

IV Очередной Всероссийский социологический конгресс

Социология и общество:  
глобальные вызовы и региональное развитие

# 4

Секция 4

**Математическое  
моделирование  
и анализ данных  
в социологии**

Абдуллин А. Р., Уфа

## Статистическое имитационное моделирование случайных процессов: методика и конкретная реализация

---

### Аннотация

В статье описана методика построения имитационной модели для стохастических, социально-экономических процессов и рассмотрен начальный вариант ее компьютерной реализации на примере моделирования числа случайных алкогольных отравлений. Модель построена на основе статистических данных по Республике Башкортостан за период 1996-2010 гг..

**Ключевые слова:** случайный процесс, прогнозирование, имитационное моделирование, численные методы, математическое моделирование, метод Монте-Карло, генератор случайных чисел, тренд, ряды Фурье, остатки, условия Маркова-Гаусса

### Постановка исследовательской задачи

Многие социально-экономические процессы являются *случайными* или могут быть рассмотрены как таковые. Например, финансовые рынки, образование очередей, распределение продаж, количества клиентов по часам, дням недели, месяцам и т. п. Корреляционно-регрессионный анализ, проведенный по статданным Республики Башкортостан [1], показал, что, во-первых, на такой социальный показатель как *ожидаемая продолжительность жизни*, помимо экономических факторов, влияют также несчастные случаи и болезни; во-вторых часть из них являются *случайными величинами*, характеризуемыми своими законами распределения. Так, числа умерших от алкогольных отравлений, утоплений и болезней новообразований имеют нормальное распределение, а от несчастных случаев на транспорте и болезней органов дыхания – логарифмически-нормальное [2].

Можно ли каким-то образом *спрогнозировать* значения подобных случайных величин? Можно. Современная наука предлагает два подхода: 1) математический и 2) эконометрический. Чисто математический подход

основан на так называемой *теории случайных функций* [3], [11] и чаще всего используется в физических и инженерных расчетах. Например, задача определения среднего времени, в течение которого электростанция не в состоянии обеспечить заявки потребителей тока вследствие случайных колебаний потребной мощности, или определения запаса прочности детали, работающей под воздействием случайной нагрузки и т. п. [11, с. 65] Решение таких задач чаще всего предполагает, что имеется *корреляционная функция* выраженная в аналитическом (!) виде, а не просто дан ее график (коррелограмма). Во втором случае, «В эконометрии принято моделировать временной ряд как случайных процесс, называемый также стохастическим процессом, под которым понимается статистическое явление, развивающееся во времени согласно законам теории вероятностей» [13, с. 348]. Эконометрический метод основан на построении разностных уравнений, известных как *модель Бокса-Дженкинсона* [4], и реализована в так называемой модели ARMA; в переводе эта аббревиатура означает *авторегрессионный процесс скользящего среднего*. «Характерной особенностью стационарных процессов типа ARMA ( $p, q$ ) является то, что корни  $\lambda_i$  характеристического уравнения  $\phi(L)=0$  находятся вне единичного круга. Если один или несколько корней лежат на единичной окружности или внутри нее, то процесс нестационарен. Теоретически можно предположить много различных типов нестационарных моделей ARMA ( $p, q$ ), однако, как показывает практика, наиболее распространенным типом нестационарных стохастических процессов являются **интегрированные процессы** или, как их еще называют, **процессы с единичным корнем**. Единичным называют корень характеристического уравнения, равный действительной единице:  $\lambda_i = 1$ » [13, с. 463]. Эконометрические модели предполагают обязательное построение *коррелограммы*, т. е. графика автокорреляционной функции.

Оба подхода, указанные выше, требуют незаурядного знания математического аппарата. В первом случае мы можем столкнуться со сложнейшими уравнениями математической физики, например с широко известными *первым и вторым уравнением Колмогорова* (уравнения *Фоккера–Планка–Колмогорова*). Однако, в этом случае, даже для профессиональных математиков «решение уравнений Колмогорова может представлять сравнительную сложную задачу» [11, с. 260]. Относительно эконометрии, отмечается следующее: «Традиционные модели временных рядов, такие как модель ARMA, не могут адекватно учесть все характеристики, которыми обладают финансовые временные ряды, и требуют расширения. Одна из характерных особенностей финансовых рынков состоит в том, что присущая рынку неопределенность изменяется во времени» [13, с. 523]. Так прогноз индекса акций «Moscow Times» с помощью модели ARIMA (были взяты ежедневные котировки с 1997 по 2003г. т. е. 1395 значений) показывает, что «Сами прогнозируемые значения остаются постоянными после первого шага, в то время как ширина доверительного интервала прогноза увеличивается. Как и ожидалось, эта модель не является эффективным инструментом для построения долгосрочных прогнозов российского финансового рынка» [8, с. 421]. Иными словами, имея данные за 6 лет, можно сделать прогноз на 1 шаг, что означает всего лишь 1 день.

Фактически же наличие нестационарности вызывает *непреодолимые* математические сложности в обоих случаях. Возникает вопрос: нельзя ли как то обойти эти сложности? Существуют ли методы, которые можно применить для решения задач, которые не решаются обычными аналитическими (математическими) методами? Такие методы существуют, и в первую очередь это *численные* (вычислительные) методы, т. е. основанные на вычислительной технике. При использовании этих методов, сложность решаемой задачи преодолевается за счет большого количества однотипных вычислений; причем, чем больше вычислений, тем точнее результат (решение). Применительно к нашей задаче – прогнозирования социально-экономических процессов – такой метод получил название *статистического имитационного моделирования*; как будет показано ниже, этот метод основан на использовании вычислительной мощи ЭВМ, за счет чего как раз и преодолеваются «непреодолимые математические сложности». Поскольку *имитационное* моделирование с одной стороны это сочетание вычислительной техники, теории вероятности и статистики, с другой – для получения информации (результатов) предполагает «прогон» построенной имитационной модели с последующей статистической оценкой полученных на ней результатов, то такое моделирование также называют *статистическим имитационным*.

