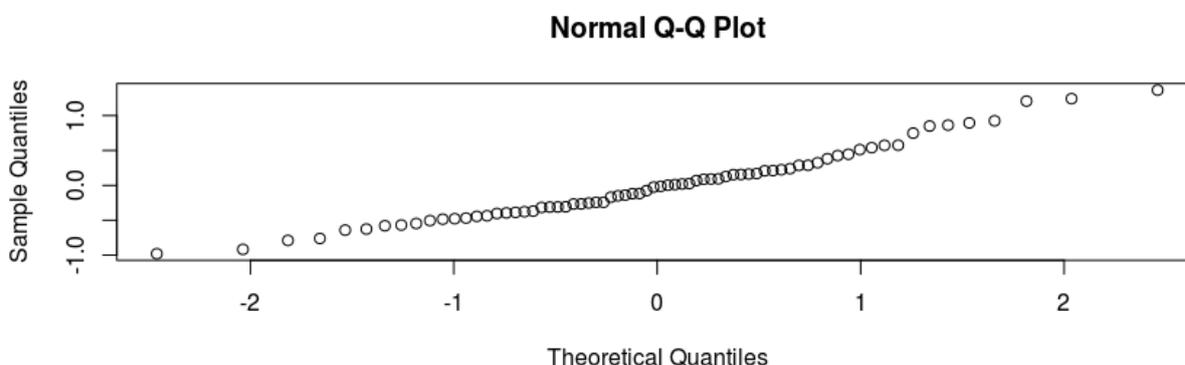


Рисунок 3. График квантилей модели 1.2

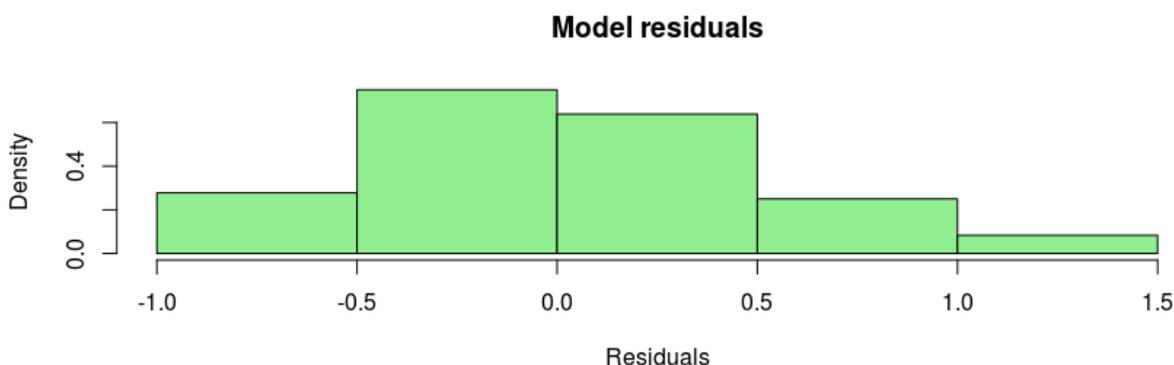


Рисунок 4. Гистограмма частот остатков модели 1.2

По гистограмме частот остатков заметна небольшая асимметричность, тем не менее, принимая во внимание график квантилей, можно сделать предположение о нормальности распределения остатков. Это подтверждается и результатами тестов, которые будут представлены в агрегированной таблице далее.

3. Линейная многофакторная регрессия по отсеянным данным.

Исходная размерность данных составляет 231x13, после исключения пропусков 112x13. Как и ранее, ниже представлена таблица порядка включения факторов в модель.

Таблица. Порядок включения факторов в модель 1.3

Фактор	включен: MYoS	включен: DR	включен: MF	остановка
UR	0.0212	0.4253	0.5325	0.6230
IH	-0.0007	0.3489	0.4987	0.5827
TTR	0.1774	0.3664	0.5546	0.6079
MYoS	0.3496	использован	использован	использован
TTRate	-0.0054	0.3515	0.5147	0.5861
DCPS	0.2135	0.3964	0.5159	0.5904
EF	0.3298	0.4282	0.5352	0.6010
PD	-0.0058	0.3453	0.4972	0.5868
MF	0.1437	0.3658	0.5865	использован
GGGI	0.1990	0.3862	0.5735	0.6114
DR	0.0430	0.5013	использован	использован
Inf	0.0647	0.3757	0.5084	0.5871

Модель 1.3:

$$QL = 4.087 + 2.186e^{-01} \cdot MYoS - 1.809e^{-01} \cdot DR + 2.085e^{-02} \cdot MF.$$

Полученная модель схожа с моделью 2, построенной по исходному набору данных. Главное отличие заключается в том, что показатель ВВП, в данном случае, заменила величина средней длительности обучения, ввиду отсутствия первого в используемой выборке.

1) Доверительные интервалы для коэффициентов:

Доверительные интервалы для коэффициентов модели 1.3			
Коэффициент	2.5%	97.5%	
Своб. член	3.573	4.601	значим
MYoS	0.168	0.269	значим
MF	0.012	0.029	значим
DR	-0.227	-0.134	значим

Все коэффициенты значимы.

2) Нормальность распределения остатков:

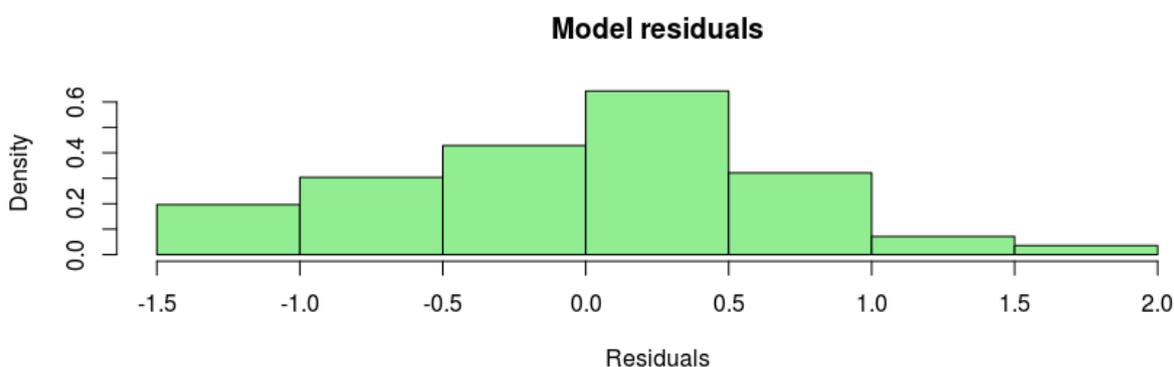


Рисунок 5. Гистограмма частот остатков модели 1.3

По гистограмме частот видна некоторая асимметричность, это отражается и на графике квантилей (см. Рисунок 6). Поэтому в данном случае особенно требуется проведение дополнительных тестов.

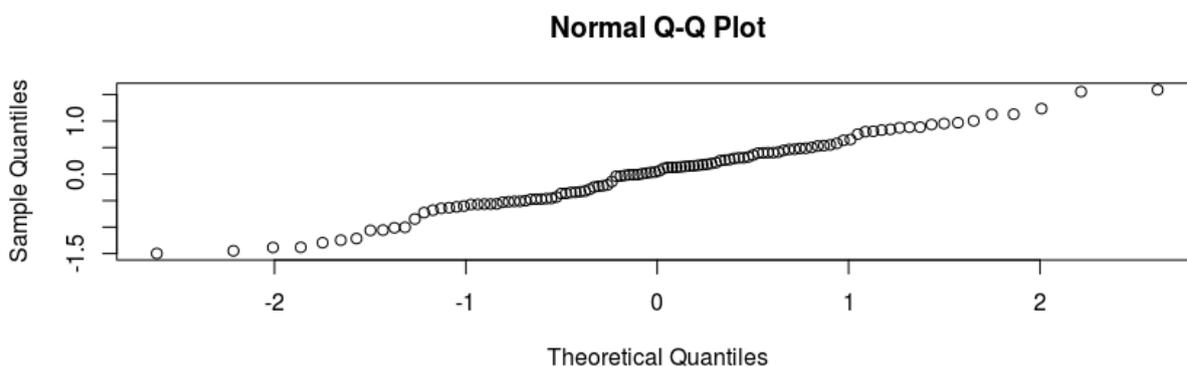


Рисунок 6. График квантилей модели 1.3

4. Результаты верификации моделей.

Результаты верификации всех трех моделей представлены в обобщенной таблице ниже.

Таблица. Результаты верификации моделей

Название теста	Модель 1.1	Модель 1.2	Модель 1.3
<i>Значимость коэффициентов</i>	Все коэффициенты значимы	Все коэффициенты значимы	Все коэффициенты значимы
Нормальность распределения остатков			
<i>Тест Шапиро-Уилка</i>	W = 0.978, p-value = 0.263; пройден	W = 0.969, p-value = 0.074; пройден	W = 0.988, p-value = 0.415; пройден
<i>Тест Андерсона-Дарлинга</i>	A = 0.53, p-value = 0.17; пройден	A = 0.636, p-value = 0.094; пройден	A = 0.396, p-value = 0.364; пройден
<i>Тест Колмогорова-Смирнова</i>	D = 0.081, p-value = 0.337; пройден	D = 0.086, p-value = 0.214; пройден	D = 0.065, p-value = 0.292; пройден
<i>Критерий χ^2 Пирсона</i>	P = 10.294, p-value = 0.245; пройден	P = 10.333, p-value = 0.324; пройден	P = 15.5, p-value = 0.161; пройден
Гомоскедастичность			
<i>Тест Гольдфельда-Квандта</i>	GDP: p-value = 0.99; пройден DR: p-value = 0.717; пройден MF: p-value = 0.894; пройден UR: p-value = 0.774; пройден	GDP: p-value = 0.989; пройден MF: p-value = 0.856; пройден DR: p-value = 0.717; пройден	MYoS: p-value = 0.384; пройден DR: p-value = 0.268; пройден MF: p-value = 0.788; пройден
<i>Тест Бройша-Пагана</i>	BP = 5.212, p-value = 0.266; пройден	BP = 2.493, p-value = 0.477; пройден	BP = 1.91, p-value = 0.591; пройден
<i>Тест Уайта</i>	W = 12.559, p-value = 0.562; пройден	W = 6.735, p-value = 0.665; пройден	W = 11.755, p-value = 0.228; пройден
Наличие систематических ошибок			
<i>Среднее остатков</i>	= 1.618e-17; пройден	= 1.678e-17; пройден	= 1.235e-16; пройден
Мультиколлинеарность			
<i>det(X^TX)</i>	= 1.696e+22; пройден	= 9.507e+18; пройден	= 2.211e+12; пройден

Независимость остатков			
Тест Дарбина-Уотсона	DW = 2.307, dL = 1.516, dU = 1.7, 4-dU < DW < 4-dL; не определено	DW = 2.283, dL = 1.561, dU = 1.675, DW < 4-dU; пройден	DW = 2.153, dL = 1.656, dU = 1.728, DW < 4-dU; пройден
Тест Льюнга-Бокса	X ² = 1.801, p-value = 0.180; пройден	X ² = 1.693, p-value = 0.193; пройден	X ² = 1.134, p-value = 0.287; пройден
Тест Бройша-Годфри	LM = 1.802, p-value = 0.179; пройден	LM = 1.635, p-value = 0.201; пройден	LM = 1.126, p-value = 0.289; пройден
Средняя ошибка аппроксимации			
A	6.69%	7.09%	10.06%

В представленной таблице цвет ячейки является индикатором принятия нулевой гипотезы, зеленый означает ее достоверность.

Видно, что все три модели удовлетворяют постулатам регрессионного анализа, а значит, и условиям теоремы Гаусса-Маркова [24], и являются пригодными для дальнейшего использования. Индикатором качества моделей являются средняя ошибка аппроксимации и скорректированный коэффициент детерминации. Видно, что наименьшей точностью обладает модель 1.3, построенная по отсеянному набору, более того, ей соответствует самое маленькое значение $R_{adj}^2 - 0.587$. Отличие первых двух моделей не так велико. Средняя ошибка аппроксимации отличается всего на 0.4%, а скорректированного коэффициента детерминации – на 0.034, составляющего 0.756 и 0.722 у первой и второй моделей соответственно. Однако, если обратить внимание на включенные факторы:

$$1.1 \quad QL = 4.8 + 3.699e^{-05} \cdot GDP - 3.587e^{-02} \cdot UR + 2.156e^{-02} \cdot MF - 8.324e^{-02} \cdot DR,$$

$$1.2 \quad QL = 4.614 + 3.938e^{-05} \cdot GDP - 2.095e^{-02} \cdot MF - 1.004e^{-01} \cdot DR,$$

становится очевидно, что первая модель наиболее полно отражает аспекты реальной жизни, другими словами, ее проще интерпретировать с точки зрения социально-экономического анализа. Таким образом, первая модель, построенная по начальному набору данных, является более предпочтительной в описанном выше смысле.

2.3.2 Реализация второго подхода

Согласно описанному ранее второму подходу, на первом этапе будет получено эвристическое решение, результатом которого является фиксированный набор стран. Далее, на втором этапе, этот набор дополняется допустимыми факторами. Под допустимостью здесь будем понимать отсутствие пропусков в данных индексов для рассмотренного списка государств. Уже по вновь сформированной выборке строится итоговая линейная регрессионная модель.

Такой подход к формированию модели продиктован стремлением использовать данные с наибольшим количеством наблюдений. Для удобства описания результатов, структура пункта будет аналогична пункту 2.3.1.

1. Линейная многофакторная регрессия по начальным данным

Используемый набор данных имеет размерность 231x35. Таблица порядка включения факторов в модель при получении эвристического решения:

Таблица. Порядок включения факторов в «промежуточную» модель

Фактор	включен: HDI	включен: Нехр	остановка
UR	0.0077	0.7143	0.7479
IH	-0.0041	0.6842	0.7355
PSaAV	0.3223	0.6756	0.7176
TTR	0.1710	0.6711	0.7214
VA	0.3374	0.6832	0.7177
GE	0.5357	0.6841	0.7178
CC	0.4652	0.6890	0.7177
RQ	0.4210	0.6766	0.7178
FiW	0.2707	0.6790	0.7176
HDI	0.6751	использован	использован
EYoS	0.5514	0.6736	0.7198
MYoS	0.5015	0.6780	0.7228

TTRate	-0.0028	0.6756	0.7202
ITRate	0.0169	0.6783	0.7130
DCPS	0.3161	0.6762	0.7230
EF	0.2990	0.6611	0.7099
WPF	0.1238	0.6810	0.7189
PD	0.0009	0.6418	0.6999
GD	0.0347	0.6769	0.7255
MF	0.2012	0.6529	0.6898
Gini	0.0460	0.6730	0.7226
GI	0.4742	0.6590	0.7106
IHDI	0.6278	0.6578	0.7050
TR	0.0475	0.6214	0.6789
GGGI	0.2435	0.6800	0.7066
Palma	-0.0005	0.6713	0.7217
QR	0.0164	0.6689	0.7177
DR	0.0660	0.6902	0.7442
GDP	0.0046	0.6664	0.7104
Inf	0.0684	0.6716	0.7112
LEB	0.5948	0.6768	0.7217
Pol	0.3507	0.5925	0.6424
CinAt	0.4185	0.5823	0.6434
Hexp	0.4723	0.7194	использован

Эвристическое решение:

$$QL = 1.792 + 4.834 \cdot HDI + 1.487e^{-04} \cdot Hexp.$$

Данное уравнение построено по выборке размерностью 149x3. Таким образом получен фиксированный набор из 149 государств. Полный их перечень представлен в приложении.

Помимо уже включенных индекса развития человеческого потенциала (*HDI*) и стоимости здравоохранения (*Hexp*), данные были пополнены следующими допустимыми факторами: *IH*, *PSaAV*, *VA*, *GE*, *CC*, *RQ*, *FiW*, *EYoS*, *MYoS*. Полученная выборка имеет размерность 149x12. Проведем построение регрессионной модели. Соответствующая таблица порядка включения факторов представлена ниже:

Таблица. Порядок включения факторов в модель 2.1

Фактор	включен: HDI	включен: Hexp	остановка
HDI	0.6820	использован	использован
Hexp	0.4723	0.7194	использован
IH	-0.0035	0.6907	0.7355

PSaAV	0.3325	0.6830	0.7176
VA	0.3426	0.6897	0.7177
GE	0.5710	0.6931	0.7178
CC	0.4794	0.6978	0.7177
RQ	0.4398	0.6849	0.7178
FiW	0.2707	0.6851	0.7176
EYoS	0.5535	0.6809	0.7198
MYoS	0.5037	0.6855	0.7228

Модель 2.1:

$$QL = 1.792 + 4.834 \cdot HDI + 1.487e^{-04} \cdot Hexp.$$

Видно, что вновь построенная регрессия совпала, возможно, случайно, с полученным ранее эвристическим решением. В текущем подпункте будет размещено лишь несколько графиков и таблиц, более строгие результаты анализа расположены в подпункте 4.

1) Доверительные интервалы для коэффициентов:

Доверительные интервалы для коэффициентов модели 2.1			
Коэффициент	2.5%	97.5%	
Своб. член	1.246	2.339	значим
HDI	3.998	5.670	значим
Hexp	8.398e-05	2.133e-04	значим

Все коэффициенты статистически значимы.

2) Нормальность распределения остатков:

По гистограмме частот (см. Рисунок 7) и графику квантилей (см. Рисунок 8) можно сделать предположение о том, что остатки получены из нормально распределенной генеральной совокупности.



Рисунок 7. Гистограмма частот остатков модели 2.1

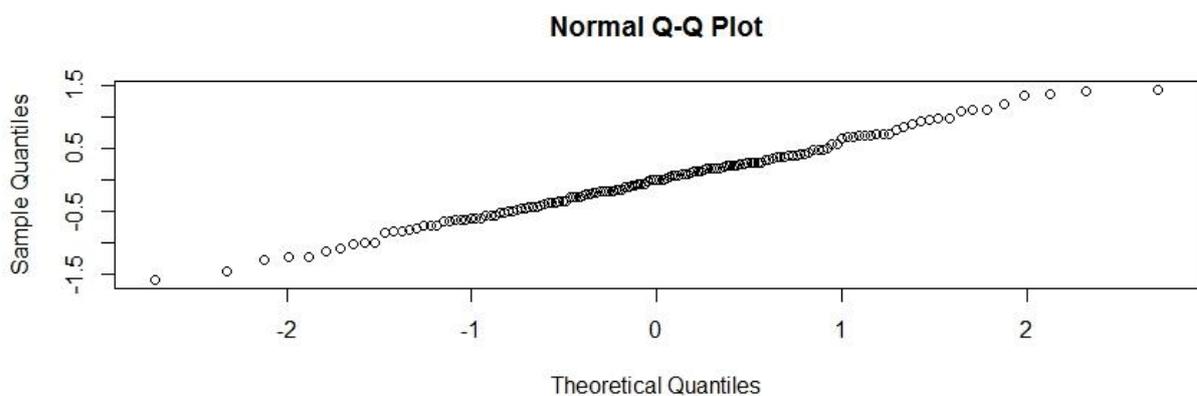


Рисунок 8. График квантилей модели 2.1

2. Линейная многофакторная регрессия по исходным данным

Размерность используемых данных 231x21. Таблица порядка включения факторов при формировании эвристического решения:

Таблица. Порядок включения факторов в «промежуточную» модель

Фактор	включен: LEB	включен: Нехр	остановка
UR	0.0077	0.6132	0.6976
IH	-0.0041	0.6005	0.7007
PSaAV	0.3223	0.6419	0.6969
TTR	0.1710	0.6039	0.6828
FiW	0.2707	0.6224	0.6856
MYoS	0.5015	0.6307	0.6979
TTRate	-0.0028	0.5910	0.6847
DCPS	0.3161	0.6002	0.6866

EF	0.2990	0.6269	0.6998
PD	0.0009	0.5701	0.6811
MF	0.2012	0.6149	0.6768
GGGI	0.2435	0.6370	0.6847
Palma	-0.0005	0.5956	0.7055
DR	0.0660	0.5939	0.6859
GDP	0.0046	0.5928	0.6870
Inf	0.0684	0.6087	0.6871
LEB	0.5948	использован	использован
Pol	0.3507	0.5463	0.6181
CinAt	0.4185	0.5837	0.6571
Hexр	0.4723	0.6850	использован

Приближенное решение:

$$QL = -2.455e^{-01} + 7.506e^{-02} \cdot LEB + 2.081e^{-04} \cdot Hexр.$$

Ему соответствует набор из 144 государств, который размещен в приложении. Список допустимых факторов: *LEB, Hexр, UR, IH, PSaAV, TTR, FiW, MYoS, DR, GDP, Inf*. Таким образом сформирована выборка размерностью 144x12. Проведем построение модели:

Таблица. Порядок включения факторов в модель 2.2

Фактор	включен: LEB	включен: Hexр	остановка
LEB	0.5969	использован	использован
Hexр	0.4628	0.6850	использован
UR	0.0120	0.6192	0.6976
IH	-0.0023	0.6014	0.7007
PSaAV	0.3214	0.6462	0.6969
TTR	0.1818	0.6073	0.6828
FiW	0.2591	0.6225	0.6856
MYoS	0.4983	0.6366	0.6979
DR	0.0670	0.5959	0.6859
GDP	0.0046	0.5948	0.6870
Inf	0.0643	0.6103	0.6871

Как и в предыдущем случае, полученная модель совпадает с эвристическим решением. Ниже представлена матрица доверительных интервалов для коэффициентов:

Доверительные интервалы для коэффициентов модели			
Коэффициент	2.5%	97.5%	
Своб. Член	-1.261	0.770	не значим
LEB	6.031e-02	8.982e-02	значим
Hexр	1.437e-04	2.726e-04	значим

Интервал, соответствующий свободному члену, содержит ноль, что говорит о статистической незначимости коэффициента. Построим модель без свободного члена.

Модель 2.2:

$$QL = 7.153e^{-02} \cdot LEB + 2.160e^{-04} \cdot Hexр.$$

1) Доверительные интервалы для коэффициентов:

Доверительные интервалы для коэффициентов модели 2.2			
Коэффициент	2.5%	97.5%	
LEB	6.971e-02	7.335e-02	значим
Hexр	1.605e-04	2.715e-04	значим

На этот раз все коэффициенты значимы.

2) Нормальность распределения остатков:

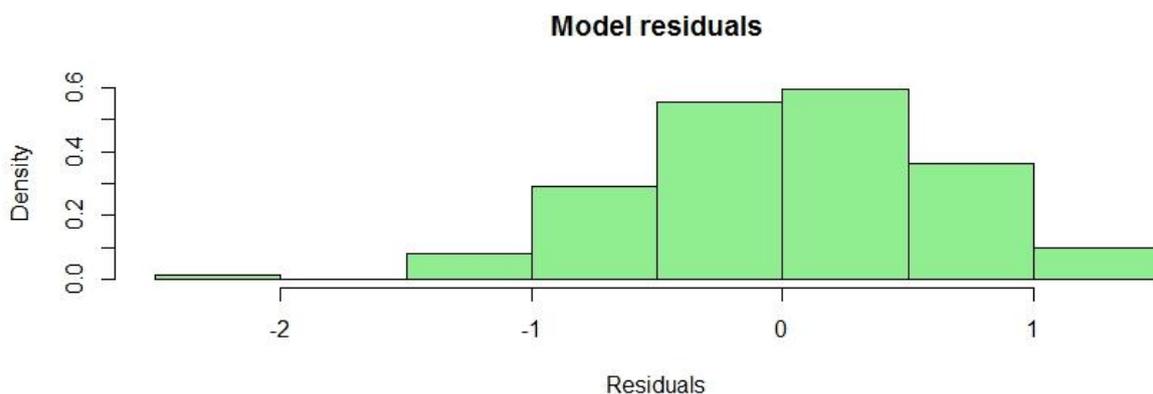


Рисунок 9. Гистограмма частот остатков модели 2.2

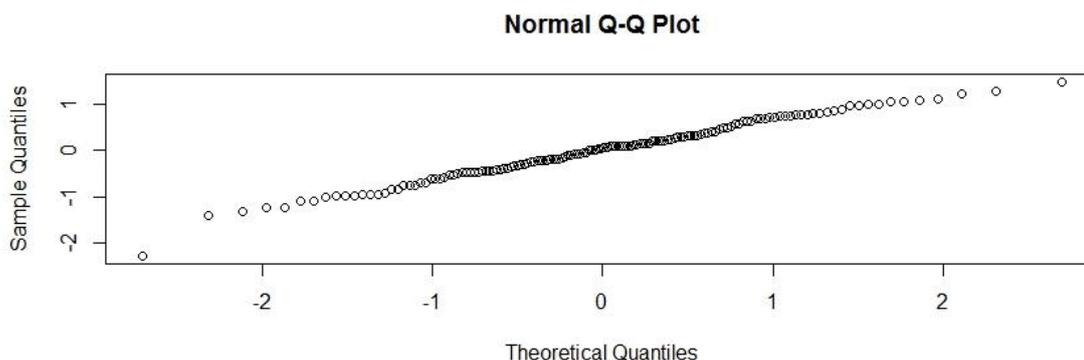


Рисунок 10. График квантилей модели 2.2

По представленным графикам можно смело утверждать, что остатки распределены нормально.