### Основная часть

Имитационное моделирование это не панацея, а лишь крайнее средство. Там где задачу можно решить аналитически, например, используя указанные выше два подхода, то лучше так и поступить. Основное назначение имитации – это экспериментирование с моделью реальной системы. Для чего это нужно? Одной из классических работ в области имитационного моделирования является работа американского специалиста *Роберта Шеннона* «Имитационное моделирование систем – искусство и наука», где утверждается, что «все имитационные модели представляют собой модели типа так называемого черного ящика» [14, с. 22]. Можно согласиться с позицией *Е.К. Масловского* (редактора перевода указанной работы), что «среди методов прикладного системного анализа имитационное моделирование является, пожалуй, самым мощным инструментом исследования сложных систем, управление которыми связано с принятием решений в условиях неопределённости» [там же, с. 5]. Эти две цитаты приведены для того, чтобы еще раз напомнить, что прогнозируя социально-экономические процессы мы имеем дело с «черным ящиком» и должны действовать в условиях неопределенности. Иначе говоря, т. е. переходя на язык эконометрии: если мы имеем процессы не «с единичным корнем», то мы вынуждены обратиться к крайнему средству – имитационному моделированию.

Для того чтобы симитировать случайный процесс надо применить *метод Монте-Карло*, который как пишет *Р. Шеннон*, является «основным принципом моделирования систем, содержащих стохастические или вероятностные элементы» [14, с. 87] и который, в силу своей успешности, для многих специалистов стал даже «синонимом термина “имитационное моделирование”» [там же]. Суть этого метода состоит в том, что данные для опыта вырабатываются искусственно при помощи генератора случайных чисел в сочетании с предварительно выявленной функцией распределения вероятностей для исследуемого процесса. Своего рода классическим примером применения метода Монте-Карло является «Задача о пьяном прохожем»: Пьяный, стоя на углу улицы, решает прогуляться, чтобы разогнать хмель. Пусть вероятности того, что достигнув очередного перекрёстка, он пойдет на север, юг, восток или запад, одинаковы. Какова вероятность того, что, пройдя 10 кварталов, пьяный окажется не далее двух кварталов от места, где он начал прогулку [там же, с. 91]?

Итак, рассмотрим какой-нибудь конкретный социальный *стохастический* процесс, например, число умерших от *несчастных случаев*, в частности от отравлений алкоголем по Республике Башкортостан за период 1996-2010 гг. Используя официальные статистические данные, построим график этого процесса (см. рис. 1).



Рис. 1. Число умерших от случайных отравлений алкоголем в РБ за период 1996-2010 гг.

На этом рисунке представлены две кривые: сам процесс и его тренд – прямая линия (на рисунке указано также и уравнение этой прямой). Квадрат коэффициента детерминации  $R^2=0,301$  говорит о том, что прямая

линия лишь на 30% описывает этот процесс, и это уже первый сигнал, указывающий на то, что мы имеем дело со случайным процессом. (Более важным и по сути, основным является показание автокорреляционной функции, анализ которой здесь опущен для простоты изложения). Ставится *задача*: сделать прогноз такого рода случайного явления на последующие годы.

Если строгий аналитический подход не дает удовлетворительного результата, то мы вынуждены перейти к имитационной модели. При этом будем считать (это наше допущение) что, любой нестационарный случайный процесс  $y(t)$ , можно представить в виде  $y(t) = A(t) x(t)$ , где  $x(t)$  – нелинейная функция стационарного случайного процесса,  $A(t)$  – детерминированный множитель. «Иными словами, такой процесс относится к нестационарным случайным процессам, выборочные функции которых обладают общим детерминированным трендом» [3, с. 25]. Такое допущение вполне оправдано для нашего случая, т. к. моделирование случайного процесса будет осуществляться методом Монте-Карло.

Как исходя из этих положений можно построить имитационную модель? В данной работе будут рассмотрены основные этапы первой части построения модели – математическая модель, и только отчасти второй (вторая часть – написание компьютерной программы; третья – проведение статистических испытаний). Забегая вперед можно отметить, что первые три этапа (первой части) не представляют собой особой сложности и широко используются в задачах прогнозирования. Поскольку мы исходим из крайнего случая, а именно полагаем, что рассматриваемый процесс имеет сложный нелинейный характер, то для его моделирования, получения аналитической функции, воспользуемся *анализом Фурье* [10, с. 193]. Итак, построение математической модели осуществим в пять этапов. Поскольку целью исследования является построение *имитационной модели*, то наш текст будем сопровождать соответствующими иллюстрациями этой модели.

### I Этап. Удаление тренда

По имеющимся статданным, используя метод наименьших квадратов (МНК), оценим и удалим тренд. Иными словами, на этом этапе мы получаем уравнение прямой  $x = a + bt$  (уравнение тренда, о чем уже говорилось выше). По этому уравнению находим значения тренда, т. е. теоретические (вычисленные) значения и вычитаем их из эмпирических данных (наших статданных). Полученная при этом разница называется *остатками* (по сути это *ошибки* расчета). Полученному уравнению на графике соответствует *прямая* линия, т. е. это тренд рассматриваемого процесса (см. рис. 2).

На этом рисунке, как и на последующих, в левой части указаны все необходимые параметры моделирования, в частности здесь указаны годы (с 1996 по 2014 гг.) и коэффициенты (**a**,**b**) уравнения прямой линии.



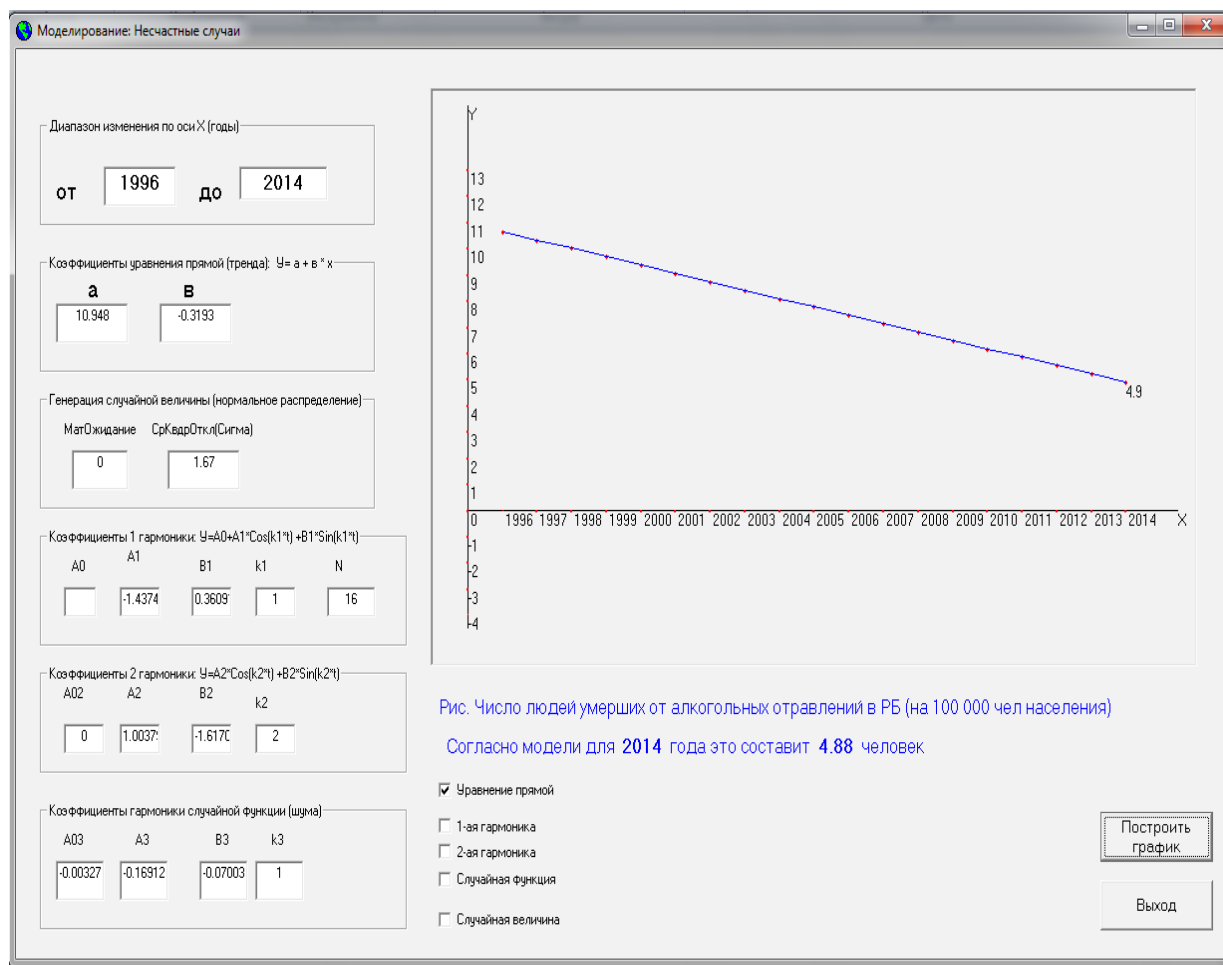


Рис. 2. Уравнение тренда

## II Этап. Анализ Фурье

Для полученных остатков, используя ряды Фурье, подберём соответствующую кривую, которая будет иметь следующий вид: ; где  $a_0$  и  $b_0$  – коэффициенты Фурье;  $x$  – время ( $t = 1, 2, 3 \dots$ );  $n$  – номер гармоники;  $L$  – число испытаний (наблюдений) (кратно  $2^k$ ). Число испытаний – это количество имеющихся статданных; оно должно быть равным только 1, 2, 4, 8, 16, 32 и т. д. Это связано с тем, что

$$f(x) = a_0 + \sum_{n=1}^{\infty} \left( a_n \cos \frac{n\pi x}{L} + b_n \sin \frac{n\pi x}{L} \right)$$

такой подход упрощает расчет коэффициентов Фурье. Расчет этих коэффициентов можно выполнить в *Excel*. Количество гармоник исследователя определяет индивидуально. Считается, что для анализа социально-экономических процессов, можно ограничиться первыми четырьмя гармониками. На графике полученному уравнению (полигармоническая функции) соответствует кривая, имеющая *нелинейный* вид (см. рис. 3).

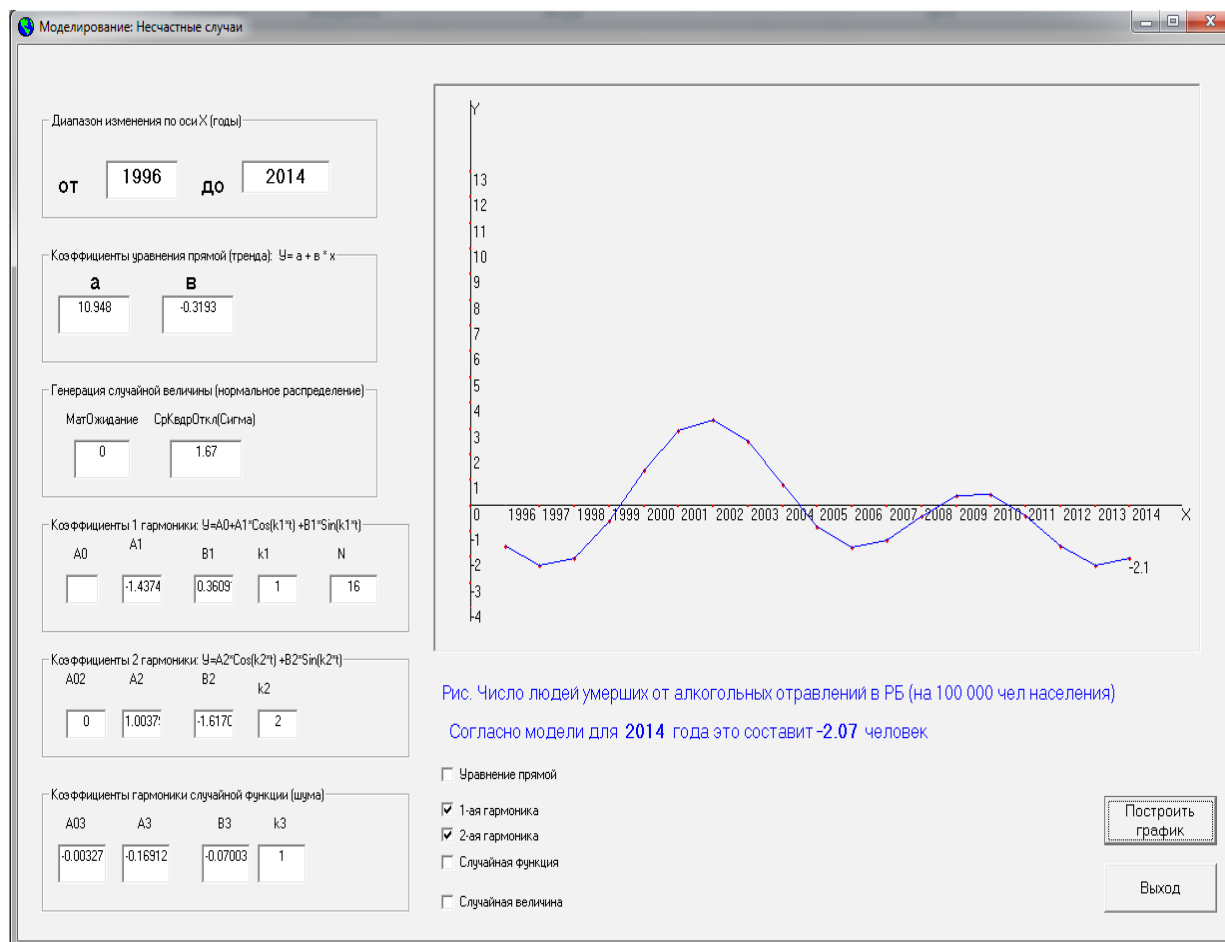


Рис. 3. Сумма двух первых гармоник (анализ рядов Фурье)