3. Линейная многофакторная регрессия по отсеянным данным

Применим описанный подход к третьему набору данных. Размерность используемой матрицы – 231x13. Таблица порядка включения факторов в модель для приближенного решения:

Таблица. Порядок включения факторов в «промежуточную» модель

Фактор	включен: MYoS	включен: DR	включен: MF	остановка
UR	0.0077	0.5411	0.6004	0.6408
IH	-0.0041	0.5037	0.5790	0.6189
TTR	0.1710	0.5096	0.6057	0.6318
MYoS	0.5015	использован	использован	использован
TTRate	-0.0028	0.4988	0.5792	0.6186
DCPS	0.3161	0.5400	0.5991	0.6315
EF	0.2990	0.5117	0.6051	0.6263
PD	0.0009	0.4449	0.5491	0.5944
MF	0.2012	0.4629	0.6214	использован
GCCI	0.2435	0.4863	0.6029	0.6217
DR	0.0660	0.5803	использован	использован
Inf	0.0684	0.5149	0.6005	0.6344

Эвристическое решение:

$$QL = 3.913 + 0.217 \cdot MYoS - 0.163 \cdot DR + 2.096e^{-02} \cdot MF.$$

Модели соответствует список из 129 государств, который также представлен в приложении. Допустимые, в этом случае, факторы: *MYoS*, *DR*, *MF*, *UR*, *IH*, *TTR*, *Inf*. Вновь проведем построение регрессионной модели. Таблица порядка включения факторов в итоговое уравнение:

Таблица. Порядок включения факторов в модель 2.3

Фактор	включен: MYoS	включен: DR	включен: MF	остановка
MYoS	0.4424	использован	использован	использован
DR	0.0366	0.5473	использован	использован
MF	0.1955	0.4589	0.6214	использован
UR	0.0298	0.4893	0.5718	0.6408
IH	-0.0014	0.4454	0.5475	0.6189
TTR	0.1911	0.4536	0.5791	0.6318
Inf	0.0957	0.4719	0.5735	0.6344

Модель 2.3:

$$QL = 3.913 + 0.217 \cdot MYoS - 0.163 \cdot DR + 2.096e^{-02} \cdot MF.$$

Итоговое уравнение в точности совпало с промежуточным решением.

1) Доверительные интервалы для коэффициентов:

Доверительные интервалы для коэффициентов модели 2.3			
Коэффициент	2.5%	97.5%	
Своб. Член	3.451	4.375	значим
MYoS	0.171	0.264	значим
DR	-0.206	-0.120	значим
MF	0.013	0.029	значим

Все коэффициенты значимы.

2) Нормальность распределения остатков:

По представленным графикам (см. Рисунок 11, Рисунок 12) можно сделать предположение о том, что остатки распределены нормально. Результаты тестов, как и ранее, будут приведены в обобщенной таблице далее.

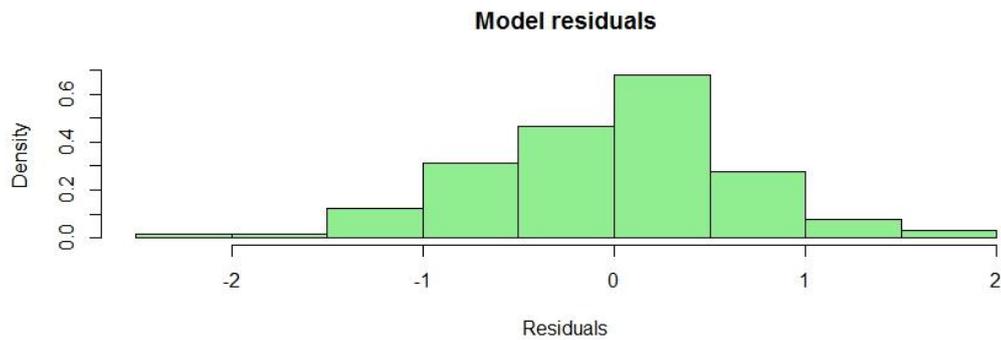


Рисунок 11. Гистограмма частот остатков модели 2.3

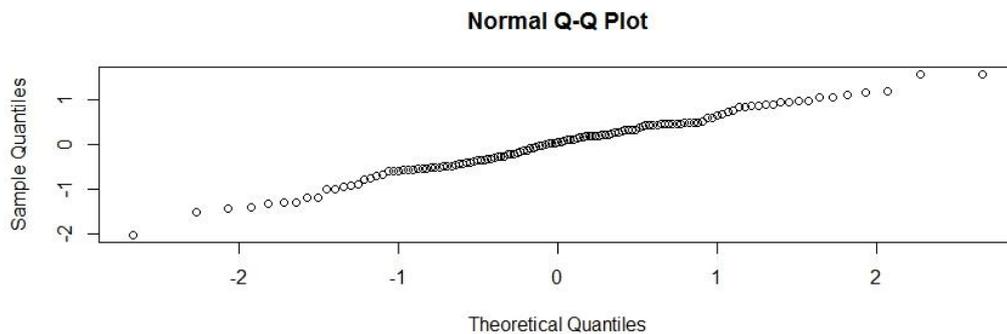


Рисунок 12. График квантилей модели 2.3

4. Результаты анализа построенных моделей.

Таблица. Результаты анализа моделей

Название теста	Модель 2.1	Модель 2.2	Модель 2.3
Значимость коэффициентов	Все коэффициенты значимы	Все коэффициенты значимы	Все коэффициенты значимы
Нормальность распределения остатков			
Тест Шапиро-Уилка	W = 0.994, p-value = 0.8754; пройден	W = 0.989, p-value = 0.367; пройден	W = 0.991, p-value = 0.547; пройден
Тест Андерсона-Дарлинга	A = 0.172, p-value = 0.927; пройден	A = 0.263, p-value = 0.698; пройден	A = 0.364, p-value = 0.434; пройден
Тест Колмогорова-Смирнова	D = 0.037, p-value = 0.876; пройден	D = 0.044, p-value = 0.724; пройден	D = 0.051, p-value = 0.568; пройден
Критерий χ^2 Пирсона	P = 10.765, p-value = 0.549; пройден	P = 12.667, p-value = 0.394; пройден	P = 17.62, p-value = 0.091; пройден
Гомоскедастичность			
Тест Гольдфельда-Квандта	HDI: p-value = 0.460; пройден Нехр: p-value = 0.471; пройден	LEB: p-value = 0.028; не пройден Нехр: p-value = 0.729; пройден	MYoS: p-value = 0.554; пройден DR: p-value = 0.649; пройден

			MF: p-value = 0.822; пройден
<i>Тест Бройша-Пагана</i>	BP = 5.655, p-value = 0.059; пройден	BP = 2.666, p-value = 0.103; пройден	BP = 2.411, p-value = 0.492; пройден
<i>Тест Уайта</i>	W = 10.936, p-value = 0.053; пройден	W = 8.582, p-value = 0.127; пройден	W = 6.296, p-value = 0.71; пройден
Наличие систематических ошибок			
<i>Среднее остатков</i>	= 1.618e-17; пройден	= 2.735e-03; пройден	= -3.097e-17; пройден
Мультиколлинеарность			
<i>det(X'X)</i>	= 1.945e+11; пройден	= 6.376; пройден	= 4.507e+12; пройден
Независимость остатков			
<i>Тест Дарбина-Уотсона</i>	DW = 2.309, dL = 1.719, dU = 1.746, DW > 4-dL; не пройден		DW = 2.109, dL = 1.681, dU = 1.744, DW < 4-dU; пройден
<i>Тест Льюнга-Бокса</i>	$\chi^2 = 3.857$, p-value = 0.0495; не пройден	$\chi^2 = 8.368$, p-value = 0.004; не пройден	$\chi^2 = 0.491$, p-value = 0.484; пройден
<i>Тест Бройша-Годфри</i>	LM = 3.825, p-value = 0.0505; пройден	LM = 8.3, p-value = 0.004; не пройден	LM = 0.495, p-value = 0.482; пройден
Средняя ошибка аппроксимации			
A	9.59%	10.31%	10.59%

В таблице выше имеет место следующая маркировка:

- Зеленый – тест пройден, принимается нулевая гипотеза (p-value > 0.05).
- Желтый – значение p-value очень близко к граничному 0.05. Тест может быть, как пройден, так и не пройден. В этом случае мы не можем что-либо смело утверждать.
- Красный – тест считается не пройденным, нулевая гипотеза отклоняется (p-value < 0.05).

По результатам проведенных тестов видно, что модель 2.2 не прошла проверку независимости остатков, то есть имеет место автокорреляция первого порядка. Также тест Гольдфельда-Квандта выявил зависимость дисперсий от фактора ожидаемой продолжительности жизни (LEB), однако тесты Бройша-Пагана и Уайта пройдены. Ввиду этого можно сделать предположение о том, что имеет место гомоскедастичность.

Модель 2.1 не прошла два из трех тестов на независимость остатков. Причем в двух случаях p-value крайне близко к граничному значению. Таким образом, предполагается, что имеет место автокорреляция первого порядка, что может быть вызвано наличием неучтенного фактора. В тестах Бройша-Пагана и Уайта величина p-value также близка к пороговому значению. Поэтому, несмотря на то, что нулевая гипотеза в обоих случаях принимается, мы не можем смело утверждать, что имеет место гомоскедастичность.

В случае первых двух моделей нарушены постулаты регрессионного анализа. Это может привести к ненадежности получаемых доверительных интервалов. Более того, значимость коэффициентов регрессии и уравнения в целом также может быть ошибочной. В итоге, данные модели не пригодны для дальнейшего использования. Для устранения этих недостатков требуется отдельное исследование и преобразование начальных данных.

В отличие от первых двух, модель 2.3, построенная по отсеянному набору данных, прошла все проверки. Однако, средняя ошибка аппроксимации составила 10,6%, что выше желаемого 10% порогового значения. Также величина скорректированного коэффициента детерминации составляет всего 0.62. Постараемся устранить эти недостатки путем проведения дополнительного анализа, а именно проверим данные на наличие выбросов.

2.3.3 Анализ на наличие выбросов

Целью анализа является выявление наблюдений, которые оказывают наибольший эффект на точность модели, а следовательно, и на ее качество. Прежде всего проверим, насколько сильно используемые данные влияют на построенную регрессию. Для этого вычислим несколько величин.

1. DFFIT

Величина $DFFIT_i$ – показывает, на сколько изменится значение, соответствующее i -му наблюдению, при условии, что оно будет удалено из выборки [33]. Для малого размера выборки значения $DFFIT_i$, большие единицы, являются подозрительными. Для большого – пороговое значение вычисляется по формуле:

$$DFFIT_{cr} = 2\sqrt{(p + 1)/n},$$

где p – количество используемых в модели факторов, n – размер выборки.

Сама величина вычисляется по формуле:

$$DFFIT_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\hat{\sigma}_{(i)}\sqrt{H_{ii}}}.$$

Здесь величина $\hat{Y}_{i(i)}$ – значение регрессии, которая была построена по выборке с исключенной (i)-ой компонентой, вычисленное по i -му наблюдению. H_{ii} – диагональный элемент «hat» матрицы или матрицы влияния [33]. Вычисленные значения представлены на графике ниже:

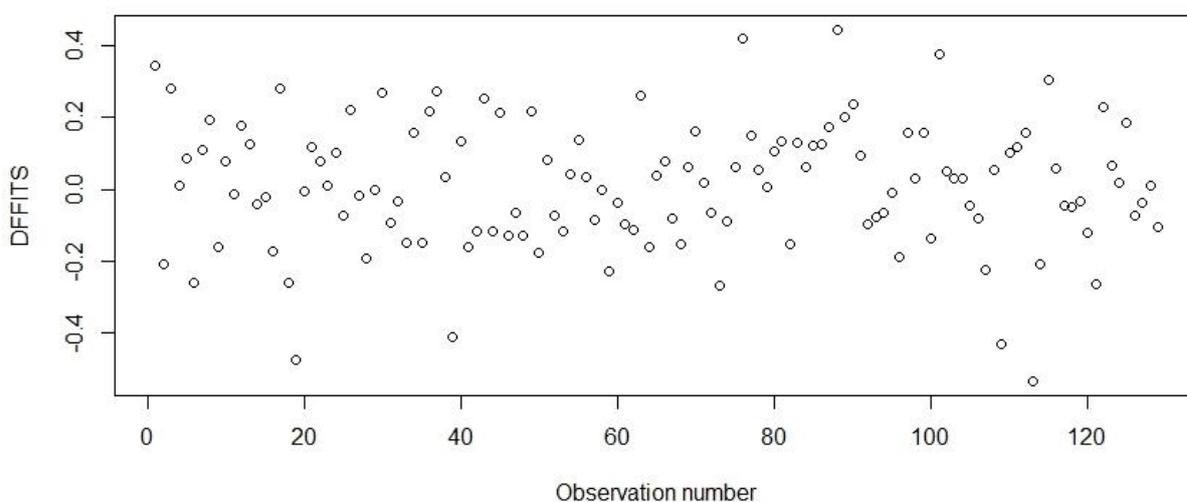


Рисунок 13. График величин DFFIT

На рисунке 13 отсутствуют явные выбросы. Пороговым значением в нашем случае является величина 0.352, поэтому к подозрительным или наиболее влиятельным можно отнести следующие наблюдения:

	QL	MYoS	DR	MF
Cambodia	3.907	4.37	7.68	70.50
Georgia	4.252	12.11	10.82	46.63
Mali	4.073	2.04	12.89	41.55
Nigeria	4.875	5.95	12.90	28.03
Saudi Arabia	6.379	8.70	3.33	7.75
Sri Lanka	4.415	10.80	6.11	21.90
Syrian Arab Republic	3.069	6.25	4.00	23.30

2. Расстояние Кука (Cook's distance)

Другой показательной величиной является расстояние Кука [33, 34, 35]. Оно отражает изменение регрессии в целом, при условии удаления i -го наблюдения. Формула представлена ниже:

$$D_i = \frac{\sum_{j=1}^n (\hat{Y}_j - \hat{Y}_{j(i)})^2}{(p + 1)\hat{\sigma}^2}.$$

Здесь \hat{Y}_j – модельное значение, соответствующее j -му наблюдению; $\hat{Y}_{j(i)}$ – значение регрессии, построенной по выборке с исключенной (i)-ой компонентой, вычисленное по j -му наблюдению; p – количество используемых в модели факторов; n – размер выборки; $\hat{\sigma}^2$ – дисперсия.

Видно (см. Рисунок 14, Рисунок 15), что все наблюдения лежат «кучно». Тем не менее, некоторые можно отнести к подозрительным, а именно: Syrian Arab Republic, Cambodia, Nigeria. Они же были отнесены к аналогичной категории по результатам анализа величины *DFFIT*.

График результатов:

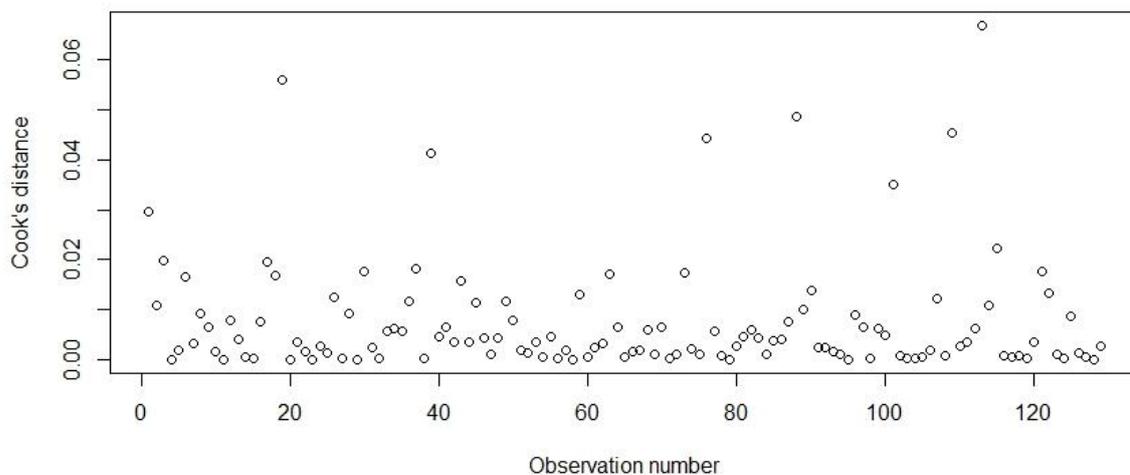


Рисунок 14. График расстояний Кука

График результатов в более наглядном представлении:

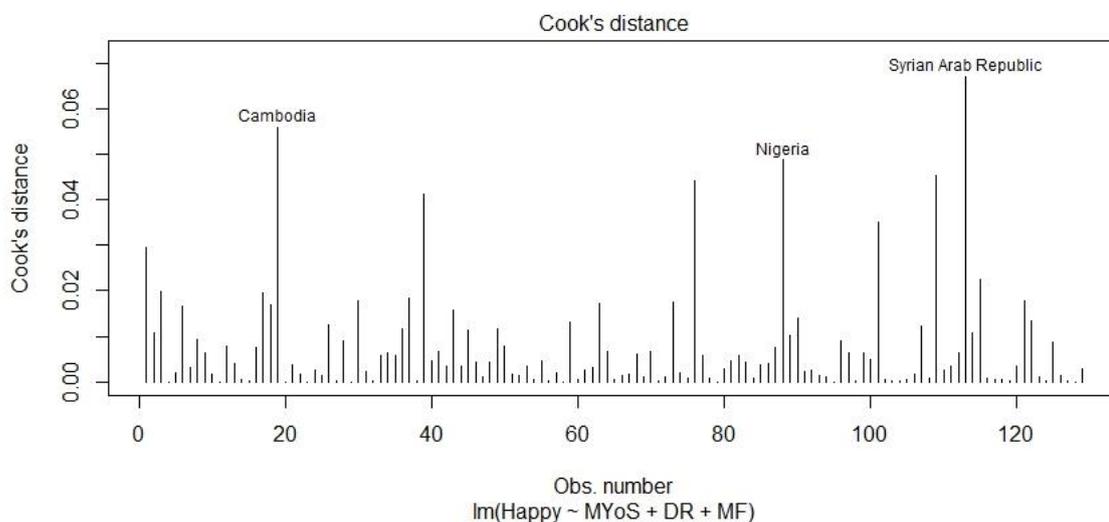


Рисунок 15. Гистограмма расстояний Кука

3. DFBETA

Величина $DFBETA$ отражает изменение коэффициентов регрессии при удалении i -го наблюдения [33]. Она вычисляется для каждого коэффициента по формуле, представленной ниже:

$$DFBETA_{j(i)} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{\hat{\sigma}_{(i)}^2 (X^T X)^{-1}_{jj}}}$$

Здесь $\hat{\beta}_j$ – j -ый коэффициент оцениваемой модели; $\hat{\beta}_{j(i)}$ – j -ый коэффициент регрессии, построенной по выборке с исключенным i -м наблюдением; X – используемая выборка.

Пороговым значением является единица для выборок малой размерности и $2/\sqrt{n}$ – большой размерности. В нашем случае пороговое значение равно 0.176. Результаты представлены ниже.

Для коэффициента, соответствующего фактору MYoS:

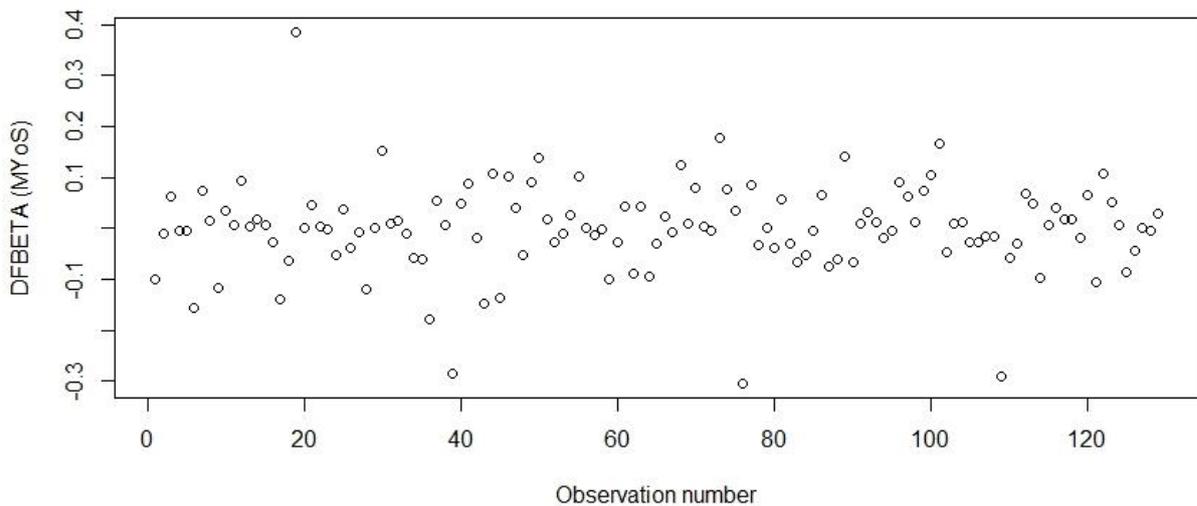


Рисунок 16. График величин DFBETA относительно фактора MYoS

Подозрительные наблюдения:

	QL	MYoS	DR	MF
Cambodia	3.907	4.37	7.68	70.50
Ethiopia	4.508	2.41	8.19	34.63
Georgia	4.252	12.11	10.82	46.63
Madagascar	3.695	6.05	6.81	46.70
Mali	4.073	2.04	12.89	41.55
Sri Lanka	4.415	10.80	6.11	21.90

Результат для коэффициента при DR:

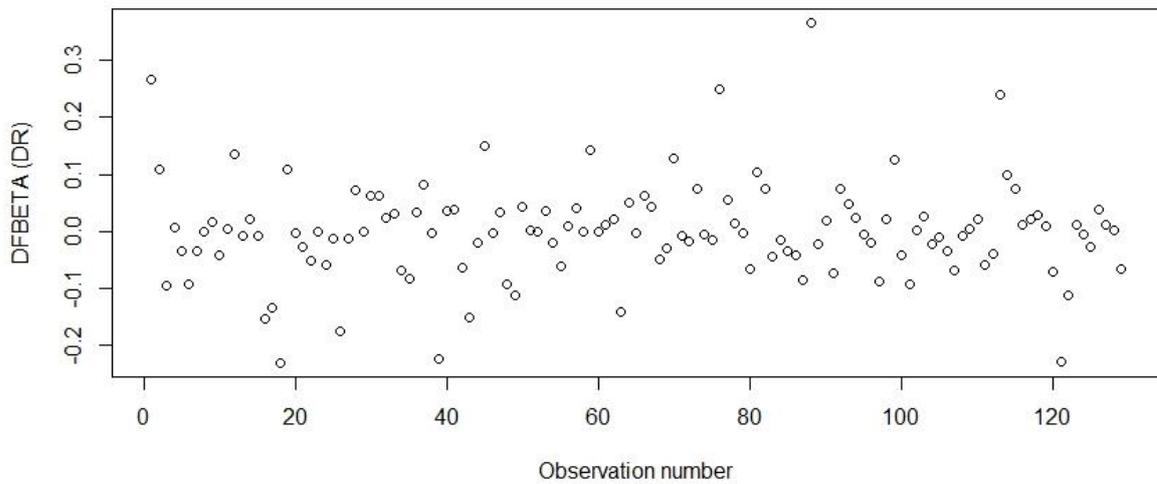


Рисунок 17. График величин DFBETA относительно фактора DR

Подозрительные наблюдения:

	QL	MYoS	DR	MF
Afghanistan	3.360	3.23	13.89	16.88
Bulgaria	4.217	10.57	14.44	53.33
Georgia	4.252	12.11	10.82	46.63
Mali	4.073	2.04	12.89	41.55
Nigeria	4.875	5.95	12.90	28.03
Syrian Arab Republic	3.069	6.25	4.00	23.30
Ukraine	4.324	11.34	14.46	47.58

Результат для коэффициента при MF:

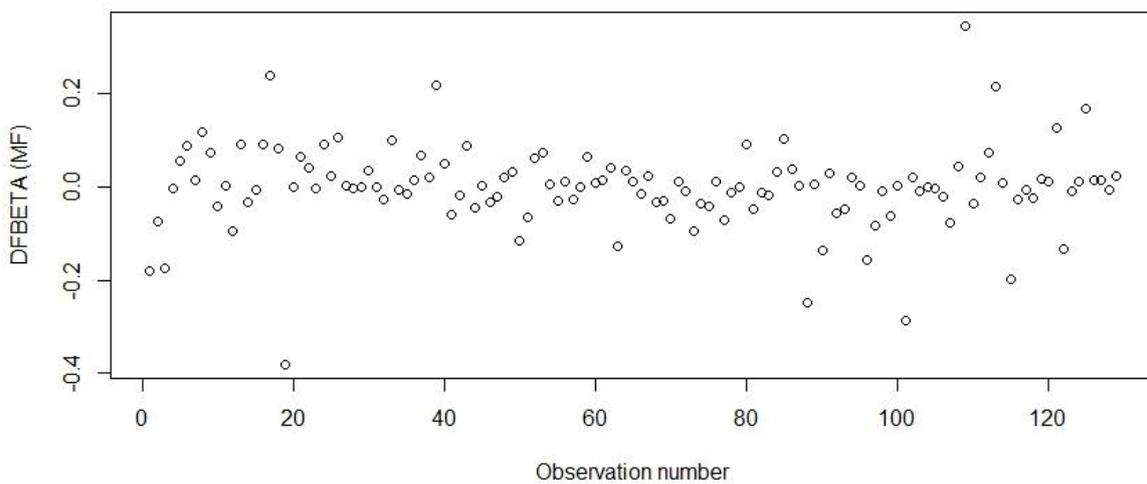


Рисунок 18. График величин DFBETA относительно фактора MF

Подозрительные наблюдения:

	QL	MYoS	DR	MF
Afghanistan	3.360	3.23	13.89	16.88
Brazil	6.952	7.66	6.58	69.30
Cambodia	3.907	4.37	7.68	70.50
Georgia	4.252	12.11	10.82	46.63
Nigeria	4.875	5.95	12.90	28.03
Saudi Arabia	6.379	8.70	3.33	7.75
Sri Lanka	4.415	10.80	6.11	21.90
Syrian Arab Republic	3.069	6.25	4.00	23.30
Thailand	6.474	7.32	7.80	30.75

По графикам легко заметить, что в каждом случае имеются наблюдения, оказывающее значительно большее влияние на фоне остальных.