### III Этап. Анализ остатков

Еще раз повторим операции первого этапа. По вновь полученному уравнению вычислим теоретические значения и вычтем их из остатков, на основе которых было получено это уравнение (Фурье). Таким образом, мы второй раз получим *остатки*. Напомним, что остатки — это разница между фактическими статданными и теми расчетными значениями, что предсказывает наша теория; если же остатки включают в уравнение регрессии, то тогда их называют *случайным членом*. При традиционном аналитическом подходе на этом построение модели заканчивают и переходят к определению ошибки прогноза (среднеквадратичному отклонению):

$$S = \sqrt{\frac{\sum_{i=1}^L [(y)_i^* - y_i]^2}{k}}$$

где  $y^*$  — расчетные (теоретические) значения;  $y$  — фактические значения;  $k$  — число степеней свободы;  $L$  — число наблюдений (испытаний) [10, с. 196].



Но можно ли на этом останавливаться, будет ли это компетентным анализом? Вот что по этому поводу пишет автор одного из самых популярных на Западе учебников эконометрики, *К. Дугерти*: «В самом деле, для того чтобы регрессионный анализ, основанный на обычном методе наименьших квадратов, давал наилучшие из всех возможных результаты, случайный член должен удовлетворять четырем условиям, известным как условия Гаусса-Маркова. Не будет преувеличением сказать, что именно понимание важности этих условий отличает компетентного исследователя, использующего регрессионный анализ, от некомпетентного. Если эти условия не выполнены, исследователь должен это сознавать. Если корректирующие действия возможны, то аналитик должен быть в состоянии их выполнить. Если ситуацию исправить невозможно, исследователь должен быть способен оценить, насколько серьезно это может повлиять на результаты» [6, с. 79–80]. Условия Гаусса-Маркова сводятся к следующим: 1) математическое ожидание (среднее значение) случайного члена (остатков) должно быть равным нулю; 2) дисперсия случайного члена должна быть постоянной; 3) отсутствие систематической связи между значениями случайного члена в любых двух наблюдениях; 4) объясняющие переменные не являются стохастическими. При этом предполагается, что остатки подчиняются нормальному распределению. Будут ли полученные в результате анализа остатки подчиняться этим условиям? Ответ на этот вопрос является решающим для проверки *адекватности* полученной модели. «Таким образом, – пишут крупнейшие американские специалисты по регрессионному анализу, – если подбираемая нами модель правильная, то остатки будут проявлять тенденцию к подтверждению сделанных предположений или по меньшей мере не будут противоречить им. Именно эта идея лежит в основе исследования остатков; мы должны сформулировать вопрос: “Не показывают ли остатки, что наши предположения ошибочны?”» [7, с. 187]. Не случайно, в этой работе анализу остатков посвящена целая глава (Глава 3. *Исследование остатков*). Все дело в том, что чаще всего, остатки не «проявляют тенденцию к подтверждению ...» и «ситуацию исправить невозможно». Это говорит о том, что мы не можем получить *удовлетворительную аналитическую зависимость* для исследуемого случайного процесса. Но что нам нужно в конечном счете: аналитическая зависимость или же все-таки прогноз? Прогноз.

### IV Этап. Закон и параметры распределения остатков

Чтобы получить прогноз исследуемого процесса, нужно поступить иначе. Если остатки не подчиняются условиям (Гаусса-Маркова) и нет возможности «заставить» их подчиниться, т. е. невозможно найти такую формулу, которая адекватно представляла бы исследуемый процесс, то тогда эти остатки надо *смоделировать*. Да, в этом случае, мы не сможем получить аналитическую зависимость, но мы сможем получить прогнозное значение, а это то, что нам как раз и нужно. В этом состоит суть метода

Монте-Карло. Хотя этот метод не позволяет получить формулу, но позволяет выявить *наиболее вероятное значение* для исследуемого случайного процесса. Поскольку остатки являются случайными величинами, то для их характеристики нужно сначала выявить: а) по какому закону они распределены; б) параметры этого распределения. Если имеются соответствующие данные, то сделать это не сложно. Например, для нормального распределения (что имеет место в подавляющем большинстве случаев) надо определить среднее значение (математическое ожидание) и среднеквадратичное отклонение ( $\sigma$ ).

## V Этап. Построение итоговой модели

Рассматривая модель как аддитивную, сложим значения полученные на I и II этапах. Результат такого сложения представлен на рис. 4.

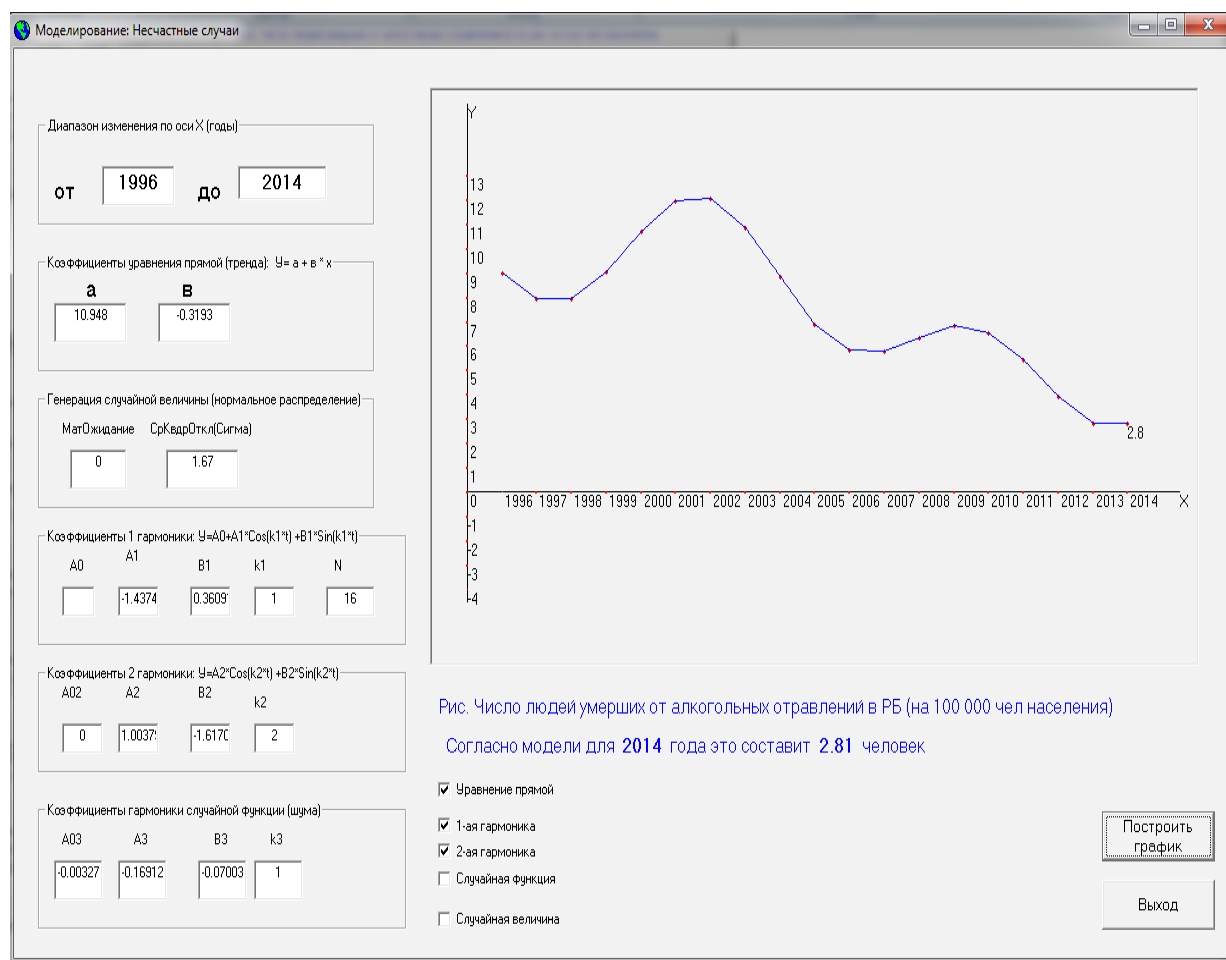


Рис. 4. Уравнение тренда и две гармоники

Если статданных было использовано много, то можно ограничиться последними (10 – 20 вполне достаточно); затем нужно добавить к ним столько значений, на сколько шагов мы хотим сделать прогноз. Поясним это на нашем примере. У нас имеются статданные за период 1996-2010 гг., т. е. за 15 лет, и нам нужно дать прогноз на последующие 4 года, т. е. до

2014 г. По формулам, полученным на I и II этапах надо вычислить значения за период 1996 –2014 гг. Затем по параметрам, полученным на IV этапе, в каждой из этих 19 точек нужно сгенерировать случайные величины. Таким образом, для исследуемого процесса мы будем иметь: а) 15 значений фактических статданных; б) 15 + 4 значений, полученных как теоретически, так и сгенерированных случайным образом. В результате мы получим кривую сгенерированную случайным образом, один из вариантов которой представлен на рис. 5.



Рис. 5. Итоговая модель (уравнение тренда, гармоника и случайная величина)

После этого сравним эти значения по общим для них 15 точкам. Если значения совпадают (коррелируют), то можно предполагать, что наша имитационная модель адекватна реальному процессу и, следовательно, полученные четыре прогнозных значения могут иметь место. Если же значения не совпадают, то будем генерировать их до тех пор, пока они не совпадут. Современная вычислительная техника позволяет проделать это сотни тысяч раз. Для повышения точности прогноза, нужно иметь как можно больше совпавших значений. Однако проведение статистических испытаний и их оценка, это уже последующие части исследования.

## Заключение

Данная работа выполняется в рамках моделирования Индекса развития человеческого потенциала (ИРЧП) и представляет, как было отмечено, первую часть этого исследования. Основные положения предложенной методологии были обсуждены на методологическом семинаре в Институте социально-политических и правовых исследований АН РБ (февраль 2012 г.). Как не трудно заметить, предложенная методология предполагает моделирование с использованием средств вычислительной техники. Написание соответствующей программы, ее апробирование и проведение статистических испытаний хотя и является следующим этапом намеченного исследования, но первые шаги уже сделаны. На указанном выше семинаре была продемонстрирована в работе начальная версия этой модели, написанная на языке *Visual Basic 6.0* – наиболее подходящем для работы со статданными [9]. Скриншоты, показывающие этапы работы этой имитационной модели как раз и были использованы в данной статье.

## Библиографический список

1. Абдуллин А.Р., Исмагилова В.С. Проблемы статистического моделирования индекса долголетия в РБ. // Инновационные технологии управления социально-экономическим развитием регионов России: Материалы III Всероссийской научно-практической конференции с международным участием. В 2-х частях. Часть I. – Уфа: ИСЭИ УНЦ РАН, 2011. – С. 274-278.
2. Абдуллин А.Р., Ямилов М.М. Анализ случайных факторов, определяющих ожидаемую продолжительность жизни в РБ. // Управление экономикой: методы, модели, технологии: Одиннадцатая Международная конференция с элементами научной школы для молодежи: сб. науч. тр./ Уфимск. гос. авиац. техн. ун-т. – Уфа: УГАТУ, 2011. – С. 287–290.
3. Бендат Дж., Пирсол А. Измерение и анализ случайных процессов. – М.: Мир, 1971. – 408 с.
4. Бокс Дж., Дженкинс Г.М. Анализ временных рядов. Прогноз и управление: В 2-х част. Ч. 1. – М.: Мир, 1974. – 405 с.
5. Бусленко Н.П., Шрейдер Ю.А. Метод статистических испытаний (Монте-Карло) и его реализация на цифровых вычислительных машинах. – М.: 1961. – 226 с.
6. Доугерти К. Введение в эконометрику. – М.: ИНФРА-М, 1999. – 402 с.
7. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: В 2-х кн. Кн. 1. – М.: Финансы и статистика, 1986. – 366 с.

8. Елисеева И.И., Курышева С.В., Костеева Т.В. и др. Эконометрика. – М.: Финансы и статистика, 2007. – 576 с.
9. Коннэлл Дж. Visual Basic 6. Введение в программирование баз данных. – М.: ДМК, 2000. – 720 с.
10. Лукинский В.С. и др. Модели и методы теории логистики. – СПб.: Питер, 20007. – 448 с.
11. Свешников А.А. Прикладные методы теории случайной функции. – М.: Наука, 1968. – 464 с.
12. Соболев И.М. Численные методы Монте-Карло. – М.: Наука, 1973. – 311 с.
13. Суслов В.И., Ибрагимов Н.М., Талышева Л.П., Цыплаков А.А. Эконометрия. – Новосибирск: СО РАН, 2005. – 744 с.
14. Шеннон Р.Ю. Имитационное моделирование систем – наука и искусство / Пер. с англ.; Под ред. Е.К. Масловского. – М.: Мир, 1978. – 418 с.