Таким образом, выявлены группы наблюдений, которые сильнее остальных воздействуют на регрессионную модель. Теперь проведем анализ на наличие выбросов в используемых данных.

Выбросом будем считать наблюдение (Y, X_1, \dots, X_l) , которое сильно выделяется на фоне остальных (например, крайне удалено от основной группы). Различают выбросы в предикторах (X) и в зависимой переменной (Y).

Для определения выбросов среди X , помимо непосредственного отображения на графике, используют величину *leverage*, которая вычисляется по формуле [35]:

$$leverage_i = H_{ii} = (X(X^T X)^{-1} X^T)_{ii}$$

Видно (см. Рисунок 19), что наблюдение в левом верхнем углу сильнее других удалено от остальных. Оно соответствует стране Афганистан. Его можно смело относить к выбросам.

Результат отражен на графике ниже:

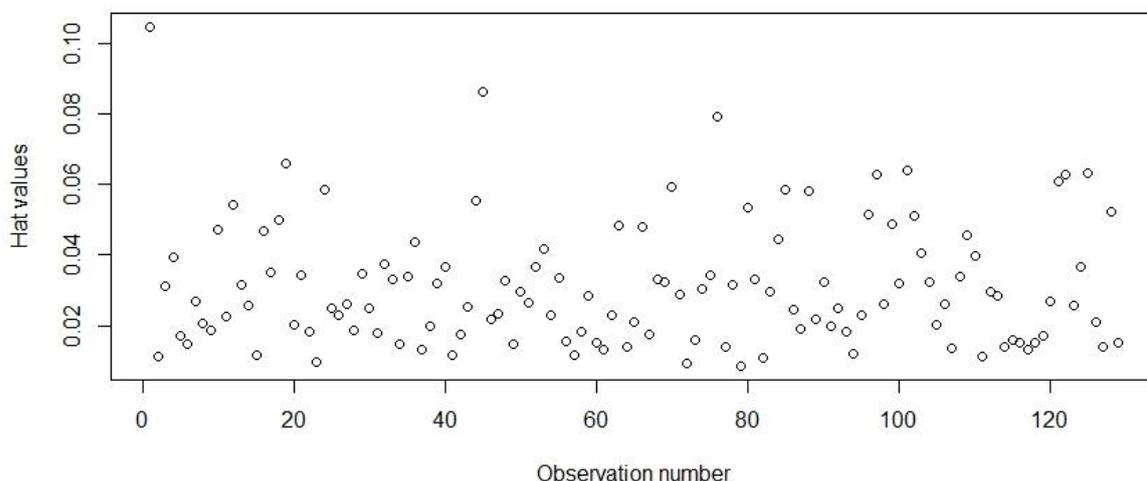


Рисунок 19. График величин leverage

Выявить выбросы в зависимой переменной можно с помощью студентизированных остатков, которые должны подчиняться распределению Стьюдента уровня $(1 - \alpha/2)$ с $(n - p - 2)$ степенями свободы. Однако, ввиду того, что данный тест придется проводить для каждого элемента выборки, то есть n раз, возможно, что при больших размерностях будут иметь место ложные срабатывания [35, 36]. Другими словами, наблюдение будет принято за выброс, даже если модель корректна. Эта проблема известна как «множественное сравнение». Для ее устранения вводится поправка Бонферрони, которая устанавливает величину $t_{кр}$, соответствующую значению квантиля распределения Стьюдента уровня $(1 - \frac{\alpha}{2n})$ с $(n - p - 2)$ степенями свободы [35, 36].

Итог применения теста:

No studentized residuals with bonferroni $p < 0.05$

Эта стандартная автоматически выдаваемая программой ремарка означает, что выбросы в исследованной зависимой переменной отсутствуют.

Помимо проведенных процедур, о влиянии наблюдений на регрессионную модель можно судить по графику, который ставит в соответствие величине *leverage* значения стандартизированных остатков. Результаты отражены на графике:

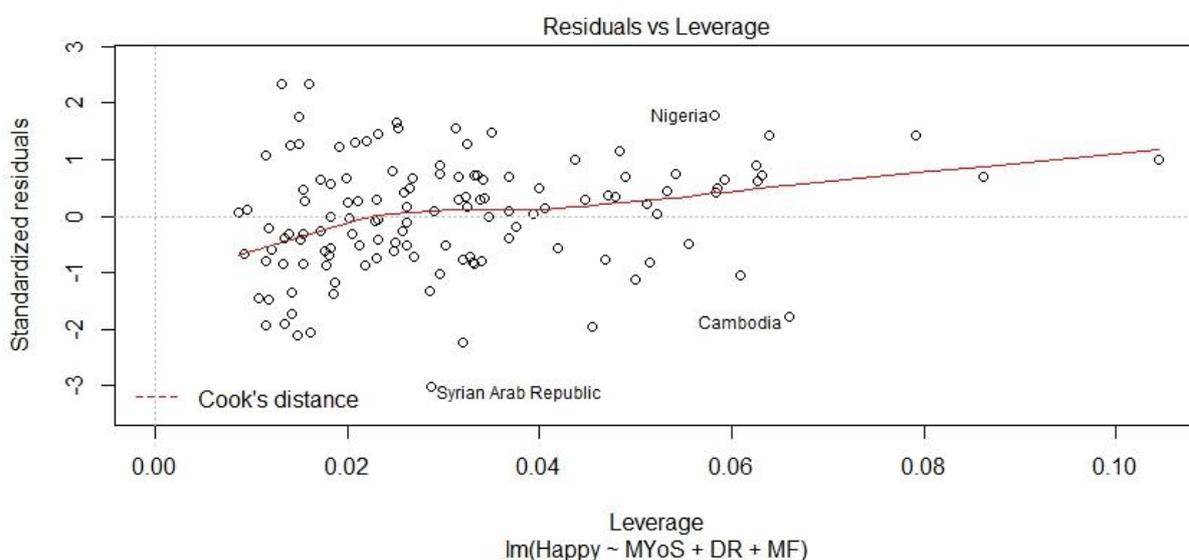


Рисунок 20. График стандартизированных остатков относительно leverage

По итогам проведенного анализа из исходной выборки исключены наиболее влиятельные наблюдения: Арабская Республика Сирия, Камбоджа, Нигерия. Также выявлен один выброс, соответствующий стране Афганистан. Обобщенная таблица результатов представлена в приложении.

2.3.4 Построение и верификация модели по результатам пункта 2.3.3

Напомним, что по итогам реализации второго подхода в пункте 2.3.2 пригодной для дальнейшего использования оказалась только модель 2.3, а именно:

$$QL = 3.913 + 0.217 \cdot MYoS - 0.163 \cdot DR + 2.096e^{-02} \cdot MF.$$

Однако, средняя ошибка аппроксимации составляла 10.6%, а скорректированный коэффициент детерминации - 0.62. Поэтому в пункте 2.3.3 был проведен дополнительный анализ на наличие выбросов в данных, по итогам которого было исключено четыре наблюдения. Построим модель методом последовательного включения факторов по вновь сформированной выборке размерностью 125x7.

Таблица. Порядок включения факторов в итоговую модель 2.3

Фактор	включен: MYoS	включен: DR	включен: MF	остановка
MYoS	0.4201	использован	использован	использован
DR	0.0372	0.5466	Использован	использован
MF	0.1870	0.4359	0.6405	использован
UR	0.0371	0.4716	0.5710	0.6606
IN	0.0001	0.4213	0.5454	0.6375
TTR	0.1646	0.4290	0.5862	0.6562
Inf	0.0828	0.4416	0.5587	0.6427

Итоговая модель 2.3:

$$QL = 4.026 + 0.211 \cdot MYoS - 0.185 \cdot DR + 2.360e^{-02} \cdot MF.$$

1) Доверительные интервалы для коэффициентов:

Доверительные интервалы для коэффициентов итоговой модели 2.3			
Коэффициент	2.5%	97.5%	
Своб. член	3.581	4.471	значим
MYoS	0.167	0.256	значим
DR	-0.229	-0.142	значим
MF	0.015	0.032	значим

Все коэффициенты статистически значимы.

2) *Нормальность распределения остатков:*

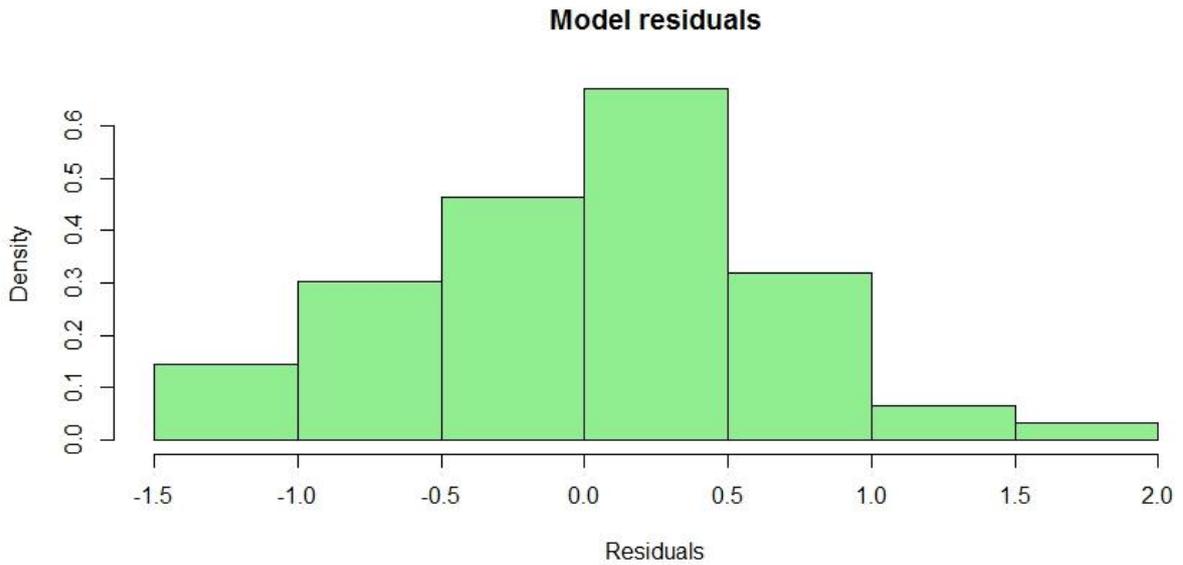


Рисунок 21. Гистограмма частот остатков итоговой модели 2.3

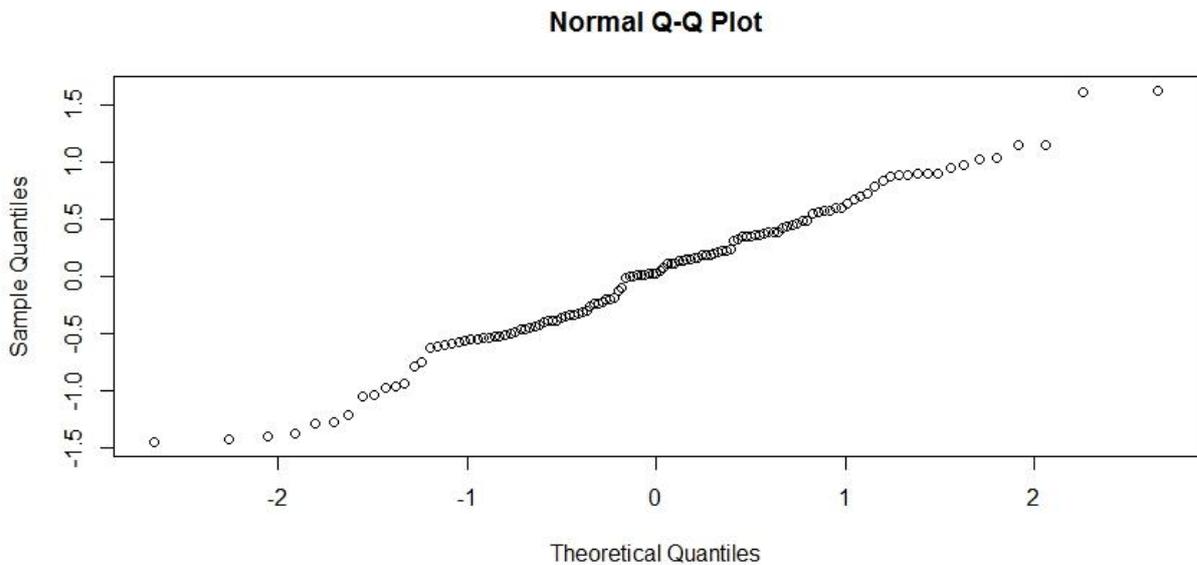


Рисунок 22. График квантилей итоговой модели 2.3

По гистограмме легко заметить асимметрию и смещение относительно нуля. Это отражается и на графике квантилей. В идеальном случае точки на нем должны располагаться строго на диагонали. Поэтому по представленным изображениям мы не можем смело говорить о нормальности распределения остатков. Как и ранее, проведем ряд тестов.

Тест Шапиро-Уилка: $W = 0.988$, $p\text{-value} = 0.362$.

Тест Андерсона-Дарлинга: $A = 0.393$, $p\text{-value} = 0.371$

Тест Колмогорова-Смирнова: $D = 0.057$, $p\text{-value} = 0.411$

Критерий хи-квадрат Пирсона: $P = 17.352$, $p\text{-value} = 0.098$

Критерий χ^2 Пирсона, в отличие от остальных, показал очень низкое, но положительное в смысле принятия нулевой гипотезы значение $p\text{-value}$. Остальные тесты, очевидно, пройдены. Отметим, что тест Шапиро-Уилка обладает наибольшей мощностью среди проведенных [37, 38]. Таким образом, мы принимаем нулевую гипотезу о нормальности распределения остатков на 5% уровне значимости.

3) Проверка гомоскедастичности:

Тест Гольдфельда-Квандта:

Зависимость от первого фактора (MYoS): $GQ = 1.141$, $p\text{-value} = 0.309$

Зависимость от второго фактора (DR): $GQ = 0.986$, $p\text{-value} = 0.522$

Зависимость от второго фактора (MF): $GQ = 0.907$, $p\text{-value} = 0.645$

Тест Бройша-Пагана: $BP = 2.052$, $p\text{-value} = 0.562$

Тест Уайта: $W = 10.997$, $p\text{-value} = 0.276$

По итогам всех тестов очевидно, что гетероскедастичность отсутствует.

4) Наличие систематических ошибок:

$E = -3.409e^{-17}$. Значение крайне близко к нулю, что говорит об отсутствии систематических ошибок в остатках.

5) Мультиколлинеарность:

$\det(X^T X) = 3.423e^{12}$. Значение отлично от нуля, следовательно, мультиколлинеарность отсутствует.

б) *Независимость остатков:*

Тест Дарбина-Уотсона: $DW = 2.195$, $dL = 1.676$, $dU = 1.740$, $4 - dU = 1.8053$

Тест Льюнга-Бокса: $X = 1.796$, $p\text{-value} = 0.180$

Тест Бройша-Годфри: $LM = 1.817$, $p\text{-value} = 0.178$

Тесты проверяли остатки на наличие автокорреляции первого порядка. В тесте Дарбина-Уотсона $dU < DW < 4 - dU$, следовательно, гипотеза о независимости случайных отклонений не отвергается на уровне значимости 0.05. Аналогичный результат дали остальные два теста. Получаем, что автокорреляция в остатках отсутствует.

Таким образом, построенная модель удовлетворяет постулатам регрессионного анализа. Средняя ошибка аппроксимации составляет 9.77%, также увеличилось значение $R_{adj} = 0.64$. Социально-экономическая интерпретация будет дана в пункте 2.4.

2.3.5 Еще один подход к построению модели

Ежегодно собирается огромное количество статистической информации. Однако, одни показатели вычисляются реже, другие вовсе перестают поддерживаться. В этом заключалась основная проблема при поиске подходящих данных. В идеальном случае, вся собранная информация должна датироваться одним годом. Однако, в реальности такой вариант не возможен, особенно на уровне проводимого в нынешней работе исследования. Ввиду этого, была собрана наиболее новая и, по возможности, полная информация. Это также значит, что часть пропусков была заполнена данными за ближайший доступный год [17], что, естественно, приводит к неточности исходной выборки. По этой причине проведем дополнительный анализ, в котором сузим допустимый временной интервал сбора показателей строго до периода с 2014 по 2016 год.

Данные по выбранной ранее в п. 2.1 зависимой переменной доступны, как за 2015, так и за 2016 год. Поэтому в текущем пункте будут рассмотрены оба набора отдельно. Итоговая выборка насчитывает 22 показателя, которые представлены в таблице ниже:

Таблица. Список показателей, подходящих под критерий отбора

Код индекса	Расшифровка	Дата
GE	Эффективность правительства	2014
CC	Контроль коррупции	2014
RQ	Качество управления	2014
FiW	Свобода в мире	2016
HDI	Индекс человеческого развития (ИРЧП)	2014
EYoS	Ожидаемый период обучения	2014
MYoS	Средняя длительность обучения	2014
TTRate	Итоговая налоговая ставка	2015
ITRate	Ставка налога на прибыль	2016
EF	Экономическая свобода	2016
WPF	Свобода прессы	2016
MF	Свобода морали	2016
GI	Индекс гендерного неравенства	2014
IHDI	ИРЧП с учетом гендерного неравенства	2014
GGGI	Глобальный индекс гендерного неравенства	2015
DR	Уровень смертности	2015
GDP	ВВП на душу населения	2015
Inf	Уровень инфляции	2015
LEB	Ожидаемая продолжительность жизни	2015
Pol	Уровень загрязнения	2015-2016
Hexp	Стоимость здравоохранения	2014
Happy	Уровень счастья	2015 или 2016

1. Построение модели на основе данных по показателю уровня счастья за 2015 год

Построение регрессионной модели проведем, как и ранее, методом последовательного включения факторов в модель. Критерием отбора будет служить скорректированный коэффициент детерминации. Пороговое значение, как и ранее, равно 0.04 [23]. Полученная модель:

$$QL = 1.394 + 2.289e^{-04} \cdot Hexp - 5.282e^{-02} \cdot LEB$$

Доверительные интервалы для коэффициентов:

Доверительные интервалы для коэффициентов			
Коэффициент	2.5%	97.5%	
Своб. член	-0.470	3.258	не значим
LEB	1.56E-04	3.02E-04	значим
Hexp	0.02695	0.07869	значим

Так как в результате анализа получено, что свободный член не значим, построим новую модель.

Модель 3.1:

$$QL = 1.990e^{-04} \cdot Hexp + 7.209e^{-02} \cdot LEB$$

Верификация модели пройдена, все коэффициенты регрессии значимы:

Доверительные интервалы для коэффициентов модели 3.1			
Коэффициент	2.5%	97.5%	
LEB	1.38E-04	2.60E-04	значим
Hexp	6.98E-02	7.44E-02	значим

Средняя ошибка аппроксимации составляет 9.66%, скорректированный коэффициент детерминации – 0.987. Результат верификации модели будет представлен далее в подпункте 3.

2. Построение модели на основе данных по показателю уровня счастья за 2016 год

Несмотря на то, что значение индекса по некоторым странам существенно отличается от показателя за 2015 год, при построении моделей наблюдается схожая картина. Формирование регрессии проведем аналогичным образом. Полученная модель:

$$QL = 3.724e^{-01} + 2.055e^{-04} \cdot Hexp + 6.689e^{-02} \cdot LEB$$

Видно, что оценка качества жизни зависит от тех же показателей: стоимость здравоохранения и ожидаемая продолжительность жизни. Более того, аналогично предыдущему подпункту, свободный член уравнения не значим. Поэтому итоговый вариант регрессионной модели имеет следующий вид, модель 3.2:

$$QL = 1.975e^{-04} \cdot Hexp + 7.204e^{-02} \cdot LEB$$

Все коэффициенты уравнения значимы.

Если сравнить полученную регрессию с моделью 3.1 прошлого подпункта:

$$QL = 1.990e^{-04} \cdot Hexp + 7.209e^{-02} \cdot LEB,$$

то видно, что в коэффициентах имеется незначительное отличие. Таким образом, можно сделать вывод о том, что использование данных за 2015 или 2016 год значительной роли не играют. Это является закономерным результатом, так как глобальные социально-экономические показатели не имеют свойства резкого изменения в течение года, за исключением кризисных ситуаций и возможных войн [12].

Средняя ошибка аппроксимации второй модели составила 9.03%, скорректированный коэффициент детерминации – 0.988.

3. Верификация моделей

Как и ранее, результаты проведенных тестов представлены в таблице ниже.

Таблица. Результаты анализа моделей

Название теста	Модель 3.1	Модель 3.2
<i>Значимость коэффициентов</i>	Все коэффициенты значимы	Все коэффициенты значимы
Нормальность распределения остатков		
<i>Тест Шапиро-Уилка</i>	W = 0.989, p-value = 0.610; пройден	W = 0.988, p-value = 0.526; пройден

<i>Тест Андерсона-Дарлинга</i>	A = 0.156, p-value = 0.953; пройден	A = 0.303, p-value = 0.567; пройден
<i>Тест Колмогорова-Смирнова</i>	D = 0.038, p-value = 0.987; пройден	D = 0.053, p-value = 0.753; пройден
<i>Критерий χ^2 Пирсона</i>	P = 2.753, p-value = 0.987; пройден	P = 12.258, p-value = 0.268; пройден
Гомоскедастичность		
<i>Тест Гольдфельда-Квандта</i>	Нехр: p-value = 0.132; пройден LEB: p-value = 0.435; пройден	Нехр: p-value = 0.183; пройден LEB: p-value = 0.370; пройден
<i>Тест Бройша-Пагана</i>	BP = 10.991, p-value = 0.052; пройден	BP = 2.101, p-value = 0.147; пройден
<i>Тест Уайта</i>	W = 14.447, p-value = 0.013; не пройден	W = 10.388, p-value = 0.065; пройден
Наличие систематических ошибок		
<i>Среднее остатков</i>	= 7.307e-3; пройден	= 1.952e-03; пройден
Мультиколлинеарность		
<i>det($X^T X$)</i>	= 1.166e+14; пройден	= 1.166e+14; пройден
Независимость остатков		
<i>Тест Льюнга-Бокса</i>	$\chi^2 = 0.015$, p-value = 0.903; пройден	$\chi^2 = 0.613$, p-value = 0.434; пройден
<i>Тест Бройша-Годфри</i>	LM = 0.014, p-value = 0.906; пройден	LM = 0.596, p-value = 0.440; пройден
Средняя ошибка аппроксимации		
A	9.66%	9.03%

Несмотря на всю внешнюю схожесть, при проверке первой модели тест Уайта отклонил нулевую гипотезу, а p-value теста Бройша-Пагана оказалось очень близко к граничному значению – 0.05. Поэтому, мы имеем право сделать предположение о наличии гетероскедастичности в остатках первой регрессии. Более того, средняя ошибка аппроксимации второй модели несколько ниже, что делает ее, в итоге, более предпочтительной для дальнейшего использования.

Итоговый вид модели 3.2:

$$QL = 1.975e^{-04} \cdot Нехр + 7.204e^{-02} \cdot LEB$$

2.4 Интерпретация моделей

По итогам проведенного в пункте 2.3 регрессионного анализа, получены три модели, которые наиболее предпочтительны для дальнейшего использования. Напомним их вид.

Модель 1.1, полученная в результате реализации первого подхода:

$$1. QL = 4.8 + 3.699e^{-05} \cdot GDP - 3.587e^{-02} \cdot UR + 2.156e^{-02} \cdot MF - 8.324e^{-02} \cdot DR.$$

Модель 2.3, полученная в результате реализации второго подхода:

$$2. QL = 4.026 + 0.211 \cdot MYoS - 0.185 \cdot DR + 2.360e^{-02} \cdot MF.$$

Модель 3.2, полученная в результате реализации подхода, описанного в п. 2.3.5:

$$3. QL = 1.975e^{-04} \cdot Hexp + 7.204e^{-02} \cdot LEB.$$

Некоторые сравнительные характеристики представлены в таблице ниже:

Таблица. Характеристики моделей

Характеристика	Модель 1	Модель 2	Модель 3
Скорректированный коэффициент детерминации	0.76	0.64	0.98
Средняя ошибка аппроксимации	6,69%	9.77%	9.03%

Из таблицы видно, что наименее точной является модель 2, более того, она описывает всего 64% вариации зависимого признака. Наилучшей, с точки зрения точности, является модель 1, но ее R_{adj}^2 ниже, чем у третьей модели. Таким образом, невозможно выделить явного «фаворита» из представленной группы уравнений.

Но если на время забыть про анализ с математической точки зрения, то какой реальный смысл несут полученные модели? Не трудно заметить, что оцениваемый параметр - «качество жизни», складывается из различного набора показателей в зависимости от уравнения.

Так, третья модель зависит от таких факторов, как стоимость здравоохранения и ожидаемая продолжительность жизни. Оба показателя относятся к характеристикам уровня здравоохранения в стране. Таким образом, может сложиться впечатление, что качество жизни населения зависит только от одной сферы жизнедеятельности, что, естественно, не так. Поэтому, несмотря на хорошую точность, практическое применение данной модели ставится под сомнение, ввиду своей содержательной стороны.

Оценка качества жизни населения, согласно модели 2.3, зависит уже от более разнообразного набора факторов, а именно: от средней длительности обучения, уровня смертности и уровня моральных свобод в обществе. Эти показатели представляют три сферы жизни: образование, здоровье и основные права. Такого рода зависимость, определенно, более адекватно отражает реальную картину мира. Особенно в естественном предположении, что каждый человек стремится прожить долгую, здоровую и полную жизнь. Так, средняя длительность обучения напрямую влияет на возможность реализации своего потенциала, тогда как уровень моральных свобод отражает удобство проживания на территории государства. Несмотря на свою информативную содержательную сторону, данная модель, напротив, обладает наихудшим качеством прогноза из представленных. Однако, она является пригодной для дальнейшего использования.

Баланс качества и содержательности наблюдается у первой модели. Главным отличием от второй является замена фактора «средней длительности обучения» на ВВП и уровень безработицы. Несмотря на то, что сам по себе показатель ВВП в качестве адекватной меры прогресса общества

все чаще подвергается сомнению, на него часто опираются при принятии решения на различных уровнях власти [6, 12, 15]. Более того, в совокупности со вторым фактором (UR), он также отражает возможность реализации своего потенциала человеком. Таким образом, оценка качества жизни населения складывается из «богатства» страны, уровня моральных свобод в обществе, уровня здравоохранения и показателя безработицы, что является закономерным результатом. Если также учесть хорошие показатели математической состоятельности этой модели, она становится наиболее предпочтительной среди всех представленных в данном исследовании.

2.5 Результаты моделирования

В нынешнем пункте будут представлены результаты моделирования с использованием моделей 1.1 и 2.3.

2.5.1 Модель 1.1

Вид уравнения:

$$QL = 4.8 + 3.699e^{-05} \cdot GDP - 3.587e^{-02} \cdot UR + 2.156e^{-02} \cdot MF - 8.324e^{-02} \cdot DR.$$

Напомним, что модель построена на данных содержащих 68 наблюдений. Для прогнозирования была сформирована выборка из 151 государства. Подробно результаты представлены в приложении, здесь же опишем часть из них.

Значение показателя для пятерки стран, получивших наивысшую оценку:

Таблица. Пятерка стран с наивысшей оценкой качества жизни по модели 1.1

Страна	LQ	GDP	UR	MF	DR
Liberia	38.14090924	900000	3.8	45.88	9.69
Central African Republic	26.39292558	600000	7.4	37.6	13.8
Qatar	10.36298327	145000	0.3	15.63	1.53
Luxembourg	9.350642996	102900	6.1	72.6	7.24
Singapore	8.570835344	85700	3	46.08	3.43

Значение показателя для пятерки стран, получивших наименьшую оценку:

Таблица. Пятерка стран с наименьшей оценкой качества жизни по модели 1.1

Страна	LQ	GDP	UR	MF	DR
Guinea-Bissau	4.399027	1500	6.9	45.63	14.33
Guinea	5.06082	1300	1.8	49.38	9.46
Mozambique	4.110002	1300	22.6	50.08	12.1
Malawi	4.928185	1200	7.5	48.83	8.41
Eritrea	4.537989	1200	7.2	26.8	7.52

Легко заметить, что в пятерке стран с наивысшей оценкой оказались небольшие государства с крайне высоким показателем ВВП на душу населения. Для сравнения: Объединенные Арабские Эмираты получили оценку 7.32 единицы с показателем GDP в 67000\$, США – 7.67 единицы при 56300\$, Великобритания – 6.57 при 41200\$ и Российская Федерация – 5.49 при 23700\$.

Ввиду описанной выше ситуации, модель 1.1 следует применять на выборке, состоящей преимущественно из крупных государств, или на данных отчищенных от наблюдений с аномально высоким показателем ВВП на душу населения, как, например, в случае с Либерией, Центральной Африканской Республикой или Катаром.

Таким образом, несмотря на высокую точность и качество, текущая модель чувствительна к составу используемых данных и требует проведения дополнительного анализа относительно показателя GDP, что ограничивает спектр ее применения.

Для сравнения проведем моделирование на отчищенной выборке, содержащей только 68 государств. Как и ранее, полные результаты размещены в приложении.

Значение показателя для пятерки стран, получивших наивысшую оценку:

1. Норвегия – 7.818
2. Австралия – 7.732
3. Швейцария – 7.716
4. США – 7.668
5. Нидерланды – 7.633

Значение показателя для пятерки стран, получивших наименьшую оценку:

1. Сербия – 4.515
2. Украина – 4.642
3. Египет – 4.719
4. Македония – 4.728
5. Эфиопия – 4.741

Россия получила оценку показателя в 5.494 единиц.

Наблюдаемые результаты лучше описывают реальную ситуацию (будет представлена в Главе 3), в сравнении с выборкой, состоящей из 150 государств.

2.5.2 Модель 2.3

Описанная выше особенность отсутствует у модели 2.3, так как показатель ВВП на душу населения заменен на среднюю длительность обучения, а остальным факторам соответствует больший весовой коэффициент. Вид модели:

$$QL = 4.026 + 0.211 \cdot MYoS - 0.185 \cdot DR + 2.360e^{-02} \cdot MF.$$

Используемая при моделировании выборка содержит информацию о 150 государствах. Полные результаты размещены в приложении, ниже представлена часть из них.

Значение показателя для пятерки стран, получивших наивысшую оценку:

Таблица. Пятерка стран с наивысшей оценкой качества жизни по модели 2.3

Страна	QL	MYoS	MF	DR
Netherlands	7.098841961	11.89	91.7	8.66
United States	7.096822897	12.94	78.2	8.15
Canada	7.021189787	13	76.58	8.42
Switzerland	6.937778788	12.82	72.38	8.13
Australia	6.89066157	12.96	61.35	7.14

Значение показателя для пятерки стран, получивших наименьшую оценку:

Таблица. Пятерка стран с наименьшей оценкой качества жизни по модели 2.3

Страна	QL	MYoS	MF	DR
Nigeria	3.553356	5.95	28.03	12.9
Central African Republic	3.246471	4.22	37.6	13.8
Mali	3.047503	2.04	41.55	12.89
Guinea-Bissau	3.043765	2.83	45.63	14.33
Afghanistan	2.53134	3.23	16.88	13.89

Согласно данной модели, Центральная Африканская Республика наоборот оказалась в пятерке с наименьшим значением показателя, что является более закономерным результатом. Российская Федерация получила оценку в 5.26 единицы и занимает 96 строчку в общем рейтинге.

Глава 3. Кластеризация

В предыдущей главе были получены модели, позволяющие оценить качество жизни населения на основе нескольких социально-экономических показателей. Но какое место занимает Россия, если рассматривать весь спектр представленных в исследовании факторов? По каким направлениям наша страна отстает от ведущей группы государств? Для ответа на эти вопросы в данном исследовании был привлечен инструментарий кластерного анализа.

Сегодня существует большое количество различных методов, решающих эту задачу. Одни лучше справляются с разбиением данных, соответствующих некоторому шаблону, другие формируют группы на основе наиболее информативных компонент [39, 40]. В текущем исследовании выбран метод, основанный на адаптивных картах Кохонена.

Самоорганизующаяся карта Кохонена – это нейронная сеть, не требующая наличия учителя, алгоритм обучения которой основан на конкуренции выходных узлов [41]. Она предназначена для решения широкого спектра задач, в том числе, кластеризации и визуализации. Такая нейронная сеть имеет ряд неоспоримых преимуществ:

- Возможность задания количества кластеров;
- Результаты кластеризации не зависят от выбора начальной точки;
- Проекция данных на пространство меньшей размерности, позволяющая лучше визуализировать результат путем формирования «карты»;
- Возможность отслеживания изменений весовых коэффициентов по ходу обработки данных;
- Не требуется формирование специальной обучающей выборки.

Кластеризацию будем проводить по исходному набору данных, дополненному полученной ранее оценкой качества жизни (по модели 1.1), так как представленные в этом случае социально-экономические показатели

очищены от схожих факторов и отражают все сферы жизни, согласно описанной в пункте 1.1 модели. Возьмем количество кластеров, равное трем, с целью оценки общей ситуации. Все вычисления были проведены в среде MATLAB [42, 43].

3.1 Результаты кластеризации

Состав полученных кластеров:

Группа 1	Группа 2	Группа 3
Australia	Argentina	Azerbaijan
Austria	Chile	Belarus
Belgium	Croatia	Bolivia
Canada	Czech Republic	Brazil
Denmark	Estonia	Bulgaria
Finland	Greece	China
France	Hungary	Colombia
Germany	Kazakhstan	Costa Rica
Iceland	Latvia	Dominican Republic
Ireland	Lithuania	Ecuador
Israel	Malaysia	Egypt. Arab Rep.
Italy	Poland	Ethiopia
Japan	Romania	Georgia
Netherlands	Russian Federation	Honduras
Norway	Slovak Republic	India
Spain	Slovenia	Indonesia
Switzerland	Turkey	Iran. Islamic Rep.
United Kingdom	Uruguay	Jordan
United States		Kenya
		Lao PDR
		Macedonia. FYR
		Mauritius
		Mexico
		Moldova
		Mongolia
		Morocco
		Peru
		Philippines
		Senegal
		Serbia
		South Africa
		Sri Lanka
		Thailand

Исследуя состав кластеров, можно предположить, что к первой группе были отнесены наиболее развитые с социально-экономической точки зрения государства, такие как США, Великобритания, Япония, Германия и другие. Напротив, наименее успешные страны оказались в третьей группе. Остальные были отнесены во вторую, среди них страны бывшего СССР, менее развитые европейские государства (в сравнении с первой группой), Россия и другие. Всего выборка насчитывала 72 наблюдения.

Рассмотрим следующий рисунок:

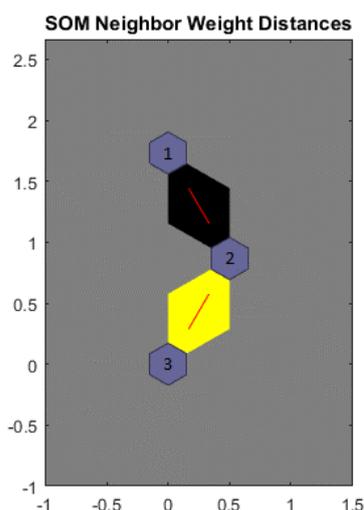


Рисунок 23. Иллюстрация расстояний между кластерами

На нем наглядно отражено расстояние между кластерами. Сами группы представлены фиолетовыми шестиугольниками. Нумерация ведется сверху вниз. Цвет между кластерами тем темнее, чем оно больше. Таким образом, видно, что расстояние между первой и второй группами крайне велико в сравнении с расстоянием между второй и третьей. Такая зависимость отражает разницу в социально-экономических показателях рассмотренных стран.

Ниже представлена таблица, содержащая средние значения факторов по каждой группе:

	QL	UR	IH	PSaAV	TTR	FiW	MYoS	TTRate	DCPS	EF	PD
Гр. 1	6.99	7.76	1.10	0.81	41.77	95.16	11.85	41.58	114.55	72.91	85.63
Гр. 2	5.83	9.28	3.53	0.40	30.11	77.11	10.96	45.89	59.38	65.25	50.73
Гр. 3	5.27	8.83	9.55	-0.38	25.07	56.17	8.14	42.76	55.89	59.35	49.43

	MF	GGGI	Palma	DR	GDP	Inf	LEB	Pol	CinAt	Hexp
Группа 1	67.03	0.77	5.75	8.65	47242.11	0.42	81.50	34.23	2.68	5415.24
Группа 2	59.41	0.70	6.56	10.34	25361.11	3.77	75.97	48.15	1.18	1064.64
Группа 3	45.98	0.69	9.14	7.58	11582.86	5.70	72.13	71.51	0.56	346.58

Если обратить особое внимание на первый показатель – QL (здесь это оценка качества жизни населения, полученная в Главе 2), то видно, что граждане государств первой группы живут заметно лучше. Эта закономерность прослеживается, если анализировать каждый индекс по отдельности. А каково место России в представленной картине мира? По каким показателям наше государство отстает от лидирующей группы?

3.2 Сравнение показателей Российской Федерации

Показатели Российской Федерации:

	QL	UR	IH	PSaAV	TTR	FiW	MYoS	TTRate	DCPS	EF	PD
Russia	5.49	5.10	9.50	-0.84	17.50	22.00	11.95	47.00	56.36	50.60	13.50

	MF	GGGI	Palma	DR	GDP	Inf	LEB	Pol	CinAt	Hexp
Russia	52.88	0.69	7.27	13.69	23700	15.40	70.47	64.49	1.36	892.852

Ниже представлена таблица, которая отражает, во сколько раз показатели России лучше или хуже соответствующих показателей лидирующей группы. В ней красным цветом отмечены индексы и соответствующие им значения, по которым наблюдается отставание. Зеленым цветом - аспекты, по которым, напротив, наблюдается некоторое опережение по сравнению с большинством других стран.

Сравнение											
Показатель	QL	UR	IH	PSaAV	TTR	FiW	MYoS	TTRate	DCPS	EF	PD
Группа 1	1.20	1.52	8.64	2.04	2.39	4.33	1.01	1.13	2.03	1.44	6.34

Сравнение										
Показатель	MF	GGGI	Palma	DR	GDP	Inf	LEB	Pol	CinAt	Hexp
Группа 1	1.27	1.10	1.26	1.58	1.99	36.58	1.16	1.88	1.97	6.07

Нетрудно заметить, что имеет место отставание практически по всем рассмотренным индексам. Лучшие характеристики наблюдаются по таким показателям, как уровень безработицы, суммарные налоговые поступления, внутренний кредит частному сектору, государственный долг, гендерное неравенство и стоимость здравоохранения. Применим полученные результаты к описанной в пункте 1.2 модели жизни:

Material living conditions:	TTR, TTRate, Palma
Productive or main activity:	UR
Health:	LEB, DR, Hexp
Education:	MYoS
Leisure and social interactions:	CinAt
Economic and physical safety:	IH, DCPS, PD, PSaAV
Governance and basic rights:	GGGI, EF, MF
Natural and living environment:	Pol
Overall experience of life:	QL
Global indexes:	GDP, FiW, Inf

Здесь красным цветом отмечены сферы жизни, у которых наблюдается отставание по всем описывающим их индексам, зеленым – напротив, наблюдается опережение, желтым – направления, о которых сложно судить однозначно.

Таким образом, немного лучше обстоят дела со сферой образования и вопросами, связанными с основной рабочей деятельностью. По остальным же

направлениям Россия скорее отстает, чем лидирует. Этот факт нашел отражение и в оценке качества жизни, полученном ранее в главе 2.

Выводы

Таким образом, по результатам проведенного исследования:

- Собраны максимально новые и полные данные по 35 социально-экономическим показателям, которые описывают основные сферы человеческой жизнедеятельности, согласно модели, представленной в пункте 1.1.
- Полученные сведения обработаны и сформированы 3 набора данных.
- Проведен регрессионный анализ, который включает:
 - Формирование различных подходов к построению моделей;
 - Построение линейных многофакторных регрессий по различным наборам данных;
 - Глубокий математический анализ каждого уравнения регрессии;
 - Социально-экономическую интерпретацию полученных результатов.
 - Предложена модель, описывающая качество жизни населения на основе находящихся в открытом доступе социально-экономических показателей.
 - Выявлены факторы, оказывающие наибольшее влияние на предложенный показатель.
- Проведена кластеризация данных по различным странам на основе нейронной сети.
- На основе построенных моделей и проведенного кластерного анализа выявлены направления, по которым Российская Федерация отстает от ведущей группы государств, и позиция, которую наша страна занимает в общемировом рейтинге.

Заключение

Главной целью нынешней работы является разработка подхода к оценке качества жизни населения на основе социально-экономических показателей, находящихся в открытом доступе. Такого рода задача продиктована стремлением получить агрегированный показатель прогресса общества без проведения дополнительных социологических опросов. Предложенный подход основан на построении многофакторной регрессионной модели методом последовательного включения факторов в условиях неполноты данных. Стоит отметить, что при формировании модели использованы как объективные (статистическая информация), так и субъективные (основанные на опросах) показатели.

Также в текущей работе предложен метод обработки социально-экономических данных, который заключается в анализе значимости показателей не только с математической, но и с содержательной точки зрения. По его итогам сформировано три набора данных, используемых в дальнейшем исследовании.

Результатом применения описанного в текущей работе подхода оценки качества жизни является формирование восьми регрессионных моделей. На их основе было выявлено, что качество жизни населения в большей степени зависит от таких показателей, как ВВП на душу населения, уровень безработицы, уровень моральных свобод в обществе, уровень смертности, средняя длительность обучения, стоимость здравоохранения и ожидаемая продолжительность жизни. Вариативность набора факторов в модели обуславливается важностью того, какой именно список стран был использован при ее формировании, ввиду сильного разброса в значениях показателей различных государств. Все модели подверглись глубокому статистическому анализу, по итогам которого пригодными для дальнейшего использования признаны две из них:

$$1. QL = 4.8 + 3.699e^{-05} \cdot GDP - 3.587e^{-02} \cdot UR + 2.156e^{-02} \cdot MF - 8.324e^{-02} \cdot DR.$$

$$2. QL = 4.026 + 0.211 \cdot MYoS - 0.185 \cdot DR + 2.360e^{-02} \cdot MF.$$

Практическое различие представленных уравнений выявлено на стадии моделирования. Ввиду того, что в первое из них непосредственно включается показатель ВВП на душу населения, наблюдается искажение модельных значений. Так, маленькие государства с небольшим населением, имеющие, возможно, множественные источники какого-либо ценного ресурса, характеризуются крайне высоким (аномальным относительно крупных государств с большой численностью населения) показателем ВВП (например, Либерия или Катар). Это, в свою очередь, значительно повышает оценку качества жизни, однако, полученный результат не отражает реального положения дел в таких государствах. Таким образом, первую модель следует применять на выборках, содержащих, преимущественно крупные государства.

Вторая модель исключает подобного рода зависимость и может быть использована на выборках больших размеров, а также по отношению к небольшим государствам.

Таким образом, сформированы две модели оценки качества жизни, каждая из которых имеет свою область применения.

Помимо этого, в работе проведен анализ, направленный на описание реальной социально-экономической картины мира, и выявление положения Российской Федерации в общемировом рейтинге. Данная часть работы основана на кластерном анализе с помощью самоорганизующейся карты Кохонена. По ее итогам сформировано три группы государств: наиболее развитые, среднего уровня развития и наименее развитые. Российская Федерация была отнесена ко второй из них. Сравнением показателей нашего государства относительно усредненных показателей ведущей группы, были

выявлены сферы жизнедеятельности, по которым наблюдается отставание: досуг и общение, окружающая среда и жилищные условия, качество жизни и общие макроэкономические показатели.

Таким образом, предложенная в данной работе методика позволяет оценить понятие качества жизни на основе общедоступных индексов и показателей. Все выявленные по результатам исследования зависимости и закономерности не противоречат общей логике происходящих в мире социально-экономических процессов.