Басимов М. М., Курган

## Математическое обеспечение рефлексивного исследования на примере межэтнических представлений<sup>1</sup>

---

### Аннотация

В статье рассмотрена адаптация авторского метода множественного сравнения для анализа перекрестных коллективных рефлексивных представлений, значимые производные величины на основе сравнительных весомостей, геометрические интерпретации результатов рефлексивного исследования.

**Ключевые слова:** рефлексивные представления, множественное сравнение, сравнительная весомость, тензор групп оценок, интегральные отличия, экстремальность

Кросс-культурное исследование определяет особенности этнической культуры, изучая коллективные представления людей по прямым оценочным суждениям. Но сравнивая этнические стереотипы восприятия одной культуры представителями других культур, и автостереотипы представителей оцениваемой культуры, неизбежно обнаруживаем несоответствие оценок.

Для получения более объективных результатов А.Б.Хромовым была предложена процедура опроса для изучения этнических особенностей через измерение перекрестных коллективных рефлексивных представлений [2]. Объектом его кросс-культурного исследования первоначально стали студенты России, США и Индии. Мы приведем только отдельные результаты этого исследования, необходимые для демонстрации различных форм результатов математического анализа данных такого опроса.

Коллективные представления студентов изучались на трех уровнях рефлексии (0, 1, 2). Нулевой уровень коллективных представлений измерялся прямыми оценками особенностей своей культуры и особенностей других культур. Первый уровень рефлексивных представлений совпадает с автостереотипами и рефлексивными представлениями других культур за представителей своей культуры. Второй уровень рефлексивного представления – многократное отражение некоторых особенностей культуры с точки зрения представителей другой культуры.

---

<sup>1</sup> Работа выполнена при поддержке РФФИ, проект № 10-06-00413а.



Для оценки культур использовался культурно-ценностный дифференциал, разработанный Г.У.Солдатовой, И.М.Кузнецовым, С.В.Рыжовой, и адаптированный А.Б.Хромовым для многоуровневого рефлексивного оценивания культур по параметрам групповых ценностных ориентаций: на группу, на власть, друг на друга, на социальные изменения [2].

Для трех этносов в результате перекрестного измерения имеем 36 прямых и рефлексивных оценок первого и второго уровня. Оценки разделим на три группы в зависимости от того, какая культура [русская (R), американская (A), индийская (I)] оценивается. Во вводимом для дальнейшего описания обозначении оценки NXYZ четыре позиции имеют следующий смысл: N – уровень рефлексии (0, 1, 2); X – кто оценивает (R, A, I); Y – кого оценивают (R, A, I); Z – с чьей точки зрения оценивают (R, A, I).

По результатам оценивания каждой этнической группы по одному из параметров имеем 12 оценок: по четыре оценки со стороны каждой из трех изучаемых культур: одна прямая оценка (обозначена по позиции Z символом «\*») и три рефлексивные оценки с точки зрения каждой из трех культур. Например, 0IR\* – прямая оценка русской культуры индийцами, 2ARI – рефлексивная оценка русской культуры американцами с позиции индийцев.

Для анализа кросс-культурного рефлексивного исследования был адаптирован авторский метод множественного сравнения [1], предложены геометрические иллюстрации для описания результатов исследования.

Изучаемая матрица данных  $D(N, M)$  рефлексивного исследования состоит из  $S$  (количество сравниваемых прямых и рефлексивных групп оценок, составляющих совокупность из  $N$  оценок по каждому параметру) блоков размерностью  $N_i \times M$ , где  $M$  – количество измеряемых параметров для каждой оценки, а  $N_i$  – количество оценок в каждой из  $S$  сравниваемых групп оценок (кто оценивает, кого оценивают, с чьей точки зрения оценивают).

Количественной характеристикой множественного сравнения является матрица сравнительной весомости  $V(M, S)$ , с использованием которой можно построить различные наглядные распределения групп оценок и изучаемых параметров, отражающие результат множественного сравнения. Матричный элемент  $V_{ij}$  определяет сравнительную значимость  $i$ -го параметра для  $j$ -ой группы оценок.

**Матрица «Сравнительная весомость – I»** строится в результате сравнительного анализа рассматриваемого множества групп оценок в рамках каждого параметра. Реализация алгоритма для одного параметра дает строку матрицы сравнительной весомости. Наглядное представление результата множественного сравнения строится в виде распределений по уровням достоверного отличия изучаемых групп оценок в рамках каждого параметра.

**Матрица «Сравнительная весомость-II»** ортогонального плана по отношению к первому варианту строится исходя из необходимости представления результатов в виде распределения по уровням изучаемых параметров в рамках каждой группы оценок. Данный вариант множественного сравнения можно охарактеризовать как неявный

по отношению к первому варианту, так как несмотря на расчеты с представлением информации в виде распределения по уровням изучаемых параметров в рамках каждой группы оценок учитываются (неявно, через стандартизацию данных на всем множестве рассматриваемых групп оценок) и отличия групп оценок между собой в рамках параметров.

**Матрица «Сравнительная весомость-III»** (обобщенный вариант) строится в результате сравнительного анализа элементов прямого произведения рассматриваемого множества групп оценок и изучаемых параметров. Элементы множества, являющегося прямым произведением множеств выделенных для исследования групп оценок и параметров, обозначим следующим образом:  $P_i/N_j$ , где  $P_i$  – параметр, а  $N_j$  – группа оценок. Всего таких элементов  $P_i/N_j$  в нашем случае будет  $M \cdot S$  ( $M$  параметров и  $S$  групп оценок).

Так как в данном случае необходимо сравнивать между собой выборки, относящиеся как к одному и тому же параметру, так и к разным параметрам, необходимо первичные оценки преобразовать в стандартные  $T$ -баллы (для совокупности всех рассматриваемых групп оценок). Элементы  $P_i/N_j$  прямого произведения множеств групп оценок и изучаемых параметров сравниваются между собой (множественное сравнение) для всех возможных пар  $P_i/N_j$  и  $P_m/N_k$ , где  $i$  и  $m$  принимают значения от 1 до  $M$ , а  $j$  и  $k$  – от 1 до  $S$  (всего  $M \cdot S(M \cdot S - 1)/2$  различных пар  $P_i/N_j$  для  $M$  параметров и  $S$  групп оценок). Сравнение проводится с использованием статистического критерия Стьюдента.