Список литературы

1. A world countries list on Nations Online.
http://www.nationsonline.org/oneworld/countries_of_the_world.htm
2. Абель Э., Бернанке Б. Макроэкономика. Изд. 5-е. СПб.: Питер, 2008. 768 с.
3. Курс экономической теории: Общие основы экономической теории. Микроэкономика. Макроэкономика. Основы национальной экономики. / Под ред. Сидоровича А. В. Изд. 2-е. М.: Изд. «Дело и Сервис», 2001. 832 с.
4. World Bank Group website. <http://www.worldbank.org/>
5. OECD website. <http://www.oecd.org/>
6. Quality of life indicators on Eurostat.
http://ec.europa.eu/eurostat/statistics-explained/index.php/Quality_of_life_indicators_-_measuring_quality_of_life
7. Eurostat website. <http://ec.europa.eu/eurostat>
8. World happiness report main web page. <http://worldhappiness.report/>
9. Gallup. Inc. website. <http://www.gallup.com/home.aspx>
10. Report by the commission on the measurement of economic performance and social progress.
<http://ec.europa.eu/eurostat/documents/118025/118123/Fitoussi+Commission+report/7bac2480-4658-439f-b022-e6542ebf714e>
11. “Measuring progress, well-being and sustainable development” leaflet.
<http://ec.europa.eu/eurostat/documents/7330775/7339383/Measuring-Progress-Well-being-sustainable-development/ef61b63c-02f0-4aa6-8155-42fdc0833018>

12. Happiness report 2016. <http://worldhappiness.report/download/>
13. Ura K., Alkire S., Zangmo T. An extensive analysis of GNH index. Thimphu: The Centre for Bhutan Studies, 2012. 217 p.
14. Смолева Е. О., Морев М. В. Удовлетворенность жизнью и уровень счастья: взгляд социолога. Вологда: ИСЭРТ РАН, 2016. 164 с.
15. OECD Guidelines on measuring subjective well-being. OECD Publishing. http://www.oecd-ilibrary.org/economics/oecd-guidelines-on-measuring-subjective-well-being_9789264191655-en
16. 5th OECD World forum on statistics, knowledge and policy. Summary report. <http://www.oecd.org/statistics/5WF%20Summary%20Report.pdf>
17. Happy planet index 2016. Methods paper. https://static1.squarespace.com/static/5735c421e321402778ee0ce9/t/578cc52b2994ca114a67d81c/1468843308642/Methods+paper_2016.pdf
18. Chaddock R. E. Principles and methods of statistics. Ed. 1. Boston: Houghton Mifflin Co., 1925. 471 p.
19. Quality of life index on The economist intelligence unit's. https://www.economist.com/media/pdf/QUALITY_OF_LIFE.pdf
20. White A. A global projection of subjective well-being: A challenge to positive psychology // Psychtalk. 2007. P. 17-20.
21. Working guidebook to the National Footprint accounts. http://www.footprintnetwork.org/content/documents/National_Footprint_Accounts_2016_Guidebook.pdf
22. Tracking global happiness report 2013. https://www.ipsos.com/sites/default/files/Tracking_Global_Happiness-2013.pdf
23. Эконометрика / Елисеева И. И., Курышева С. В., Костеева Т. В., Михайлов Б. А. и др.; под ред. Елисеева И. И. М.: Финансы и статистика, 2003. 344 с.

24. Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика. Начальный курс. Изд. 6-е. М.: Дело, 2004. 576 с.
25. R manual. <https://cran.r-project.org/manuals.html>
26. Буре В. М., Евсеев Е. А. Основы эконометрики. СПб.: СПбГУ, 2004.
27. Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2006. 816 с.
28. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Основы моделирования и первичная обработка данных. М.: Финансы и статистика, 1983. 471 с.
29. Буре В.М., Парилина Е.М., Седаков А.А., Шевкопляс Е.В. Прикладная статистика в R, STATISTICA и Excel. Описательная статистика. Оценивание параметров. Статистические критерии. СПб.: СПбГУ, 2011. 104 с.
30. Lilliefors H. On the Kolmogorov-Smirnov test for normality with mean and variance unknown // Journal of the American Statistical Association. 1967. Vol. 62, No. 318. P. 309-402.
31. Breusch T. S., Pagan A. R. A simple test for heteroscedasticity and random coefficient variation // Econometrica. 1979. Vol. 47, No. 5. P. 1287-1294
32. Куфель Т. Эконометрика. Решение задач с применением пакета программ GRETL. М.: Горячая линия – телеком, 2007. 200 с.
33. Eyduran E., Ozdemir T., Alarslan E. Importance of diagnostics in multiple regression analysis // Journal of Applied Sciences. 2005. No. 5, P. 1792-1796.
34. Practical regression and ANOVA using R by Faraway J. <https://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>
35. Regression diagnostics on Tropej. http://www.oxfordjournals.org/our_journals/tropej/online/ma_chap1.pdf

36. Armstrong R. When to use the Bonferroni correction // *Ophthalmic & Physiological Optics*. 2014. Vol. 34. C. 502-508.
37. Razali N., Yap. B. W. Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests // *Journal of Statistical Modeling and Analysis*. 2011. Vol. 2, No. 1. P. 21-33.
38. Yap B. W., Sim C. H. Comparisons of various types of normality tests // *Journal of Statistical Computation and Simulation*. 2011. Vol. 81, No. 12. P. 2141-2155.
39. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: классификация и снижение размерности. М.: Финансы и статистика, 1989. 607 с.
40. Мандель И. Д. Кластерный анализ. М.: Финансы и статистика, 1988. 176 с.
41. Kohonen T. The self-organizing map // *Proceedings of the IEEE*. 1990. Vol. 78, No. 9. P. 1464-1480.
42. Ballabio D., Vasighi M. A MATLAB toolbox for self-organizing maps and supervised neural network learning strategies // *Chemometrics and Intelligent Laboratory Systems*. 2012. No. 118. C. 24-32.
43. MATLAB documentation. <https://www.mathworks.com/help/>

Приложение

Список используемых статистических баз данных.

<http://info.worldbank.org>
<https://freedomhouse.org>
<http://databank.worldbank.org>
<http://www.heritage.org>
<http://hdr.undp.org>
<http://reports.weforum.org>
<http://www.tradingeconomics.com>
<http://worldhappiness.report>
<http://apps.who.int/nha/database>
<http://www.fundalib.org>
<https://rsf.org>

Описание показателей.

UR - *Unemployment rate*. Уровень безработицы.

IH - *Intentional homicides per 100000 people*. Количество убийств на 100000 населения. Преступления, которые попадают под определение «убийство», могут немного варьироваться в зависимости от страны.

PSaAV - *Political Stability and Absence of Violence/Terrorism*. Политическая стабильность и отсутствие насилия/терроризма. Отражает представления о вероятности политической нестабильности и/или политически мотивированного насилия, включая терроризм.

TTR – *Total Tax Revenue*. Суммарные (итоговые) налоговые поступления. Налоги и другие доходы, получаемых правительством страны в течение указанного периода, выраженные в процентах от ВВП. Налоги включают персональные и корпоративные подоходные налоги, налоги на добавочную стоимость, налоги на акцизы и тарифы. Другие доходы включают в себя социальные взносы такие, как плата за соц. обеспечение и страхование, и чистые доходы от гос. предприятий.

VA – *Voice and Accountability*. Право голоса и прозрачность. Индекс отражает в какой степени граждане страны имеют возможность участвовать в выборе своего правительства, а также свободу выражения мнений, свободу ассоциаций и свободу СМИ.

GE – *Government Effectiveness*. Эффективность правительства. Отражает восприятие качества гос. услуг, качество муниципальных служб и степень их независимости от политического давления, качество выработки и реализации политики, а также доверие к намеченному курсу.

CC – Control of Corruption. Контроль коррупции. Отражает восприятие степени, в которой гос. власть осуществляется в целях личной выгоды, включая мелкие и крупные формы коррупции, а также “захват” государства со стороны элит и частных интересов.

RQ – Regulatory Quality. Качество управления. Отражает восприятие способности правительства разрабатывать и осуществлять обоснованную политику и нормативные акты, которые позволяют стимулировать развитие частного сектора.

FiW – Freedom in the World. Свобода в мире. Многофакторный индекс, который включает в себя анализ избирательного процесса, политического плюрализма и участия, функционирование правительства, свободы выражения, ассоциаций и организаций, верховенство закона и права личности.

HDI – Human Development Index. Индекс человеческого развития (индекс развития человеческого потенциала). Многофакторный индекс, который включает в себя такие аспекты жизни человека как: здоровье (а именно, индекс LEB), образование (индексы MYoS, EYoS), доходы (ВНД на душу населения). Имеется пример расчёта данного индекса.

EYoS – Expected Years of Schooling. Ожидаемый период обучения (лет).

MYoS – Mean Years of Schooling. Средняя длительность обучения (лет).

TTRate – Total Tax Rate. Итоговая (общая) налоговая ставка. Сумма налогов и обязательных отчислений, уплачиваемых предприятиями после учета допустимых вычетов и льгот (доля от коммерческой прибыли). Налоги удержанные (например, налог на доходы физ. Лица) или собранные, которые перечисляются в налоговые органы (например, налог на добавочную стоимость, налог с продаж товаров и услуг) исключаются.

ITRate – Income Tax Rate. Ставка налога на прибыль.

DCPS – Domestic Credit to Private Sector. Внутренний кредит частному сектору (проценты от ВВП). Он относится к финансовым ресурсам, предоставляемым частному сектору финансовыми корпорациями, которые устанавливают требования о погашении. Для некоторых стран учитываются кредиты государственными предприятиям.

EF – Economic Freedom. Экономическая свобода. Индекс включает в себя 10 аспектов экономической свободы: права собственности, свобода от коррупции, бюджетная свобода, государственные расходы, свобода бизнеса, свобода труда, денежные свободы, свобода торговли, свобода инвестиций, финансовая свобода.

WPF – World Press Freedom. Свобода прессы.

PD – Public Debt to GDP. Государственный долг. Кумулятивная величина, отражающая все государственные заимствования за вычетом бюджетных излишков.

GD – Government Debt to GDP. Аналогичен PD (см. выше).

MF – index of Moral Freedom. Свобода морали. Индекс представляет собой среднеарифметическое показателей, которые можно разделить на 5 групп: религиозные свободы, биоэтические показатели, показатели, отражающие доступность наркотиков,

индикаторы, связанные с полом и семьей, а также индикаторы, связанные с сексуальной сферой жизни.

Gini — *Gini index*. Индекс Джини. Индекс отражает неравномерность распределения доходов среди населения согласно кривой Лоренца. Значения от 1 до 0 — идеальное (равномерное) распределение. 1 — полное (абсолютное) неравенство.

GI — *Gender Inequality index*. Индекс гендерного неравенства. Показатель отражает неравенство в достижениях между мужчинами и женщинами в трех направлениях: репродуктивное здоровье, расширенные возможности и рынок труда.

IHDI — *Inequality-adjusted HDI*. ИРЧП с учетом полового неравенства. Индекс представляет собой HDI дополненный показателем GI.

TR — *Tax Revenue*. Налоговые поступления. Отличие данного индекса от TTR в том, что здесь не учитываются штрафы, большинство взносов соц. страхования. Возвраты и корректировки ошибочно взысканных налогов учитываются как отрицательный доход.

GGGI — *Global Gender Gap Index*. Глобальный индекс гендерного неравенства. Индекс отражает относительную разницу между мужчинами и женщинами в 4 ключевых аспектах: здоровье, образование, экономика и политика.

Palma — *Palma ratio*. Соотношение Пальма. Индекс определяется как отношение доходов 10% самых богатых к 40% самых бедных. Данное представление основывается на работе чилийского экономиста Габриэля Пальма, который показал, что 50% ВВП распределяется между средних слоев, а остальные 50% между 10% самых богатых и 40% самых бедных.

QR — *Quintile Ratio*. Квантильное соотношение. Индекс определяется как отношение распределения доходов среди 10% самых богатых к 10% самых бедных.

DR — *Death Rate*. Уровень смертности. Отражает среднее ежегодное количество смертей на 1000 человек (на момент середины года).

GDP — *Gross Domestic Product per capita*. ВВП на душу населения.

Inf — *Inflation rate*. Уровень инфляции.

LEB — *Life Expectancy at Birth*. Ожидаемая продолжительность жизни.

Pol (*Pollution rate*) – уровень загрязнения. Данный индекс оценивает суммарное загрязнение региона. Различные типы загрязнения получили различные веса. Самым большим весом обладает загрязнение воздуха, далее загрязнение воды (ее доступность). Это два наиболее значимых фактора загрязнения. Также индекс учитывает загрязнение мусором, уровень шума, чистоту и качество парков и др. аспекты загрязнения.

CinAt (*Cinema Attendance*) – посещаемость кинотеатров. Индекс отражает количество посещения на душу населения.

Hexp (*Health Expenditure*) – стоимость здравоохранения, выраженная в долларах США. Индекс рассчитывается как суммарные расходы государственного и частного секторов по отношению к общей численности населения. Он охватывает предоставление

медицинских услуг (профилактических и лечебных мероприятий), услуг по планированию семьи, деятельность в области неотложной помощи, но не включает в себя обеспечение водоснабжения и санитарию.

Happy (*Happiness*) – общий уровень счастья. Этот индекс отражает общий уровень счастья населения и включает в себя следующие аспекты жизни: ВВП на душу населения, социальная поддержка, ожидаемая продолжительность здоровой жизни, свобода выбора, щедрость, восприятие коррупции.

Корреляционная матрица по исходным данным.

	UR	IN	PSaAV	TTR	FIW	MVos	TTRate	DCPS	EF	PD	MF	GGGI	Palma	DR	GDP	Inf	LEB	Poi	Cinat	Некр	Happy
UR																					
IN	-0.03																				
PSaAV	-0.02	-0.25																			
TTR	0.05	-0.25	0.56																		
FIW	0.14	-0.18	0.68	0.54																	
MVos	0.01	-0.35	0.55	0.48	0.48																
TTRate	-0.10	0.12	-0.15	0.04	0.02	-0.12															
DCPS	0.00	-0.13	0.34	0.19	0.33	0.34	-0.14														
EF	-0.04	-0.20	0.61	0.30	0.63	0.53	-0.39	0.48													
PD	0.23	-0.15	0.22	0.28	0.36	0.13	0.09	0.34	0.12												
MF	0.05	-0.03	0.54	0.59	0.66	0.43	0.17	0.23	0.40	0.16											
GGGI	-0.09	-0.06	0.54	0.55	0.53	0.47	-0.05	0.21	0.36	0.04	0.51										
Palma	0.13	0.76	-0.27	-0.28	-0.10	-0.33	0.16	0.07	-0.15	-0.18	0.03	-0.06									
DR	0.21	-0.25	0.26	0.43	0.22	0.49	-0.08	-0.01	0.05	0.08	0.31	0.27	-0.36								
GDP	-0.09	-0.33	0.70	0.56	0.60	0.70	-0.06	0.57	0.66	0.27	0.53	0.57	-0.33	0.15							
Inf	-0.11	0.11	-0.48	-0.21	-0.31	-0.11	0.34	-0.27	-0.56	-0.10	-0.17	-0.19	0.04	0.11	-0.31						
LEB	0.05	-0.33	0.57	0.48	0.60	0.62	0.06	0.46	0.59	0.43	0.43	0.33	-0.30	0.03	0.77	-0.31					
Poi	0.05	0.26	-0.71	-0.57	-0.63	-0.65	0.09	-0.38	-0.57	-0.25	-0.55	-0.63	0.33	-0.42	-0.76	-0.31	-0.63				
Cinat	-0.18	-0.13	0.54	0.36	0.55	0.53	-0.02	0.51	0.58	0.24	0.50	0.57	-0.11	0.00	0.81	-0.23	0.62	-0.67			
Некр	-0.14	-0.24	0.60	0.54	0.56	0.55	-0.08	0.62	0.58	0.27	0.52	0.59	-0.25	0.08	0.92	-0.28	0.66	-0.69	0.78		
Happy	-0.29	-0.17	0.59	0.47	0.53	0.51	0.09	0.44	0.57	0.11	0.60	0.57	-0.10	-0.09	0.78	-0.24	0.70	-0.64	0.72	0.73	

Корреляционная матрица по отсеянным данным.

	Happy	UR	IN	TR	Myos	TRate	DCPS	EF	PD	MF	GGGI	DR	InfI
Happy													
UR	-0.16												
IN	-0.09	-0.01											
TR	0.43	0.16	-0.24										
Myos	0.60	0.18	-0.27	0.50									
TRate	0.06	0.03	0.09	0.04	-0.05								
DCPS	0.47	0.06	-0.20	0.29	0.45	-0.17							
EF	0.58	0.02	-0.14	0.30	0.58	-0.38	0.51						
PD	0.06	0.23	-0.05	0.21	0.16	0.03	0.31	0.07					
MF	0.39	0.18	-0.04	0.49	0.43	0.23	0.26	0.33	0.20				
GGGI	0.46	0.01	-0.01	0.45	0.46	0.06	0.19	0.35	0.08	0.55			
DR	-0.23	0.30	-0.14	0.30	0.26	0.09	-0.10	-0.05	0.09	0.40	0.26		
InfI	-0.27	-0.08	0.05	-0.19	-0.16	0.30	-0.33	-0.49	-0.13	-0.13	-0.15	0.13	

Исходные данные

Country	Happy	UR	IH	PSaAV	TTR	FiW	MYos	TTRate	DCPS	EF	PD	MF	GGGI	Palma	DR	GDP	Inf	LEB	Pol	CinAt	Hexp
Argentina	6.65	8.20	7.60	0.08	24.80	79.00	9.83	137.40	14.30	43.80	45.80	64.45	0.73	10.62	7.33	22400	27.60	77.69	58.46	1.31	605.1878
Australia	7.313	6.00	1.00	1.08	34.30	98.00	12.96	47.60	137.67	80.30	44.30	61.35	0.73	5.85	7.14	65400	1.90	82.15	22.88	3.93	6031.107
Austria	7.119	5.00	0.50	1.29	50.60	95.00	10.83	51.70	87.67	71.70	83.40	71.13	0.73	4.57	9.42	47500	0.90	81.39	30.77	1.98	5580.494
Azerbaijan	5.291	5.20	2.50	-0.50	25.20	16.00	11.16	39.80	38.45	60.20	11.80	42.68	0.68	5.10	7.07	18700	3.90	72.20	87.04	0.01	471.4144
Belarus	5.802	5.90	3.60	0.12	35.20	17.00	11.98	51.80	28.83	48.80	36.40	46.08	0.73	3.83	13.36	17800	15.00	72.48	33.7	1.38	450.214
Belgium	6.929	8.50	1.80	0.71	49.50	96.00	11.27	58.40	61.69	68.40	107.00	79.35	0.75	4.99	9.63	44100	0.50	80.88	56.61	2.14	4884.066
Bolivia	5.822	2.70	12.10	-0.36	48.60	68.00	8.15	83.70	57.71	47.40	37.50	65.30	0.75	15.20	6.52	6500	4.20	68.86	83.14	0.28	208.7842
Brazil	6.952	6.80	24.60	-0.01	35.10	81.00	7.66	69.20	67.86	56.50	67.30	69.30	0.69	16.87	6.58	15800	10.60	73.53	61.18	0.82	947.4277
Bulgaria	4.217	11.60	1.60	0.08	37.80	80.00	10.57	27.00	56.82	65.90	26.40	53.33	0.72	6.41	14.44	18400	0.10	74.39	60.35	0.73	661.8465
Canada	7.404	6.90	1.40	1.18	37.20	99.00	13.00	21.10	124.41	78.00	95.40	76.58	0.74	5.77	8.42	45900	1.20	81.76	29.48	3.94	5291.746
Chile	6.705	6.40	3.60	0.49	20.20	95.00	9.79	28.90	111.22	77.70	17.40	59.40	0.70	12.61	6.00	23800	4.40	78.61	67.21	1.31	1137.356
China	5.245	4.70	0.80	-0.46	21.30	16.00	7.54	67.80	155.33	52.00	16.70	40.00	0.68	10.08	7.53	14300	1.50	75.41	85.54	0.48	419.7342
Colombia	6.481	10.10	27.90	-1.12	29.30	63.00	7.35	69.70	47.13	70.80	46.90	74.98	0.73	17.54	5.40	14000	4.40	75.48	64.44	1.01	569.1853
Costa Rica	7.087	8.30	10.00	0.63	14.50	90.00	8.37	58.00	60.16	67.40	59.70	53.08	0.73	12.81	4.55	15500	1.10	78.40	53.09	1.41	970.002
Croatia	5.488	16.70	0.80	0.60	41.90	87.00	11.03	20.00	65.47	59.10	89.50	59.13	0.71	5.27	12.18	21300	-0.30	76.61	30.42	1.06	1050.334
Czech Republic	6.596	6.20	0.70	0.97	39.60	95.00	12.32	50.40	51.24	73.20	41.40	80.50	0.69	3.88	10.34	31500	0.40	78.48	42.49	1.14	1378.521
Denmark	7.526	6.60	1.00	0.94	55.60	98.00	12.73	24.50	174.04	75.30	47.20	66.33	0.77	3.96	10.25	45800	0.60	79.25	29.93	2.53	6463.243
Dominican Republic	5.155	15.00	17.40	0.19	16.00	70.00	7.56	42.40	27.54	61.00	44.70	38.90	0.69	10.29	4.55	14900	0.70	77.97	75.33	0.49	268.9943
Ecuador	5.976	4.60	8.20	-0.01	35.50	59.00	7.59	33.00	27.38	48.60	32.40	64.75	0.74	12.00	5.06	11300	4.00	76.56	59.32	0.97	579.1924
Egypt, Arab Rep.	4.362	13.20	3.40	-1.58	26.70	27.00	6.55	45.00	26.47	56.00	91.70	16.88	0.60	4.36	4.77	11500	10.40	73.70	96.05	0.17	177.7684
Estonia	5.517	7.70	3.10	0.76	38.20	94.00	12.48	49.40	69.54	77.20	10.20	69.40	0.75	5.63	12.40	28700	0.50	76.47	16.45	2.17	1248.28
Ethiopia	4.508	5.20	8.10	-1.24	15.80	15.00	2.41	32.10	17.71	51.50	45.80	34.63	0.64	5.26	8.19	1700	10.30	61.48	66.74	0.02	26.64778
Finland	7.413	8.60	1.60	1.28	57.00	100.00	10.29	37.90	96.34	72.60	61.20	60.68	0.85	4.04	9.83	41200	-0.20	80.77	17.84	1.59	4612.29
France	6.478	9.90	1.20	0.36	51.70	91.00	11.13	62.70	95.73	62.30	98.20	69.93	0.76	5.09	9.16	41400	0.10	81.75	46.31	3.4	4958.989
Georgia	4.252	13.40	2.70	-0.23	28.20	64.00	12.11	16.40	49.76	72.60	38.80	46.63	0.69	8.92	10.82	9500	4.10	75.95	69.9	0.18	302.6006
Germany	6.994	5.00	0.90	0.93	45.00	95.00	13.07	48.80	78.13	74.40	71.70	78.03	0.78	4.72	11.42	47400	0.20	80.57	29.91	1.71	5410.635
Greece	5.033	26.30	1.10	0.02	27.70	83.00	10.26	49.60	113.01	53.20	171.30	61.38	0.69	6.42	11.09	25600	-1.40	80.43	47.22	0.93	1743.038
Honduras	4.871	3.90	84.60	-0.51	17.20	45.00	5.48	44.30	56.11	57.70	47.40	42.45	0.69	23.46	5.17	5000	4.20	71.00	75.2	1.16	212.3152
Hungary	5.145	7.80	1.50	0.70	47.90	79.00	11.64	48.40	36.42	66.00	75.50	59.88	0.67	4.54	12.73	26000	0.10	75.69	46.7	1.12	1036.624
Iceland	7.501	5.00	0.30	1.24	41.30	100.00	10.59	29.60	92.06	73.30	81.90	52.95	0.88	3.78	6.28	46600	1.70	82.97	13.1	4.78	4661.621
India	4.404	3.60	3.20	-0.96	10.80	77.00	5.39	60.60	52.65	56.20	51.70	57.03	0.66	5.01	7.32	6300	5.60	68.13	77.02	1.73	74.9946
Indonesia	5.314	6.20	0.50	-0.37	14.10	65.00	7.59	29.70	39.07	59.40	27.70	21.43	0.68	5.72	6.37	11300	6.70	72.45	75.14	0.16	99.41036
Iran, Islamic Rep.	4.813	12.80	3.90	-0.91	14.10	17.00	8.17	44.10	54.41	43.50	13.20	16.63	0.58	7.02	5.94	17800	15.30	71.15	84.21	0.09	350.7386
Ireland	6.907	11.60	1.10	1.07	34.50	96.00	12.20	25.90	64.78	77.30	101.20	57.33	0.81	5.25	6.48	54300	-0.20	80.68	30.85	3.55	4239.154
Israel	7.267	6.10	1.70	-0.99	25.50	80.00	12.54	30.60	67.36	70.70	64.40	46.28	0.71	10.28	5.15	34300	-0.60	82.27	61.46	2.15	2910.289
Italy	5.977	12.50	0.80	0.50	48.20	89.00	10.10	64.80	88.37	61.20	135.80	64.25	0.73	6.88	10.19	35800	0.30	82.12	56.56	1.79	3257.573
Japan	5.921	3.70	0.29	1.02	35.00	96.00	11.52	51.30	194.30	73.10	227.90	50.85	0.67	5.39	9.51	38200	0.70	84.74	34.04	1.39	3702.953
Jordan	5.303	11.10	2.45	-0.56	22.80	36.00	9.91	29.50	70.25	68.30	79.20	35.88	0.59	5.14	3.79	12400	-0.70	74.35	82.72	0.1	358.9125
Kazakhstan	5.919	4.10	7.82	0.05	17.50	24.00	11.42	29.20	37.73	63.60	17.60	51.08	0.72	4.05	8.21	24700	5.30	70.55	72	0.89	538.7828
Kenya	4.356	9.20	5.90	-1.27	16.80	51.00	6.27	37.10	34.89	57.50	48.60	37.33	0.72	10.99	6.89	3300	6.40	63.77	62.2	0.0251	77.69659
Lao PDR	4.876	1.40	5.90	0.46	24.70	12.00	5.00	25.30	20.76	49.80	48.60	44.63	0.71	5.81	7.63	5400	1.30	63.88	68.56	0.02	32.56967
Latvia	5.56	10.00	3.90	0.55	33.80	86.00	11.48	35.90	48.69	70.40	38.50	59.63	0.75	6.88	14.31	24500	0.70	74.23	37.14	1.31	920.7038
Lithuania	5.813	11.30	5.50	0.78	31.90	91.00	12.38	42.60	41.73	75.20	38.80	52.63	0.74	5.68	14.27	28000	-1.00	74.69	30.94	1.13	1063.421
Macedonia, FYR	5.121	27.90	1.60	0.25	26.80	57.00	9.26	12.90	50.86	67.50	40.30	54.13	0.70	9.29	9.08	14000	-0.10	76.02	85.29	0.231	353.9253
Lithuania	5.813	11.30	5.50	0.78	31.90	91.00	12.38	42.60	41.73	75.20	38.80	52.63	0.74	5.68	14.27	28000	-1.00	74.69	30.94	1.13	1063.421
Macedonia, FYR	5.121	27.90	1.60	0.25	26.80	57.00	9.26	12.90	50.86	67.50	40.30	54.13	0.70	9.29	9.08	14000	-0.10	76.02	85.29	0.231	353.9253
Malaysia	6.005	2.00	2.30	0.34	16.90	45.00	9.96	40.00	125.24	71.50	53.50	25.08	0.66	11.33	5.03	26600	2.10	74.75	67.53	2.26	455.8255
Mauritius	5.648	7.70	2.70	0.74	19.10	90.00	8.54	22.40	104.33	74.70	61.10	45.75	0.65	5.90	6.91	19500	3.40	75.40	47.38	1.08	482.4542
Mexico	6.778	4.90	15.70	-0.76	22.30	65.00	8.47	51.70	32.69	65.20	45.20	75.53	0.70	11.13	5.26	18500	2.70	75.65	64.64	2.25	677.19
Moldova	5.897	3.40	3.20	-0.10	36.80	60.00	11.19	40.20	34.79	57.40	51.90	52.88	0.74	4.63	12.59	5000	9.10	70.42	57.24	0.22	228.8467
Mongolia	4.907	4.80	7.50	0.87	32.40	86.00	9.28	24.40	54.75	59.40	72.00	48.08	0.71	6.20	6.35	12500	8.30	69.29	90.96	0.54	195.3335
Morocco	5.151	10.20	1.00	-0.39	23.80	41.00	4.37	49.10	64.46	61.30	73.40	27.08	0.59	7.34	4.81	8300	1.70	76.71	66.87	0.06	190.054
Netherlands	7.339	6.90	0.70	1.05	44.80	99.00	11.89	41.00	111.15	74.60	68.90	91.70	0.78	4.46	8.66	49300	0.30	81.23	32.79	2.03	5693.86
Norway	7.498	3.40	0.60	1.13	55.40	100.00	12.63	39.50	84.85	70.80	39.30	59.63	0.85	4.00	8.12	68400	2.00	81.70	18.21	2.62	9522.222
Peru	5.743	4.20	6.70	-0.52	34.60	71.00	9.01	35.90	36.84	67.40	19.80	59.08	0.68	11.87	6.01	12300	3.60	73.48	87.91	1.28	358.5769
Philippines	5.279	7.10	9.90	-0.70	16.00	65.00	8.88	42.90	41.88	63.10	44.80	39.63	0.79	8.43	6.11	7500	1				

Отсеянные данные

Country	Happy	UR	IH	TTR	MYoS	TTRate	DCPS	EF	PD	MF	GGGI	DR	Infl
Albania	4.655	16.1	4	25.7	9.26	36.5	35.28	65.9	73.3	53.5	0.7	6.58	2.2
Algeria	6.355	9.5	1.5	28.2	7.61	72.7	21.6	50.1	9.3	20.63	0.63	4.31	5.1
Angola	3.866	6.8	10.8	34.7	4.73	48.4	27.22	48.9	56.7	36.5	0.64	11.49	10.1
Argenti	6.65	8.2	7.6	24.8	9.83	137.4	14.3	43.8	45.8	64.45	0.73	7.33	27.6
Armenia	4.36	17.1	2.02	24	10.87	19.9	45.6	67	47.5	49.58	0.67	9.34	5.6
Australia	7.313	6	1	34.3	12.96	47.6	137.67	80.3	44.3	61.35	0.73	7.14	1.9
Austria	7.119	5	0.5	50.6	10.83	51.7	87.67	71.7	83.4	71.13	0.73	9.42	0.9
Azerbaijan	5.291	5.2	2.5	25.2	11.16	39.8	38.45	60.2	11.8	42.68	0.68	7.07	3.9
Bahrain	6.218	4.1	0.5	16.7	9.42	13.5	71.21	74.3	66.7	21.63	0.64	2.69	2
Bangladesh	4.643	4.3	2.8	11	5.07	31.6	43.93	53.3	31.9	28.9	0.7	5.61	5.7
Belarus	5.802	5.9	3.6	35.2	11.98	51.8	28.83	48.8	36.4	46.08	0.73	13.36	15
Belgium	6.929	8.5	1.8	49.5	11.27	58.4	61.69	68.4	107	79.35	0.75	9.63	0.5
Bolivia	5.822	2.7	12.1	48.6	8.15	83.7	57.71	47.4	37.5	65.3	0.75	6.52	4.2
Botswa	3.974	18.2	15.4	38.8	8.87	25.1	33.85	71.1	17	39.88	0.71	13.39	4
Brazil	6.952	6.8	24.6	35.1	7.66	69.2	67.86	56.5	67.3	69.3	0.69	6.58	10.6
Bulgaria	4.217	11.6	1.6	37.8	10.57	27	56.82	65.9	26.4	53.33	0.72	14.44	0.1
Cambodia	3.907	0.4	6.5	18.8	4.37	21	63.1	57.9	33.9	70.5	0.62	7.68	1.4
Cameroon	4.513	4.3	2.8	14.1	5.96	48.8	15.95	54.2	31.7	46.25	0.68	10.11	2
Cada	7.404	6.9	1.4	37.2	13	21.1	124.41	78	95.4	76.58	0.74	8.42	1.2
Chile	6.705	6.4	3.6	20.2	9.79	28.9	111.22	77.7	17.4	59.4	0.7	6	4.4
Chi	5.245	4.7	0.8	21.3	7.54	67.8	155.33	52	16.7	40	0.68	7.53	1.5
Colombia	6.481	10.1	27.9	29.3	7.35	69.7	47.13	70.8	46.9	74.98	0.73	5.4	4.4
Costa Rica	7.087	8.3	10	14.5	8.37	58	60.16	67.4	59.7	53.08	0.73	4.55	1.1
Croatia	5.488	16.7	0.8	41.9	11.03	20	65.47	59.1	89.5	59.13	0.71	12.18	-0.3
Cyprus	5.546	15.6	1.1	39.9	11.62	24.4	250.76	68.7	108.4	53.73	0.67	6.62	-0.3
Czech Republic	6.596	6.2	0.7	39.6	12.32	50.4	51.24	73.2	41.4	80.5	0.69	10.34	0.4
Denmark	7.526	6.6	1	55.6	12.73	24.5	174.04	75.3	47.2	66.33	0.77	10.25	0.6
Dominican Republic	5.155	15	17.4	16	7.56	42.4	27.54	61	44.7	38.9	0.69	4.55	0.7
Ecuador	5.976	4.6	8.2	35.5	7.59	33	27.38	48.6	32.4	64.75	0.74	5.06	4
Egypt. Arab Rep.	4.362	13.2	3.4	26.7	6.55	45	26.47	56	91.7	16.88	0.6	4.77	10.4
El Salvador	6.068	6.2	64.2	20	6.51	38.7	44.86	65.1	64.9	37.7	0.71	5.69	-0.9
Estonia	5.517	7.7	3.1	38.2	12.48	49.4	69.54	77.2	10.2	69.4	0.75	12.4	0.5
Ethiopia	4.508	5.2	8.1	15.8	2.41	32.1	17.71	51.5	45.8	34.63	0.64	8.19	10.3
Finland	7.413	8.6	1.6	57	10.29	37.9	96.34	72.6	61.2	60.58	0.85	9.83	-0.2
France	6.478	9.9	1.2	51.7	11.13	62.7	95.73	62.3	98.2	69.93	0.76	9.16	0.1
Georgia	4.252	13.4	2.7	28.2	12.11	16.4	49.76	72.6	38.8	46.63	0.69	10.82	4.1
Germany	6.994	5	0.9	45	13.07	48.8	78.13	74.4	71.7	78.03	0.78	11.42	0.2
Gha	4.276	2.4	1.7	21.6	7	32.7	20.27	63	76	49	0.7	7.22	16.9
Greece	5.033	26.3	1.1	27.7	10.26	49.6	113.01	53.2	171.3	61.38	0.69	11.09	-1.4
Guatemala	6.324	2.9	31.2	11.5	5.6	37.5	34.37	61.8	24.2	43.45	0.67	4.77	2.2
Honduras	4.871	3.9	84.6	17.2	5.48	44.3	56.11	57.7	47.4	42.45	0.69	5.17	4.2
Hungary	5.145	7.8	1.5	47.9	11.64	48.4	36.42	66	75.5	59.88	0.67	12.73	0.1
Iceland	7.501	5	0.3	41.3	10.59	29.6	92.06	73.3	81.9	52.95	0.88	6.28	1.7
India	4.404	3.6	3.2	10.8	5.39	60.6	52.65	56.2	51.7	57.03	0.66	7.32	5.6
Indonesia	5.314	6.2	0.5	14.1	7.59	29.7	39.07	59.4	27.7	21.43	0.68	6.37	6.7
Iran. Islamic Rep.	4.813	12.8	3.9	14.1	8.17	44.1	54.41	43.5	13.2	16.63	0.58	5.94	15.3
Ireland	6.907	11.6	1.1	34.5	12.2	25.9	64.78	77.3	101.2	57.33	0.81	6.48	-0.2
Israel	7.267	6.1	1.7	25.5	12.54	30.6	67.36	70.7	64.4	46.28	0.71	5.15	-0.6
Italy	5.977	12.5	0.8	48.2	10.1	64.8	88.37	61.2	135.8	64.25	0.73	10.19	0.3
Jamaica	5.51	13.2	42.88	26.8	9.67	35.2	30.44	67.5	122.5	53.88	0.7	6.7	4.2
Japan	5.921	3.7	0.29	35	11.52	51.3	194.3	73.1	227.9	50.85	0.67	9.51	0.7
Jordan	5.303	11.1	2.45	22.8	9.91	29.5	70.25	68.3	79.2	35.88	0.59	3.79	-0.7
Kazakhstan	5.919	4.1	7.82	17.5	11.42	29.2	37.73	63.6	17.6	51.08	0.72	8.21	5.3
Kenya	4.356	9.2	5.9	16.8	6.27	37.1	34.89	57.5	48.6	37.33	0.72	6.89	6.4
Korea. Rep.	5.835	3.5	0.7	20.9	11.89	33.2	140.57	71.7	34.9	44.88	0.65	6.75	0.7
Kuwait	6.239	3	0.4	49.6	7.21	13	99.67	62.7	9.5	15.93	0.65	2.18	3.4
Lao PDR	4.876	1.4	5.9	24.7	5	25.3	20.76	49.8	48.6	44.63	0.71	7.63	1.3
Latvia	5.56	10	3.9	33.8	11.48	35.9	48.69	70.4	38.5	59.63	0.75	14.31	0.7
Lebanon	5.129	6.4	4.68	20.1	7.92	30.3	106.6	59.5	138.8	33.9	0.6	4.88	-3.5
Liberia	3.622	3.8	3.3	33.8	4.09	47.8	20.78	52.2	6.7	45.88	0.65	9.69	8.2
Lithuania	5.813	11.3	5.5	31.9	12.38	42.6	41.73	75.2	38.8	52.63	0.74	14.27	-1
Luxembourg	6.871	6.1	0.7	40.8	11.71	20.1	95.36	73.9	21.7	72.6	0.74	7.24	0.1
Macedonia. FYR	5.121	27.9	1.6	26.8	9.26	12.9	50.86	67.5	40.3	54.13	0.7	9.08	-0.1
Malawi	4.156	7.5	1.8	18.6	4.29	34.5	12.2	51.8	61.1	48.83	0.7	8.41	22.2
Malaysia	6.005	2	2.3	16.9	9.96	40	125.24	71.5	53.5	25.08	0.66	5.03	2.1
Mali	4.073	8.1	11.2	22	2.04	48.3	22.55	56.5	39	41.55	0.6	12.89	1.5
Malta	6.488	5.9	1.4	41.3	10.33	41.3	111.8	66.7	60.6	45.2	0.67	9.09	1.2
Mauritius	5.648	7.7	2.7	19.1	8.54	22.4	104.33	74.7	61.1	45.75	0.65	6.91	3.4
Mexico	6.778	4.9	15.7	22.3	8.47	51.7	32.69	65.2	45.2	75.53	0.7	5.26	2.7

Moldova	5.897	3.4	3.2	36.8	11.19	40.2	34.79	57.4	51.9	52.88	0.74	12.59	9.1
Mongolia	4.907	4.8	7.5	32.4	9.28	24.4	54.75	59.4	72	48.08	0.71	6.35	8.3
Morocco	5.151	10.2	1	23.8	4.37	49.1	64.46	61.3	73.4	27.08	0.59	4.81	1.7
Nepal	4.793	2.7	2.9	19.9	3.31	29.5	64.92	50.9	30	49.58	0.66	6.56	7.2
Netherlands	7.339	6.9	0.7	44.8	11.89	41	111.15	74.6	68.9	91.7	0.78	8.66	0.3
New Zealand	7.334	5.6	0.9	43.1	12.5	34.3	142.33	81.6	33.5	65.25	0.78	7.36	0.7
Nicaragua	5.992	5.3	11.3	16	6	63.9	37.06	58.6	42.6	36.33	0.78	5.08	4.2
Nigeria	4.875	7.5	10.3	2.9	5.95	33.3	14.54	57.5	11.7	28.03	0.64	12.9	9.5
Norway	7.498	3.4	0.6	55.4	12.63	39.5	84.85	70.8	39.3	59.63	0.85	8.12	2
Pakistan	5.132	5.2	7.7	15.6	4.73	32.6	15.44	55.9	64.8	18.05	0.56	6.49	4.5
Pama	6.701	4.3	17.2	21.4	9.35	37.2	86.8	64.8	39.6	51.75	0.77	4.81	0.5
Paraguay	5.538	4.5	8.8	18.2	7.7	35	57.12	61.5	19.9	54.13	0.67	4.68	3.4
Peru	5.743	4.2	6.7	34.6	9.01	35.9	36.84	67.4	19.8	59.08	0.68	6.01	3.6
Philippines	5.279	7.1	9.9	16	8.88	42.9	41.88	63.1	44.8	39.63	0.79	6.11	1.8
Poland	5.835	9.2	0.7	17	11.82	40.3	53.91	69.3	43.4	50.08	0.72	10.19	-0.8
Portugal	5.123	14.2	0.9	45.5	8.25	41	120.06	65.1	129	83.8	0.73	11.02	0.6
Qatar	6.375	0.3	1.1	40.2	9.07	11.3	68.65	70.7	39.9	15.63	0.65	1.53	1.6
Romania	5.528	7	1.5	32.9	10.78	42	29.89	65.6	39.9	52	0.69	11.9	-0.6
Russian Federation	5.856	5.1	9.5	17.5	11.95	47	56.36	50.6	13.5	52.88	0.69	13.69	15.4
Rwanda	3.515	0.6	4.6	21.9	3.74	33	10.41	63.1	33.7	36.58	0.79	8.96	2.3
Saudi Arabia	6.379	5.6	0.8	28.3	8.7	15	56.63	62.1	7.8	7.75	0.61	3.33	2.3
Senegal	4.219	10	8.1	24.3	2.48	47.3	33.56	58.1	55.7	47.63	0.7	8.46	-0.4
Serbia	5.177	22.2	1.3	40.8	10.46	39.7	44.14	62.1	75	53.13	0.72	13.66	1.9
Singapore	6.739	3	0.3	15	10.63	18.4	129.75	87.8	105.6	46.08	0.71	3.43	-0.5
Slovak Republic	6.078	13.3	1.1	40.2	12.23	51.2	53.88	66.6	52.5	62.33	0.68	9.74	-0.2
Slovenia	5.768	9.5	0.7	40.7	11.89	31	50.23	60.6	62.7	70	0.78	11.37	-0.4
South Africa	4.459	25.1	33	26.5	9.94	28.8	150.02	61.9	45.4	61.7	0.76	9.91	4.8
Spain	6.361	24.7	0.7	38.8	9.58	50	118.87	68.5	101	78.6	0.74	9.04	-0.6
Sri Lanka	4.415	4.6	2.79	12.3	10.8	55.2	30.83	59.9	74.7	21.9	0.69	6.11	0.8
Switzerland	7.509	4.5	0.5	32.8	12.82	28.8	174.15	81	34	72.38	0.79	8.13	-1.1
Tajikistan	4.996	10.9	1.46	30.2	10.36	81.8	22.17	51.3	6.5	48.38	0.68	6.18	8.9
Thailand	6.474	0.9	3.9	19.1	7.32	27.5	151.26	63.9	50.6	30.75	0.71	7.8	-0.8
Trinidad and Tobago	6.168	4	30.22	31.8	10.88	32.2	31.44	62.9	42.1	45.75	0.72	8.56	7.5
Tunisia	5.045	13.3	3.1	24.6	6.84	59.9	79.6	57.6	52.6	42.58	0.63	5.98	4.7
Turkey	5.389	9.2	4.3	24.3	7.56	40.9	80	62.1	33.1	50.78	0.62	5.88	7.5
Uganda	3.739	3.8	11.8	13.2	5.42	36.5	15.23	59.3	34.8	41.88	0.71	10.69	4.9
Ukraine	4.324	7.7	4.3	28.2	11.34	52.2	56.97	46.8	94.9	47.58	0.7	14.46	49
United Arab Emirates	6.573	3.6	0.7	32.5	9.49	15.9	76.48	72.6	52.1	15.38	0.65	1.97	3.7
United Kingdom	6.725	6.3	0.9	38.4	13.05	32	134.47	76.4	90.6	58.05	0.76	9.35	0.1
United States	7.104	6.2	3.9	18.1	12.94	43.9	190.36	75.4	73.6	78.2	0.74	8.15	0.2
Uruguay	6.545	7	7.8	29	8.45	41.8	30.02	68.8	68.6	88.75	0.68	9.45	8.6
Vietm	5.061	2.3	3.3	19.9	7.45	39.4	111.93	54	52.7	33.88	0.69	5.93	0.9
Zimbabwe	4.193	5.4	7.5	26.8	7.25	32.8	27.11	38.2	205.3	44.08	0.71	10.13	0.1

Фиксированный список государств, по которому строится модель 2.1 п. 2.3.2

1	Afghanistan	44	Ethiopia	87	Mauritania	130	Tajikistan
2	Albania	45	Finland	88	Mauritius	131	Tanzania
3	Algeria	46	France	89	Mexico	132	Thailand
4	Angola	47	Gabon	90	Moldova	133	Togo
5	Argentina	48	Georgia	91	Mongolia	134	Trinidad and Tobago
6	Armenia	49	Germany	92	Montenegro	135	Tunisia
7	Australia	50	Ghana	93	Morocco	136	Turkey
8	Austria	51	Greece	94	Myanmar	137	Turkmenistan
9	Azerbaijan	52	Guatemala	95	mibia	138	Uganda
10	Bahrain	53	Guinea	96	Nepal	139	Ukraine
11	Bangladesh	54	Guinea-Bissau	97	Netherlands	140	United Arab Emirates
12	Belarus	55	Haiti	98	New Zealand	141	United Kingdom
13	Belgium	56	Honduras	99	Nicaragua	142	United States
14	Belize	57	Hungary	100	Niger	143	Uruguay
15	Benin	58	Iceland	101	Nigeria	144	Uzbekistan
16	Bhutan	59	India	102	Norway	145	Venezuela. RB
17	Bolivia	60	Indonesia	103	Pakistan	146	Vietnam
18	Bosnia and Herzegovina	61	Iran. Islamic Rep.	104	Panama	147	Yemen. Rep.
19	Botswana	62	Iraq	105	Paraguay	148	Zambia
20	Brazil	63	Ireland	106	Peru	149	Zimbabwe
21	Bulgaria	64	Israel	107	Philippines		
22	Burki Faso	65	Italy	108	Poland		
23	Burundi	66	Jamaica	109	Portugal		
24	Cambodia	67	Japan	110	Qatar		
25	Cameroon	68	Jordan	111	Romania		
26	Canada	69	Kazakhstan	112	Russian Federation		
27	Chad	70	Kenya	113	Rwanda		
28	Chile	71	Korea. Rep.	114	Saudi Arabia		
29	China	72	Kuwait	115	Senegal		
30	Colombia	73	Kyrgyz Republic	116	Serbia		
31	Comoros	74	Lao PDR	117	Sierra Leone		
32	Congo. Dem. Rep.	75	Latvia	118	Singapore		
33	Congo. Rep.	76	Lebanon	119	Slovak Republic		
34	Costa Rica	77	Liberia	120	Slovenia		
35	Croatia	78	Libya	121	South Africa		
36	Cyprus	79	Lithuania	122	South Sudan		
37	Czech Republic	80	Luxembourg	123	Spain		
38	Denmark	81	Macedonia. FYR	124	Sri Lanka		
39	Dominican Republic	82	Madagascar	125	Sudan		
40	Ecuador	83	Malawi	126	Suriname		
41	Egypt. Arab Rep.	84	Malaysia	127	Sweden		
42	El Salvador	85	Mali	128	Switzerland		
43	Estonia	86	Malta	129	Syrian Arab Republic		

Фиксированный список государств, по которому строится модель 2.2 п. 2.3.2

1	Afghanistan	44	Ethiopia	87	Mauritania	130	Trinidad and Tobago
2	Albania	45	Finland	88	Mauritius	131	Tunisia
3	Algeria	46	France	89	Mexico	132	Turkey
4	Angola	47	Gabon	90	Moldova	133	Turkmenistan
5	Argentina	48	Georgia	91	Mongolia	134	Uganda
6	Armenia	49	Germany	92	Morocco	135	Ukraine
7	Australia	50	Ghana	93	mibia	136	United Arab Emirates
8	Austria	51	Greece	94	Nepal	137	United Kingdom
9	Azerbaijan	52	Guatemala	95	Netherlands	138	United States
10	Bahrain	53	Guinea	96	New Zealand	139	Uruguay
11	Bangladesh	54	Guinea-Bissau	97	Nicaragua	140	Uzbekistan
12	Belarus	55	Haiti	98	Niger	141	Vietnam
13	Belgium	56	Honduras	99	Nigeria	142	Yemen. Rep.
14	Belize	57	Hungary	100	Norway	143	Zambia
15	Benin	58	Iceland	101	Pakistan	144	Zimbabwe
16	Bhutan	59	India	102	Panama		
17	Bolivia	60	Indonesia	103	Paraguay		
18	Bosnia and Herzegovina	61	Iran. Islamic Rep.	104	Peru		
19	Botswana	62	Iraq	105	Philippines		
20	Brazil	63	Ireland	106	Poland		
21	Bulgaria	64	Israel	107	Portugal		
22	Burki Faso	65	Italy	108	Qatar		
23	Burundi	66	Jamaica	109	Romania		
24	Cambodia	67	Japan	110	Russian Federation		
25	Cameroon	68	Jordan	111	Rwanda		
26	Canada	69	Kazakhstan	112	Saudi Arabia		
27	Chad	70	Kenya	113	Senegal		
28	Chile	71	Korea. Rep.	114	Serbia		
29	China	72	Kuwait	115	Sierra Leone		
30	Colombia	73	Kyrgyz Republic	116	Singapore		
31	Comoros	74	Lao PDR	117	Slovak Republic		
32	Congo. Dem. Rep.	75	Latvia	118	Slovenia		
33	Congo. Rep.	76	Lebanon	119	South Africa		
34	Costa Rica	77	Liberia	120	Spain		
35	Croatia	78	Libya	121	Sri Lanka		
36	Cyprus	79	Lithuania	122	Sudan		
37	Czech Republic	80	Luxembourg	123	Suriname		
38	Denmark	81	Macedonia. FYR	124	Switzerland		
39	Dominican Republic	82	Madagascar	125	Syrian Arab Republic		
40	Ecuador	83	Malawi	126	Tajikistan		
41	Egypt. Arab Rep.	84	Malaysia	127	Tanzania		
42	El Salvador	85	Mali	128	Thailand		
43	Estonia	86	Malta	129	Togo		