Исходя из стандартизированных оценок вычисляем средние значения  $X_{ij}$  и средние квадратические отклонения  $Y_{ij}$  по каждому измеряемому параметру ( $i$  принимает значения от 1 до  $M$ ) для всех заданных групп оценок ( $j$  принимает значения от 1 до  $S$ ). Имея  $M \cdot S$  средних значений  $X_{ij}$  (и столько же средних квадратических отклонений  $Y_{ij}$ ), проведем их множественное сравнение. Сравниваются попарно все элементы  $P_i/N_j$  прямого произведения множеств выделенных групп оценок и изучаемых параметров. При обнаружении достоверных отличий между элементами  $P_i/N_j$  прямого произведения множеств выделенных групп оценок и параметров выявляется сколько уровней достоверного отличия будет получено для них. Распределение элементов  $P_i/N_j$  по уровням производится так.

Вычисляются критерии Стьюдента для всех возможных пар элементов  $P_i/N_j$  прямого произведения множеств групп оценок и параметров. После чего расчетные значения критериев сравниваются с критическими значениями критерия Стьюдента для уровня значимости, например,  $\alpha=0.05$ . Пользуясь таблицей теоретического распределения Стьюдента выделяются те пары элементов  $P_i/N_j$  прямого произведения множеств выделенных групп оценок и изучаемых параметров, для которых наблюдается достоверное отличие их средних значений.

Для каждого элемента  $P_i/N_j$  (всего  $M \cdot S$ ) прямого произведения множеств групп оценок и параметров, выявляется сколько элементов  $P_i/N_j$   $B(i)$  имеют значения достоверно меньшие и сколько элементов  $P_i/N_j$   $A(i)$  имеют значения достоверно большие, чем элемент  $P_i/N_j$  под номе-

ром  $i$  ( $i$  принимает значения от 1 до  $M \cdot S$ ). Вычисляется разность  $B(i) - A(i)$ , показывающая насколько больше элементов  $P\$/N\%$  находится ниже данного элемента  $P\$/N\%$  под номером  $i$ , чем выше, при сравнении  $M \cdot S$  элементов  $P\$/N\%$  и выявленном при этом достоверном отличии между ними. Минимальному значению разности  $B(i) - A(i)$  соответствует нижний уровень – первый, максимальному значению – высший уровень. Количество уровней очевидно может изменяться от 1 до  $M \cdot S$  для списка из  $M \cdot S$  элементов  $P\$/N\%$  прямого произведения множеств выделенных групп оценок и изучаемых параметров.

Значения  $B(i) - A(i)$  могут быть как положительными, так и отрицательными, изменяясь в интервале от  $1 - M \cdot S$  до  $M \cdot S - 1$ . Значение  $1 - M \cdot S$  соответствует случаю, когда элемент  $P\$/N\%$  с номером  $i$  меньше по величине всех других элементов  $P\$/N\%$  полученного списка, а значение  $M \cdot S - 1$  – если больше по величине всех остальных элементов  $P\$/N\%$  прямого произведения множеств групп оценок и параметров.

Заметим, что в этом случае (в отличие от двух предыдущих) нулю равна только сумма сравнительных весомостей всех  $M \cdot S$  элементов  $P\$/N\%$  прямого произведения.

Начало и конец распределения таких пар (см. таблицу 1).

В обозначенном выше кросс-культурном исследовании с использованием процедуры изучения этнических особенностей через измерение перекрестных коллективных рефлексивных представлений, при сравнении множества прямых и рефлексивных оценок (пар «оценка – параметр групповых ценностных ориентаций») – 36 оценок по 4 параметрам и использовании метода множественного сравнения для всех упорядоченных пар «группа оценок – параметр» максимальное и минимальное значения сравнительной весомости получились следующими:  $VES/\max = 143$  и  $VES/\min = -132$ .

Таблица 1

Уровень	Параметр – рефлексивная оценка	
106	gr/0II*	143
105	gr/2RII law/2IAA	127
104	gr/2RIA	123
103	gr/1RIR	119
102	gr/0RI*	113
-----	-----	-----
5	gr/2AAR int/2AAR	-114
4	int/1RAR	-115
3	gr/1IAI	-126
2	law/0RR*	-130
1	int/0RA*	-132

Выбирая из пар «оценка – параметр групповых ценностных ориентаций» пары, относящиеся к одному изучаемому параметру, получим распределения по уровням сравнительной весомости 36 оценок, относящихся к этому параметру. Эти 36 прямые и рефлексивные оценки разделим на три группы в зависимости от того какая культура (российская, американская, индийская) оценивается. Для примера приведем такую таблицу для параметра «Законопослушность» (law) (см. таблицу 2).

Таблица 2

США	Индия	Россия
NXYZ S/V	NXYZ S/V	NXYZ S/V
2IAA 127	0RI* 108	2IRR 93
0IA* 110	2RII 108	2ARR 33
1IAI 103	1RIR 107	0IR* 7
2RAA 65	2RIA 97	0AR* -29
0RA* 56	2AII 40	2IRA -37
1RAR 49	0AI* 22	1ARA -39
2IAR 36	1AIA 1	2RRI -46
2RAI 24	0II* -11	1IRI -57
1AAA -56	2IIR -20	2ARI -65
0AA* -76	2IIA -27	2RRA -83
2AAI -80	2AIR -39	1RRR -92
2AAR -107	1III -48	0RR* -130

Напомним, что в обозначении оценки NXYZ четыре позиции имеют следующий смысл: N – уровень рефлексии, X – кто оценивает, Y – кого оценивают, Z – с чьей точки зрения оценивают (прямая оценка обозначена по позиции Z символом «\*»); а заголовок в таблице S/V – сравнительная весомость.

Далее для изучения несоответствия спектра прямых и рефлексивных оценок предпочтительней перейти от рассмотрения оценок в рамках каждого изучаемого параметра к рассмотрению этих оценок на множестве всех параметров.

Тензор групп оценок  $TS(S,S)$  описывает интегральное сравнение групп оценок на множестве изучаемых параметров (gr, ch, int, law). Тензор групп оценок представляет симметричную квадратную матрицу  $TS(S,S)$  размерности  $S*S$  (S строк и S столбцов). Матричный элемент, принадлежащий строке i и столбцу j, дает численную характеристику «расстояния» между группами оценок под номерами i и j в пространстве изучаемых параметров, представленными ко векторами сравнительной весомости, определенными для замкнутого множества рассматриваемых групп оценок.