Фиксированный список государств, по которому строится модель 2.3 п. 2.3.2

1	Afghanistan	44	Guinea	87	Nicaragua
2	Albania	45	Guinea-Bissau	88	Nigeria
3	Algeria	46	Haiti	89	Norway
4	Angola	47	Honduras	90	Pakistan
5	Argentina	48	Hungary	91	Panama
6	Armenia	49	Iceland	92	Paraguay
7	Australia	50	India	93	Peru
8	Austria	51	Indonesia	94	Philippines
9	Azerbaijan	52	Iran. Islamic Rep.	95	Poland
10	Bahrain	53	Iraq	96	Portugal
11	Bangladesh	54	Ireland	97	Qatar
12	Belarus	55	Israel	98	Romania
13	Belgium	56	Italy	99	Russian Federation
14	Bolivia	57	Jamaica	100	Rwanda
15	Bosnia and Herzegovina	58	Japan	101	Saudi Arabia
16	Botswana	59	Jordan	102	Senegal
17	Brazil	60	Kazakhstan	103	Serbia
18	Bulgaria	61	Kenya	104	Singapore
19	Cambodia	62	Korea. Rep.	105	Slovak Republic
20	Cameroon	63	Kuwait	106	Slovenia
21	Canada	64	Kyrgyz Republic	107	South Africa
22	Chile	65	Lao PDR	108	Spain
23	China	66	Latvia	109	Sri Lanka
24	Colombia	67	Lebanon	110	Sudan
25	Comoros	68	Liberia	111	Suriname
26	Costa Rica	69	Libya	112	Switzerland
27	Croatia	70	Lithuania	113	Syrian Arab Republic
28	Cyprus	71	Luxembourg	114	Tajikistan
29	Czech Republic	72	Macedonia. FYR	115	Thailand
30	Denmark	73	Madagascar	116	Trinidad and Tobago
31	Dominican Republic	74	Malawi	117	Tunisia
32	Ecuador	75	Malaysia	118	Turkey
33	Egypt. Arab Rep.	76	Mali	119	Turkmenistan
34	El Salvador	77	Malta	120	Uganda
35	Estonia	78	Mauritania	121	Ukraine
36	Ethiopia	79	Mauritius	122	United Arab Emirates
37	Finland	80	Mexico	123	United Kingdom
38	France	81	Moldova	124	United States
39	Georgia	82	Mongolia	125	Uruguay
40	Germany	83	Morocco	126	Uzbekistan
41	Ghana	84	Nepal	127	Vietnam
42	Greece	85	Netherlands	128	Yemen. Rep.
43	Guatemala	86	New Zealand	129	Zimbabwe

Таблица результатов проверки данных на выбросы и наблюдения, оказывающие наиболее сильное влияние на регрессионную модель

Country	dfbetas.1	dfbetas.MYoS	dfbetas.DR	dfbetas.MF	dffit	Cook distance	Leverage
Afghanistan	0.0635	-0.0991	0.2676	-0.1805	0.3438	0.0295	0.1045
Albania	-0.0529	-0.0100	0.1078	-0.0749	-0.2108	0.0109	0.0115
Algeria	0.2014	0.0620	-0.0946	-0.1734	0.2823	0.0197	0.0313
Angola	0.0023	-0.0047	0.0065	-0.0023	0.0095	0.0000	0.0394
Argentina	-0.0042	-0.0036	-0.0350	0.0567	0.0856	0.0018	0.0171
Armenia	0.0779	-0.1556	-0.0917	0.0892	-0.2608	0.0165	0.0148
Australia	-0.0353	0.0749	-0.0349	0.0145	0.1121	0.0032	0.0268
Austria	-0.0825	0.0151	-0.0002	0.1173	0.1922	0.0092	0.0208
Azerbaijan	-0.0010	-0.1175	0.0177	0.0731	-0.1609	0.0065	0.0188
Bahrain	0.0438	0.0354	-0.0411	-0.0414	0.0797	0.0016	0.0471
Bangladesh	-0.0132	0.0066	0.0035	0.0032	-0.0144	0.0001	0.0228
Belarus	-0.0840	0.0931	0.1363	-0.0949	0.1768	0.0078	0.0541
Belgium	-0.0578	0.0025	-0.0074	0.0906	0.1256	0.0040	0.0316
Bolivia	-0.0082	0.0165	0.0208	-0.0321	-0.0405	0.0004	0.0258
Bosnia and Herzegovina	0.0015	0.0062	-0.0084	-0.0061	-0.0221	0.0001	0.0118
Botswana	0.0417	-0.0271	-0.1520	0.0905	-0.1721	0.0074	0.0469
Brazil	0.0526	-0.1400	-0.1342	0.2380	0.2812	0.0196	0.0350
Bulgaria	0.1302	-0.0642	-0.2300	0.0829	-0.2604	0.0169	0.0501
Cambodia	-0.1326	0.3853	0.1094	-0.3803	-0.4771	0.0559	0.0659
Cameroon	-0.0010	0.0022	-0.0019	-0.0002	-0.0038	0.0000	0.0203
Canada	-0.0576	0.0448	-0.0272	0.0634	0.1197	0.0036	0.0342
Chile	0.0117	0.0037	-0.0499	0.0420	0.0789	0.0016	0.0182
China	0.0067	-0.0018	0.0003	-0.0032	0.0116	0.0000	0.0096
Colombia	0.0219	-0.0514	-0.0595	0.0901	0.1034	0.0027	0.0584
Comoros	-0.0552	0.0382	-0.0126	0.0246	-0.0737	0.0014	0.0250
Costa Rica	0.1064	-0.0371	-0.1755	0.1061	0.2236	0.0124	0.0231
Croatia	0.0107	-0.0062	-0.0132	0.0021	-0.0188	0.0001	0.0262
Cyprus	0.0180	-0.1185	0.0734	-0.0030	-0.1911	0.0091	0.0186
Czech Republic	-0.0002	0.0001	0.0000	0.0002	0.0004	0.0000	0.0347
Denmark	-0.1690	0.1516	0.0639	0.0339	0.2682	0.0177	0.0252
Dominican Republic	-0.0719	0.0094	0.0634	0.0003	-0.0946	0.0022	0.0181
Ecuador	-0.0116	0.0152	0.0232	-0.0270	-0.0351	0.0003	0.0375
Egypt. Arab Rep.	-0.1170	-0.0112	0.0310	0.0993	-0.1509	0.0057	0.0331
El Salvador	0.1348	-0.0584	-0.0674	-0.0052	0.1590	0.0063	0.0150
Estonia	0.1076	-0.0606	-0.0817	-0.0151	-0.1507	0.0057	0.0339
Ethiopia	0.1425	-0.1800	0.0327	0.0142	0.2162	0.0117	0.0437
Finland	-0.1121	0.0540	0.0833	0.0678	0.2753	0.0183	0.0132
France	-0.0156	0.0063	-0.0014	0.0200	0.0359	0.0003	0.0200
Georgia	0.1738	-0.2851	-0.2237	0.2195	-0.4126	0.0412	0.0320
Germany	-0.0956	0.0486	0.0355	0.0504	0.1342	0.0045	0.0368
Ghana	-0.0894	0.0877	0.0395	-0.0586	-0.1627	0.0066	0.0119

Greece	0.0577	-0.0176	-0.0641	-0.0171	-0.1165	0.0034	0.0178
Guatemala	0.2004	-0.1471	-0.1496	0.0877	0.2532	0.0158	0.0253
Guinea	-0.0507	0.1069	-0.0206	-0.0441	-0.1179	0.0035	0.0556
Guinea-Bissau	0.0192	-0.1354	0.1490	0.0034	0.2134	0.0114	0.0862
Haiti	-0.0808	0.1033	-0.0016	-0.0321	-0.1306	0.0043	0.0218
Honduras	-0.0514	0.0390	0.0331	-0.0194	-0.0636	0.0010	0.0233
Hungary	0.0804	-0.0520	-0.0936	0.0212	-0.1305	0.0043	0.0328
Iceland	0.0180	0.0921	-0.1113	0.0310	0.2187	0.0118	0.0150
India	-0.0775	0.1390	0.0423	-0.1144	-0.1762	0.0078	0.0296
Indonesia	0.0479	0.0188	0.0024	-0.0639	0.0823	0.0017	0.0265
Iran. Islamic Rep.	-0.0377	-0.0271	-0.0006	0.0618	-0.0743	0.0014	0.0368
Iraq	-0.0949	-0.0104	0.0356	0.0751	-0.1189	0.0036	0.0419
Ireland	-0.0071	0.0273	-0.0185	0.0047	0.0440	0.0005	0.0230
Israel	0.0019	0.1010	-0.0619	-0.0303	0.1366	0.0047	0.0336
Italy	-0.0142	0.0016	0.0103	0.0132	0.0334	0.0003	0.0156
Jamaica	-0.0143	-0.0144	0.0411	-0.0261	-0.0860	0.0019	0.0116
Japan	0.0010	-0.0018	-0.0009	0.0009	-0.0026	0.0000	0.0182
Jordan	-0.1058	-0.1003	0.1438	0.0635	-0.2284	0.0130	0.0286
Kazakhstan	0.0094	-0.0267	-0.0012	0.0088	-0.0384	0.0004	0.0154
Kenya	-0.0772	0.0439	0.0111	0.0136	-0.0988	0.0024	0.0134
Korea. Rep.	0.0067	-0.0888	0.0221	0.0426	-0.1138	0.0033	0.0230
Kuwait	0.2028	0.0440	-0.1404	-0.1275	0.2624	0.0172	0.0484
Kyrgyz Republic	-0.0166	-0.0954	0.0499	0.0359	-0.1622	0.0065	0.0142
Lao PDR	0.0250	-0.0311	-0.0015	0.0108	0.0395	0.0004	0.0211
Latvia	-0.0468	0.0243	0.0640	-0.0163	0.0775	0.0015	0.0479
Lebanon	-0.0605	-0.0080	0.0429	0.0249	-0.0827	0.0017	0.0176
Liberia	-0.0631	0.1246	-0.0482	-0.0325	-0.1540	0.0059	0.0332
Libya	0.0493	0.0086	-0.0301	-0.0307	0.0632	0.0010	0.0324
Lithuania	-0.0918	0.0786	0.1293	-0.0676	0.1617	0.0066	0.0593
Luxembourg	-0.0047	0.0037	-0.0075	0.0111	0.0176	0.0001	0.0289
Macedonia. FYR	0.0075	-0.0034	-0.0160	-0.0094	-0.0648	0.0011	0.0092
Madagascar	-0.1772	0.1784	0.0744	-0.0942	-0.2670	0.0174	0.0161
Malawi	-0.0446	0.0767	-0.0051	-0.0345	-0.0892	0.0020	0.0303
Malaysia	0.0236	0.0352	-0.0145	-0.0418	0.0611	0.0009	0.0342
Mali	0.0986	-0.3044	0.2506	0.0130	0.4225	0.0442	0.0792
Malta	-0.0237	0.0843	0.0556	-0.0723	0.1516	0.0057	0.0141
Mauritania	0.0370	-0.0338	0.0137	-0.0131	0.0538	0.0007	0.0316
Mauritius	0.0028	0.0002	-0.0019	-0.0001	0.0061	0.0000	0.0087
Mexico	0.0155	-0.0381	-0.0655	0.0900	0.1050	0.0028	0.0533
Moldova	-0.0697	0.0566	0.1037	-0.0467	0.1353	0.0046	0.0332
Mongolia	-0.0535	-0.0305	0.0748	-0.0114	-0.1531	0.0058	0.0108
Morocco	0.1242	-0.0669	-0.0446	-0.0186	0.1300	0.0042	0.0297
Nepal	0.0374	-0.0536	-0.0148	0.0311	0.0609	0.0009	0.0447
Netherlands	-0.0466	-0.0052	-0.0340	0.1034	0.1228	0.0038	0.0586
New Zealand	-0.0412	0.0661	-0.0417	0.0393	0.1251	0.0039	0.0247
Nicaragua	0.1537	-0.0736	-0.0860	0.0031	0.1730	0.0075	0.0192

Nigeria	0.0308	-0.0599	0.3658	-0.2471	0.4451	0.0487	0.0583
Norway	-0.0800	0.1408	-0.0220	0.0069	0.2019	0.0101	0.0220
Pakistan	0.1890	-0.0661	0.0181	-0.1356	0.2365	0.0139	0.0325
Panama	0.0362	0.0080	-0.0723	0.0309	0.0947	0.0023	0.0199
Paraguay	-0.0503	0.0319	0.0740	-0.0549	-0.0990	0.0025	0.0249
Peru	-0.0190	0.0110	0.0488	-0.0471	-0.0774	0.0015	0.0182
Philippines	-0.0331	-0.0179	0.0240	0.0201	-0.0648	0.0011	0.0122
Poland	0.0036	-0.0058	-0.0039	0.0035	-0.0085	0.0000	0.0232
Portugal	0.0483	0.0921	-0.0188	-0.1558	-0.1894	0.0090	0.0515
Qatar	0.0951	0.0640	-0.0869	-0.0831	0.1595	0.0064	0.0628
Romania	-0.0135	0.0115	0.0213	-0.0097	0.0289	0.0002	0.0262
Russian Federation	-0.0876	0.0733	0.1251	-0.0621	0.1577	0.0062	0.0489
Rwanda	-0.0793	0.1040	-0.0412	0.0031	-0.1390	0.0048	0.0320
Saudi Arabia	0.2052	0.1652	-0.0934	-0.2863	0.3767	0.0352	0.0639
Senegal	0.0264	-0.0466	0.0029	0.0197	0.0508	0.0007	0.0511
Serbia	-0.0151	0.0079	0.0266	-0.0095	0.0310	0.0002	0.0406
Singapore	0.0096	0.0113	-0.0223	0.0012	0.0290	0.0002	0.0325
Slovak Republic	0.0252	-0.0264	-0.0094	-0.0030	-0.0452	0.0005	0.0204
Slovenia	0.0551	-0.0266	-0.0336	-0.0220	-0.0832	0.0017	0.0262
South Africa	0.0839	-0.0147	-0.0670	-0.0764	-0.2241	0.0123	0.0135
Spain	-0.0126	-0.0149	-0.0082	0.0448	0.0525	0.0007	0.0339
Sri Lanka	-0.0913	-0.2924	0.0045	0.3453	-0.4307	0.0453	0.0455
Sudan	0.0770	-0.0584	0.0222	-0.0369	0.1033	0.0027	0.0399
Suriname	0.0759	-0.0298	-0.0580	0.0212	0.1167	0.0034	0.0115
Switzerland	-0.0709	0.0688	-0.0393	0.0727	0.1587	0.0063	0.0296
Syrian Arab Republic	-0.4743	0.0473	0.2393	0.2141	-0.5349	0.0669	0.0287
Tajikistan	-0.0362	-0.0987	0.0982	0.0095	-0.2087	0.0107	0.0142
Thailand	0.1613	0.0076	0.0738	-0.1974	0.3044	0.0223	0.0160
Trinidad and Tobago	-0.0097	0.0397	0.0124	-0.0268	0.0593	0.0009	0.0153
Tunisia	-0.0355	0.0189	0.0212	-0.0068	-0.0465	0.0005	0.0134
Turkey	-0.0277	0.0184	0.0293	-0.0227	-0.0497	0.0006	0.0151
Turkmenistan	-0.0112	-0.0189	0.0089	0.0175	-0.0344	0.0003	0.0172
Uganda	-0.0309	0.0658	-0.0716	0.0129	-0.1195	0.0036	0.0269
Ukraine	0.1274	-0.1049	-0.2274	0.1274	-0.2660	0.0177	0.0609
United Arab Emirates	0.1256	0.1067	-0.1122	-0.1323	0.2308	0.0133	0.0626
United Kingdom	-0.0344	0.0514	0.0112	-0.0103	0.0672	0.0011	0.0259
United States	-0.0085	0.0062	-0.0054	0.0111	0.0190	0.0001	0.0368
Uruguay	-0.0320	-0.0870	-0.0279	0.1697	0.1848	0.0086	0.0632
Uzbekistan	-0.0157	-0.0435	0.0381	0.0155	-0.0745	0.0014	0.0212
Vietnam	-0.0283	0.0007	0.0113	0.0141	-0.0375	0.0004	0.0140
Yemen. Rep.	0.0098	-0.0053	0.0010	-0.0056	0.0118	0.0000	0.0522
Zimbabwe	-0.0151	0.0292	-0.0648	0.0225	-0.1043	0.0027	0.0154

Результаты моделирования согласно модели 1.1 (151 государство).

Номер	Страна	LQ	GDP	UR	MF	DR
1	Liberia	38.14090924	900000	3.8	45.88	9.69
2	Central African Republic	26.39292558	600000	7.4	37.6	13.8
3	Qatar	10.36298327	145000	0.3	15.63	1.53
4	Luxembourg	9.350642996	102900	6.1	72.6	7.24
5	Singapore	8.570835344	85700	3	46.08	3.43
6	Brunei Darussalam	7.72335456	79700	3.8	18.75	3.52
7	Monaco	7.885692747	78700	2	47.08	9.24
8	Kuwait	7.525330648	72200	3	15.93	2.18
9	Norway	7.818281556	68400	3.4	59.63	8.12
10	United Arab Emirates	7.317058774	67000	3.6	15.38	1.97
11	Australia	7.732693532	65400	6	61.35	7.14
12	San Marino	7.126113709	62100	8.7	48.43	8.45
13	Switzerland	7.716277055	59300	4.5	72.38	8.13
14	United States	7.668145136	56300	6.2	78.2	8.15
15	Saudi Arabia	6.508853544	54600	5.6	7.75	3.33
16	Ireland	7.089425027	54300	11.6	57.33	6.48
17	Bahrain	6.789456706	51200	4.1	21.63	2.69
18	Netherlands	7.632729913	49300	6.9	91.7	8.66
19	Austria	7.127478885	47500	5	71.13	9.42
20	Germany	7.106087823	47400	5	78.03	11.42
21	Iceland	6.96353411	46600	5	52.95	6.28
22	Oman	6.415946726	46200	7.2	20.63	3.36
23	Canada	7.200889704	45900	6.9	76.58	8.42
24	Denmark	6.834599313	45800	6.6	66.33	10.25
25	Belgium	7.035914513	44100	8.5	79.35	9.63
26	France	6.721804139	41400	9.9	69.93	9.16
27	United Kingdom	6.571559871	41200	6.3	58.05	9.35
28	Finland	6.503652256	41200	8.6	60.58	9.83
29	Japan	6.385273328	38200	3.7	50.85	9.51
30	Andorra	6.447999031	37200	4	46.13	6.96
31	Korea. Rep.	6.437964655	36700	3.5	44.88	6.75
32	New Zealand	6.740003022	36400	5.6	65.25	7.36
33	Italy	6.213151487	35800	12.5	64.25	10.19
34	Spain	6.158463856	35200	24.7	78.6	9.04
35	Malta	6.090001332	34700	5.9	45.2	9.09
36	Israel	6.419281198	34300	6.1	46.28	5.15
37	Equatorial Guinea	6.012915606	33300	7.9	43.88	8.19
38	Trinidad and Tobago	6.143848429	32800	4	45.75	8.56
39	Czech Republic	6.617996741	31500	6.2	80.5	10.34
40	Cyprus	5.994692372	31000	15.6	53.73	6.62
41	Slovenia	6.16526716	30900	9.5	70	11.37
42	Slovak Republic	5.947447319	29500	13.3	62.33	9.74
43	Estonia	6.049777275	28700	7.7	69.4	12.4

44	Lithuania	5.377462823	28000	11.3	52.63	14.27
45	Portugal	6.208691086	27800	14.2	83.8	11.02
46	Seychelles	6.158075033	27000	3	48.25	6.89
47	Malaysia	5.834352461	26600	2	25.08	5.03
48	Poland	5.678236434	26400	9.2	50.08	10.19
49	Hungary	5.713553932	26000	7.8	59.88	12.73
50	Bahamas. The	5.597034203	25600	15.4	45.88	7.05
51	Greece	5.203962815	25600	26.3	61.38	11.09
52	Kazakhstan	5.984676856	24700	4.1	51.08	8.21
53	Latvia	5.442233076	24500	10	59.63	14.31
54	Chile	6.232240897	23800	6.4	59.4	6
55	Russian Federation	5.494471532	23700	5.1	52.88	13.69
56	Argentina	6.114064438	22400	8.2	64.45	7.33
57	Uruguay	6.482438127	21800	7	88.75	9.45
58	Croatia	5.250021484	21300	16.7	59.13	12.18
59	Panama	6.134387019	20900	4.3	51.75	4.81
60	Romania	5.44165445	20600	7	52	11.9
61	Turkey	5.833828775	20500	9.2	50.78	5.88
62	Mauritius	5.656443932	19500	7.7	45.75	6.91
63	Azerbaijan	5.637013506	18700	5.2	42.68	7.07
64	Lebanon	5.583234473	18600	6.4	33.9	4.88
65	Mexico	6.499397128	18500	4.9	75.53	5.26
66	Bulgaria	5.01250275	18400	11.6	53.33	14.44
67	Belarus	5.128346908	17800	5.9	46.08	13.36
68	Iran. Islamic Rep.	4.863418394	17800	12.8	16.63	5.94
69	Botswana	4.547218894	17700	18.2	39.88	13.39
70	Suriname	5.685142657	16700	5.6	45.38	6.13
71	Thailand	5.377067424	16100	0.9	30.75	7.8
72	Brazil	6.08713872	15800	6.8	69.3	6.58
73	Turkmenistan	5.285382667	15600	10.5	36.88	6.13
74	Costa Rica	5.841449123	15500	8.3	53.08	4.55
75	Iraq	4.751543996	15500	16.4	13	3.77
76	Libya	4.846187806	15100	19.2	22	3.58
77	Dominican Republic	5.273134622	14900	15	38.9	4.55
78	Algeria	5.07795395	14400	9.5	20.63	4.31
79	China	5.396096676	14300	4.7	40	7.53
80	Colombia	6.12287117	14000	10.1	74.98	5.4
81	Macedonia. FYR	4.728413131	14000	27.9	54.13	9.08
82	Maldives	5.118390489	13600	11.6	25.75	3.89
83	Serbia	4.515291669	13600	22.2	53.13	13.66
84	South Africa	4.900807304	13400	25.1	61.7	9.91
85	Mongolia	5.598374989	12500	4.8	48.08	6.35
86	Jordan	5.318695269	12400	11.1	35.88	3.79
87	Peru	5.877998631	12300	4.2	59.08	6.01
88	St. Lucia	4.906080457	12000	20	46.25	7.42
89	Albania	5.268541623	11900	16.1	53.5	6.58

90	Tunisia	5.172359566	11600	13.3	42.58	5.98
91	Egypt. Arab Rep.	4.718788413	11500	13.2	16.88	4.77
92	Ecuador	6.027997353	11300	4.6	64.75	5.06
93	Indonesia	4.927431145	11300	6.2	21.43	6.37
94	Sri Lanka	5.012905684	11200	4.6	21.9	6.11
95	St. Vincent and the Gredines	4.986020102	11000	18.8	48.75	7.18
96	Cuba	5.524199773	10200	3	50.88	7.72
97	Bosnia and Herzegovina	4.564410773	10200	27.9	55.63	9.75
98	Swaziland	4.136863329	9800	22.3	41.88	13.56
99	Georgia	4.775537532	9500	13.4	46.63	10.82
100	Paraguay	5.741727729	8800	4.5	54.13	4.68
101	Jamaica	5.256097596	8800	13.2	53.88	6.7
102	Armenia	4.788920342	8400	17.1	49.58	9.34
103	El Salvador	5.223888367	8300	6.2	37.7	5.69
104	Morocco	4.924643842	8300	10.2	27.08	4.81
105	Ukraine	4.642016451	8000	7.7	47.58	14.46
106	Guatemala	5.528041613	7900	2.9	43.45	4.77
107	Angola	4.667803668	7600	6.8	36.5	11.49
108	Philippines	5.168664209	7500	7.1	39.63	6.11
109	Guyana	5.17642728	7200	11.1	51.83	7.32
110	Bolivia	5.808915437	6500	2.7	65.3	6.52
111	Nigeria	4.2982903	6400	7.5	28.03	12.9
112	India	5.524310164	6300	3.6	57.03	7.32
113	Vietnam	5.180057	6100	2.3	33.88	5.93
114	Uzbekistan	5.151463549	6100	10.6	43.93	5.3
115	Lao PDR	5.276746176	5400	1.4	44.63	7.63
116	Tonga	5.424149787	5100	1.1	40.75	4.85
117	Syrian Arab Republic	4.770653824	5100	10.8	23.3	4
118	Honduras	5.330026967	5000	3.9	42.45	5.17
119	Nicaragua	5.155328373	5000	5.3	36.33	5.08
120	Moldova	4.955232502	5000	3.4	52.88	12.59
121	Pakistan	4.643670024	4900	5.2	18.05	6.49
122	Sudan	4.259985208	4500	14.8	21.43	7.66
123	Mauritania	3.791305183	4500	31	28.73	8.2
124	Ghana	5.328539957	4300	2.4	49	7.22
125	Bangladesh	4.935077066	3600	4.3	28.9	5.61
126	Cambodia	5.796013391	3500	0.4	70.5	7.68
127	Kyrgyz Republic	5.081719498	3400	8.1	46.38	6.65
128	Kenya	4.82343369	3300	9.2	37.33	6.89
129	Djibouti	2.762320444	3300	60	29.5	7.73
130	Cameroon	4.919826406	3200	4.3	46.25	10.11
131	Papua New Guinea	5.211101646	2800	2.5	43.63	6.53
132	Tajikistan	5.041327922	2800	10.9	48.38	6.18
133	Yemen. Rep.	3.998749619	2800	17.4	11.23	6.28
134	Nepal	5.318633778	2500	2.7	49.58	6.56
135	Senegal	4.856556905	2500	10	47.63	8.46

136	Zimbabwe	4.791215398	2100	5.4	44.08	10.13
137	Uganda	4.754558724	2100	3.8	41.88	10.69
138	Solomon Islands	5.408274518	2000	3.9	46.13	3.85
139	Afghanistan	3.755281316	2000	9.1	16.88	13.89
140	Haiti	4.922423953	1800	6.8	44.13	7.83
141	Rwanda	4.887972706	1800	0.6	36.58	8.96
142	Mali	4.398963639	1800	8.1	41.55	12.89
143	Gambia. The	5.004163648	1700	7	45.8	7.15
144	Ethiopia	4.741303113	1700	5.2	34.63	8.19
145	Comoros	4.602415277	1600	6.5	28.13	7.57
146	Madagascar	5.166441691	1500	3.6	46.7	6.81
147	Guinea-Bissau	4.399026832	1500	6.9	45.63	14.33
148	Guinea	5.060819589	1300	1.8	49.38	9.46
149	Mozambique	4.110002307	1300	22.6	50.08	12.1
150	Malawi	4.928185434	1200	7.5	48.83	8.41
151	Eritrea	4.537989248	1200	7.2	26.8	7.52

Результаты моделирования согласно модели 1.1 (68 государств).

Номер	Страна	LQ
1	Norway	7.818281556
2	Australia	7.732693532
3	Switzerland	7.716277055
4	United States	7.668145136
5	Netherlands	7.632729913
6	Canada	7.200889704
7	Austria	7.127478885
8	Germany	7.106087823
9	Ireland	7.089425027
10	Belgium	7.035914513
11	Iceland	6.96353411
12	Denmark	6.834599313
13	France	6.721804139
14	Czech Republic	6.617996741
15	United Kingdom	6.571559871
16	Finland	6.503652256
17	Mexico	6.499397128
18	Uruguay	6.482438127
19	Israel	6.419281198
20	Japan	6.385273328
21	Chile	6.232240897
22	Italy	6.213151487
23	Slovenia	6.16526716
24	Spain	6.158463856
25	Colombia	6.12287117
26	Brazil	6.08713872
27	Estonia	6.049777275
28	Ecuador	6.027997353
29	Kazakhstan	5.984676856
30	Slovak Republic	5.947447319
31	Peru	5.877998631
32	Costa Rica	5.841449123
33	Turkey	5.833828775
34	Bolivia	5.808915437
35	Hungary	5.713553932
36	Poland	5.678236434
37	Mauritius	5.656443932
38	Azerbaijan	5.637013506
39	Mongolia	5.598374989
40	India	5.524310164
41	Russian Federation	5.494471532
42	Latvia	5.442233076

Номер	Страна	LQ
44	Lithuania	5.377463
45	Thailand	5.377067
46	Honduras	5.330027
47	Jordan	5.318695
48	Dominican Republic	5.273135
49	Croatia	5.250021
50	Greece	5.203963
51	Tunisia	5.17236
52	Philippines	5.168664
53	Belarus	5.128347
54	Sri Lanka	5.012906
55	Bulgaria	5.012503
56	Moldova	4.955233
57	Indonesia	4.927431
58	Morocco	4.924644
59	South Africa	4.900807
60	Iran. Islamic Rep.	4.863418
61	Senegal	4.856557
62	Kenya	4.823434
63	Georgia	4.775538
64	Ethiopia	4.741303
65	Macedonia. FYR	4.728413
66	Egypt. Arab Rep.	4.718788
67	Ukraine	4.642016
68	Serbia	4.515292

Результаты моделирования согласно модели 2.3 (150 государство).

Номер	Страна	QL	MYoS	MF	DR
1	Netherlands	7.098841961	11.89	91.7	8.66
2	United States	7.096822897	12.94	78.2	8.15
3	Canada	7.021189787	13	76.58	8.42
4	Switzerland	6.937778788	12.82	72.38	8.13
5	Australia	6.89066157	12.96	61.35	7.14
6	Luxembourg	6.873327577	11.71	72.6	7.24
7	New Zealand	6.844636023	12.5	65.25	7.36
8	Israel	6.815249422	12.54	46.28	5.15
9	Ireland	6.757478906	12.2	57.33	6.48
10	Singapore	6.725665876	10.63	46.08	3.43
11	Mexico	6.624600715	8.47	75.53	5.26
12	Czech Republic	6.613782173	12.32	80.5	10.34
13	Norway	6.598498386	12.63	59.63	8.12
14	Cyprus	6.523883157	11.62	53.73	6.62
15	Germany	6.513752092	13.07	78.03	11.42
16	Belgium	6.496292964	11.27	79.35	9.63
17	United Kingdom	6.421868497	13.05	58.05	9.35
18	Chile	6.385737156	9.79	59.4	6
19	Denmark	6.382703338	12.73	66.33	10.25
20	Uzbekistan	6.376689637	10.86	43.93	5.3
21	Tonga	6.353377419	10.71	40.75	4.85
22	Iceland	6.350726029	10.59	52.95	6.28
23	Colombia	6.348805797	7.35	74.98	5.4
24	Korea. Rep.	6.347966779	11.89	44.88	6.75
25	Panama	6.332849305	9.35	51.75	4.81
26	France	6.331513603	11.13	69.93	9.16
27	Slovak Republic	6.27715107	12.23	62.33	9.74
28	Argentina	6.266697899	9.83	64.45	7.33
29	Jordan	6.265867861	9.91	35.88	3.79
30	Austria	6.248171417	10.83	71.13	9.42
31	Spain	6.230647403	9.58	78.6	9.04
32	Cuba	6.229311235	11.51	50.88	7.72
33	Ecuador	6.221151087	7.59	64.75	5.06
34	Tajikistan	6.212765098	10.36	48.38	6.18
35	Peru	6.2113839	9.01	59.08	6.01
36	Costa Rica	6.205231788	8.37	53.08	4.55
37	Uruguay	6.155223114	8.45	88.75	9.45
38	Kyrgyz Republic	6.127014058	10.59	46.38	6.65
39	Kazakhstan	6.124110423	11.42	51.08	8.21
40	Bahamas. The	6.115030025	10.94	45.88	7.05
41	Jamaica	6.100222328	9.67	53.88	6.7
42	Slovenia	6.083950919	11.89	70	11.37
43	Azerbaijan	6.082309027	11.16	42.68	7.07

44	Bolivia	6.081741664	8.15	65.3	6.52
45	Paraguay	6.064219385	7.7	54.13	4.68
46	Brazil	6.061410754	7.66	69.3	6.58
47	United Arab Emirates	6.030750574	9.49	15.38	1.97
48	Bahrain	6.029923419	9.42	21.63	2.69
49	Qatar	6.029450307	9.07	15.63	1.53
50	Albania	6.026809488	9.26	53.5	6.58
51	Estonia	6.003500696	12.48	69.4	12.4
52	Mongolia	5.945765333	9.28	48.08	6.35
53	Japan	5.898692555	11.52	50.85	9.51
54	Seychelles	5.877104875	9.41	48.25	6.89
55	Andorra	5.850028812	9.58	46.13	6.96
56	Turkmenistan	5.849084097	9.88	36.88	6.13
57	Trinidad and Tobago	5.819185057	10.88	45.75	8.56
58	Poland	5.817825027	11.82	50.08	10.19
59	Finland	5.808901626	10.29	60.58	9.83
60	Malaysia	5.791507185	9.96	25.08	5.03
61	Italy	5.788574937	10.1	64.25	10.19
62	Armenia	5.762793934	10.87	49.58	9.34
63	South Africa	5.746485704	9.94	61.7	9.91
64	Turkey	5.73295232	7.56	50.78	5.88
65	St. Lucia	5.716784132	9.34	46.25	7.42
66	Philippines	5.706238153	8.88	39.63	6.11
67	Portugal	5.704869987	8.25	83.8	11.02
68	Dominican Republic	5.699232083	7.56	38.9	4.55
69	Guyana	5.69575672	8.53	51.83	7.32
70	Sri Lanka	5.693750074	10.8	21.9	6.11
71	Georgia	5.680857414	12.11	46.63	10.82
72	Brunei Darussalam	5.670532724	8.77	18.75	3.52
73	St. Vincent and the Gredines	5.668055661	8.62	48.75	7.18
74	Mauritius	5.630406978	8.54	45.75	6.91
75	Lebanon	5.596127053	7.92	33.9	4.88
76	Malta	5.591586142	10.33	45.2	9.09
77	Greece	5.587725047	10.26	61.38	11.09
78	Oman	5.581757	8	20.63	3.36
79	Suriname	5.578148307	7.65	45.38	6.13
80	Macedonia. FYR	5.577957891	9.26	54.13	9.08
81	Hungary	5.539942211	11.64	59.88	12.73
82	Kuwait	5.522633548	7.21	15.93	2.18
83	Croatia	5.495262408	11.03	59.13	12.18
84	Solomon Islands	5.466835258	5.04	46.13	3.85
85	Saudi Arabia	5.431324221	8.7	7.75	3.33
86	Libya	5.427374649	7.31	22	3.58
87	Tunisia	5.368590279	6.84	42.58	5.98
88	Guatemala	5.351347583	5.6	43.45	4.77
89	Romania	5.326032808	10.78	52	11.9

90	Ghana	5.323959234	7	49	7.22
91	Algeria	5.323070019	7.61	20.63	4.31
92	Moldova	5.305519155	11.19	52.88	12.59
93	Vietnam	5.301501657	7.45	33.88	5.93
94	Bosnia and Herzegovina	5.292421794	8.33	55.63	9.75
95	Russian Federation	5.262196899	11.95	52.88	13.69
96	Lithuania	5.239643391	12.38	52.63	14.27
97	Dominica	5.239276275	7.86	43.13	7.91
98	El Salvador	5.237408089	6.51	37.7	5.69
99	Honduras	5.228171432	5.48	42.45	5.17
100	Nicaragua	5.21036967	6	36.33	5.08
101	Latvia	5.207133697	11.48	59.63	14.31
102	Cambodia	5.18996887	4.37	70.5	7.68
103	Belarus	5.169242214	11.98	46.08	13.36
104	China	5.16821051	7.54	40	7.53
105	Syrian Arab Republic	5.15599908	6.25	23.3	4
106	India	5.154490559	5.39	57.03	7.32
107	Maldives	5.147531868	5.84	25.75	3.89
108	Madagascar	5.144825041	6.05	46.7	6.81
109	Iran. Islamic Rep.	5.044727349	8.17	16.63	5.94
110	Iraq	4.983026808	6.38	13	3.77
111	Serbia	4.958575524	10.46	53.13	13.66
112	Indonesia	4.955616956	7.59	21.43	6.37
113	Kenya	4.955335416	6.27	37.33	6.89
114	Egypt. Arab Rep.	4.925072556	6.55	16.88	4.77
115	Ukraine	4.865271553	11.34	47.58	14.46
116	Thailand	4.853264645	7.32	30.75	7.8
117	Bulgaria	4.841876493	10.57	53.33	14.44
118	Bangladesh	4.740014529	5.07	28.9	5.61
119	Lao PDR	4.721821664	5	44.63	7.63
120	Zimbabwe	4.72091959	7.25	44.08	10.13
121	Equatorial Guinea	4.701749118	5.48	43.88	8.19
122	Morocco	4.697418122	4.37	27.08	4.81
123	Papua New Guinea	4.68021381	3.95	43.63	6.53
124	Nepal	4.679756011	3.31	49.58	6.56
125	Haiti	4.647545516	4.88	44.13	7.83
126	Malawi	4.526136658	4.29	48.83	8.41
127	Cameroon	4.503057395	5.96	46.25	10.11
128	Gambia. The	4.377473546	2.82	45.8	7.15
129	Botswana	4.359665	8.87	39.88	13.39
130	Comoros	4.258890868	4.6	28.13	7.57
131	Pakistan	4.24877718	4.73	18.05	6.49
132	Uganda	4.178129815	5.42	41.88	10.69
133	Liberia	4.176784173	4.09	45.88	9.69
134	Senegal	4.10578003	2.48	47.63	8.46
135	Djibouti	4.100835102	3.84	29.5	7.73

136	Eritrea	4.080285198	3.86	26.8	7.52
137	Rwanda	4.018656216	3.74	36.58	8.96
138	Swaziland	4.005271924	7.12	41.88	13.56
139	Mauritania	3.980677137	3.77	28.73	8.2
140	Guinea	3.948910744	2.42	49.38	9.46
141	Ethiopia	3.834201797	2.41	34.63	8.19
142	Sudan	3.775304116	3.14	21.43	7.66
143	Angola	3.756834016	4.73	36.5	11.49
144	Yemen. Rep.	3.676321099	2.6	11.23	6.28
145	Mozambique	3.651261348	3.25	50.08	12.1
146	Nigeria	3.55335595	5.95	28.03	12.9
147	Central African Republic	3.246470778	4.22	37.6	13.8
148	Mali	3.047503206	2.04	41.55	12.89
149	Guinea-Bissau	3.043764788	2.83	45.63	14.33
150	Afghanistan	2.531339922	3.23	16.88	13.89

Коды функций на языке R

Функция `auto.regression()` возвращает объект типа список, который содержит матрицу `Radj`, отражающую порядок включения факторов в модель на основе прироста скорректированного коэффициента детерминации, а так же соответствующий матрице набор данных - `Data`, по которому следует строить линейную многофакторную регрессию.

Входные аргументы:

`@initial_data` (тип `dataframe`), матрица исходных данных, которая содержит в первом столбце имена наблюдений, во втором – данные по регрессору, в оставшейся части набор рассматриваемых факторов. В случае, если в объекте присутствуют пропуски, функция возвращает эвристическое решение согласно описанному в пункте 2.2 подходу;

`@accuracy` – пороговое значение прироста скорректированного коэффициента детерминации (по умолчанию 0.05);

`@correlation_bound` – минимальное значение определителя матрицы $X^T X$ с целью избегания случая появления мультиколлинеарности по ходу построения модели (по умолчанию 0.25);

`@output_file_name` – имя выходного файла, в который будут записаны результаты.

Код функции:

```
auto.regression <- function(initial_data, accuracy = 0.05, correlation_bound = 0.25,
output_file_name) {
  num_of_countries <- dim(initial_data)[1]
  num_of_indexes <- dim(initial_data)[2] - 2

  idx_names <- colnames(initial_data[3:(num_of_indexes + 2)])
  C <- as.matrix(initial_data[,1])
  Y <- as.matrix(initial_data[,2])
  X <- as.matrix(initial_data[,3:(num_of_indexes + 2)])

  Radj <- matrix(data = NA, nrow = num_of_indexes, ncol = num_of_indexes, byrow = FALSE,
    dimnames = list(idx_names, c(rep("stopped", num_of_indexes))))
```

```

X_check = X;
previous_indexes = NULL
const_name_1 = "include:"
const_name_0 = "skipped:"
maximum_location = 0
zero = rep(0, num_of_countries)
previous_maximum = 0
previous_names = ""

for (i in 1:num_of_indexes) {
  for (j in 1:num_of_indexes) {
    if (identical(as.numeric(X_check[,j]), zero, num.eq = TRUE)) { Radj[j, i] = 0; next }

    merger <- cbind(previous_indexes, X[, j])
    tmp <- na.omit(cbind(Y, merger))
    lin_model <- lm(tmp[,1] ~ tmp[, 2:dim(tmp)[2]])
    Radj[j,i] = summary(lin_model)$adj.r.squared
  }

  maximum = max(Radj[,i])
  if (maximum - previous_maximum < accuracy) break
  maximum_location = which(Radj[,i] == maximum)

  tmp <- na.omit(cbind(previous_indexes, X[, maximum_location]))
  tmp.1 <- cbind(rep(1, nrow(tmp)), as.matrix(tmp))

  if (det(t(tmp.1)%*%tmp.1) < det_bound) {
    maximum_name <- paste(const_name_0, row.names(Radj)[maximum_location], sep = " ");
    colnames(Radj)[i + 1] <- maximum_name;
    X_check[, maximum_location] = zero;
    break
  }
}

```

```

maximum_name <- paste(const_name_1, row.names(Radj)[maximum_location], sep = " ")
colnames(Radj)[i] <- maximum_name

previous_indexes <- cbind(previous_indexes, X[,maximum_location])
previous_names <- cbind(previous_names, row.names(Radj)[maximum_location])
X_check[,maximum_location] = zero
previous_maximum = maximum
}

final_data <- cbind(Y, previous_indexes)
row.names(final_data) <- C
colnames(final_data) <- previous_names

final_data <- as.data.frame(na.omit(final_data))

Radj <- t(na.omit(t(Radj)))
Radj <- format(round(Radj, 4), nsmall = 4)
Radj <- trimws(Radj)

for (i in 1:num_of_indexes)
  for (j in 1:dim(Radj)[2])
    if (Radj[i,j] == "0.0000") Radj[i,j] = "used"

write.table(Radj, output_file_name, append = FALSE, row.names = TRUE, col.names = TRUE)
write("\n\n", output_file_name, append = TRUE)
write.table(final_data, output_file_name, append = TRUE, row.names = TRUE, col.names = TRUE)
write("\n\n", output_file_name, append = TRUE)
write.table(cor(final_data), output_file_name, append = TRUE, row.names = TRUE, col.names = TRUE)

lst <- list(Radj, final_data)
names(lst) <- c("Radj", "Data")
return(lst)
}

```

Функция `add.index()` возвращает матрицу, содержащую максимально доступный набор факторов для фиксированного набора наблюдений.

Входные аргументы:

`@initial_data` – (тип `dataframe`) набор исходных данных, содержащий все наблюдения и все рассматриваемые факторы. Структура матрицы аналогична структуре аргумента «`@initial_data`» функции `auto.regression()`;

`@radj_data` – матрица формируемая функцией `auto.regression()`;

`@output_file_name` – имя выходного файла, в который будут записаны результаты.

Код функции:

```
add.index <- function(initial_data, radj_data, output_file_name) {
  row.numbers <- nrow(radj_data)
  country.position <- rep(0, row.numbers)
  result_data <- radj_data
  new.added <- 0

  for (i in 1:row.numbers) {
    country.position[i] = which(initial_data[,1] == row.names(radj_data[i,]))
  }

  for (i in 2:ncol(initial_data)) {
    flag <- 0
    if (colnames(initial_data)[i] %in% colnames(radj_data)) next

    for (j in country.position) {
      if (is.na(initial_data[j,i]) == TRUE) break
      else { flag <- flag + 1 }
    }
    if (flag == row.numbers) {
      result_data <- cbind(result_data, initial_data[country.position,i])
      new.added <- new.added + 1
    }
  }
}
```

```

last.elem <- ncol(result_data)
colnames(result_data)[last.elem] <- colnames(initial_data)[i]
}
}
result_data <- cbind(rownames(radj_data),result_data)
colnames(result_data)[1] <- "Country"
rownames(result_data) <- country.position
write.table(result_data, output_file_name, append = FALSE, row.names = TRUE, col.names =
TRUE)
lst <- list(result_data, new.added)
names(lst) <- c("data", "new.indexes")

return(lst)
}

```

Функция `white.test()` является модификацией функции `bptest()` из пакета «`lmtest`», которая проводит тест Уайта основываясь на входной матрицы данных. Функция возвращает значение статистики и величину `p-value`.

Входные аргументы:

`@database` – (тип `dataframe`) матрица содержащая данные, по которым должна строиться линейная регрессионная модель. Первый столбец матрицы соответствует регрессору.

Код функции:

```
white.test <- function(database) {
  name <- names(database)
  size <- ncol(database)

  str.part1 <- paste(name[2:size], collapse = " + ")

  tmp.start <- rep("I(", (size - 1))
  tmp.end <- rep("^2)", (size - 1))
  str.part3 <- paste(tmp.start, name[2:size], tmp.end, sep = "", collapse = " + ")

  if (size > 2) {
    comb <- combn(name[2:size], 2)
    str.part2 <- paste(comb[1, ], comb[2, ], sep = "*", collapse = " + ")
    frmla2 <- as.formula(paste("~ ", str.part2, " + ", str.part3))
  } else {
    str.part2 <- "";
    frmla2 <- as.formula(paste("~ ", str.part3))
  }

  frmla1 <- as.formula(paste(name[1], " ~ ", str.part1))

  white.lm <- lm(frmla1, data = database)
  bp <- bptest(white.lm, frmla2, data = database)

  return(bp)
}
```

Функция `gq.test()` является модификацией функции `gqtest()` из пакета «`lmtest`», которая реализует применение теста Гольфельда-Квандта для каждого фактора автоматически на основе матрицы входных данных. Функция возвращает список, содержащий результаты теста для каждого фактора.

Входные аргументы:

`@data` – (тип `dataframe`) матрица данных, по которым должна строиться линейная регрессионная модель. Первый столбец матрицы соответствует регрессору.

Код функции:

```
gq.test <- function(data) {
  if (is.data.frame(data)) {
    size <- ncol(data)

    if (size >= 2) {
      ordered.data <- data[order(data[, 2]),]
      ordered.lm <- lm(as.matrix(ordered.data[,1]) ~ as.matrix(ordered.data[,2:size]))
      res <- gqtest(ordered.lm)
      output <- list(res)

      if (size >= 3) {
        for (i in 3:size) {
          ordered.data <- data[order(data[, i]),]
          ordered.lm <- lm(as.matrix(ordered.data[,1]) ~ as.matrix(ordered.data[,2:size]))
          res <- gqtest(ordered.lm)
          output <- c(output, list(res))
        }
        names(output) <- names(data)[2:size]
        return(output)
      }
    } else {
      stop("ERROR! DATA should contain at least one factor!")
    }
  } else {
    stop("ERROR! DATA should be data.frame type!")
  }
}
```