Компоненты тензора групп оценок  $TS_{ij}$  вычисляются следующим образом: суммируются абсолютные значения разностей элементов двух столбцов матрицы сравнительной весомости  $V(M,S)$  с номерами i и j для M параметров [2].

Для наглядности можно распечатать тензор групп оценок покомпонентно в порядке убывания всевозможных парных различий между группами оценок на множестве диагностируемых параметров. Всего компонент у тензора групп оценок  $S^*S$ , из них  $S$  компонент, расположенных по главной диагонали, равны нулю (каждая группа оценок тождественна сама себе). Оставшиеся  $S^*(S-1)$  компонент образуют симметричную структуру. Значит, в покомпонентной распечатке достаточно привести в порядке убывания  $S^*(S-1)/2$  компонент.

Тензор 36 групп оценок представляет собой матрицу  $36 \times 36$ , определяющую «расстояния» между группами в четырехмерном пространстве изучаемых параметров (gr, ch, int, law), которая состоит из 630 различных парных отличий между 36 группами оценок. Для анализа нам понадобятся только те компоненты, которые определяют расстояния между группами оценок, совпадающими по оцениваемой культуре (русской, американской, индийской). Таким образом, следует последовательно рассмотреть три подмножества компонент тензора, каждое из которых содержит по 66 компонент. Для примера приведем попарные (66 пар) интегральные (по четырем компонентам) различия между различными оценками русской культуры.

Таблица 3

**Попарные (66 пар) интегральные (по четырем компонентам) различия между различными оценками русской культуры (#R#)**

0RR* 0AR* 428	0RR* 2IRR 425	0RR* 1ARA 407	1RRR 0AR* 400	1RRR 1ARA 379	2RRA 0AR* 379	2ARI 2IRR 374	1RRR 2IRR 373	0AR* 2IRR 371	2RRA 2IRR 370
1ARA 2IRR 370	0RR* 2ARI 359	2RRA 1ARA 358	0RR* 2IRA 343	0RR* 0IR* 337	2ARI 0IR* 336	0AR* 0IR* 333	1ARA 0IR* 332	1RRR 2ARI 331	2RRI 0AR* 312
2RRA 2ARI 310	2ARR 2ARI 298	0AR* 2ARR 295	2ARR 1ARA 294	1RRR 2IRA 291	2RRI 1ARA 291	2RRA 2IRA 288	1RRR 0IR* 285	0RR* 1IRI 284	2IRR 1IRI 283
2RRA 0IR* 282	0RR* 2ARR 281	2RRI 2ARI 281	1RRR 1IRI 256	2RRI 2IRR 251	0AR* 2IRA 245	0IR* 1IRI 245	2RRA 2ARR 242	2RRA 1IRI 235	1RRR 2ARR 229
2ARI 2IRA 208	2ARR 1IRI 207	1ARA 2IRA 204	2ARR 2IRA 202	2RRI 1IRI 190	1ARA 1IRI 183	2IRR 2IRA 180	1IRI 2IRA 179	0RR* 2RRI 178	2RRI 2IRA 165
2RRI 2ARR 163	2ARI 1IRI 161	2RRI 0IR* 159	0IR* 2IRR 150	0AR* 1IRI 144	2ARR 2IRR 144	0IR* 2IRA 128	0AR* 2ARI 127	1RRR 2RRI 126	2ARR 0IR* 124
2RRI 2RRA 123	0RR* 1RRR 64	2ARI 1ARA 64	0AR* 1ARA 63	0RR* 2RRA 61	1RRR 2RRA 43				

Для получения значимых для описания и интерпретации результатов предпочтительней рассматривать в рамках каждой культуры эти 66 пар оценок по отдельным характерным подмножествам. Рассмотрим примеры таких подмножеств.



1. Интегральные отличия прямой и трех рефлексивных оценок, данных одной культуре представителями другой культуры (в обозначении оценки  $NXYZ$  одинаковы позиции:  $X$  – кто оценивает,  $Y$  – кого оценивают). Приведем для примера интегральные различия между оценками (прямыми и рефлексивными) российской культуры американцами ( $AR\#$ ).

Таблица 4

2ARR 2ARI 298	0AR* 2ARR 295	2ARR 1ARA 294	0AR* 2ARI 127	2ARI 1ARA 64	0AR* 1ARA 63	S=1141
---------------------	---------------------	---------------------	---------------------	--------------------	--------------------	--------

Геометрически соотношения трех рефлексивных оценок на плоскости (три точки в четырехмерном пространстве) можно представить в виде «Треугольника рефлексивных оценок русской культуры американцами»:



А если вернуться к матрице сравнительной весомости, то каждая из четырех оценок (одна прямая и три рефлексивные) характеризуется своеобразным психологическим портретом в рамках изучаемых четырех параметров. При этом психологический портрет каждой из 36 групп оценок рассматривается в единой системе, с общим началом координат – уровнем нулевой сравнительной весомости. Приведем для примера психологические портреты групп оценок российской культуры по прямому и рефлексивным оцениваниям, данным американцами (см. таблицу 5).

Таблица 5

0AR*	1ARA	2ARI	2ARR
law -29 gr -37 ch -83 int -103	law -39 ch -53 gr -58 int -101	ch -48 law -65 gr -66 int -76	law 33 int 29 gr 3 ch -22

2. Подмножества, определяемые парой этносов, которые дают оценку третьему этносу. При рассмотрении трех культур таких подмножеств будет девять. Приведем в качестве примера попарные (16 пар) интегральные (по четырем компонентам) различия между оценками (прямыми и рефлексивными) российской культуры, данными россиянами и американцами ( $RR\#$  и  $AR\#$ ) (см. таблицу 6).

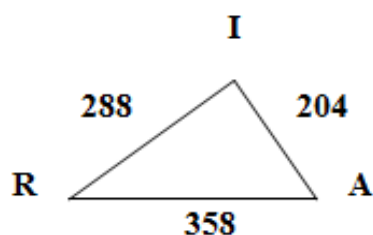
Таблица 6

0RR*	0RR*	0AR*	1RRR	2RRA	0RR*	2RRA	1RRR	2RRI	2RRA
0AR*	1ARA	1RRR	1ARA	0AR*	2ARI	1ARA	2ARI	0AR*	2ARI
428	407	400	379	379	359	358	331	312	310
2RRI	0RR*	2RRI	2RRA	1RRR	2RRI				
1ARA	2ARR	2ARI	2ARR	2ARR	2ARR				
291	281	281	242	229	163				



3. Последние из предлагаемых для описания и анализа подмножества – это сопоставление оценщиков в четырехмерном пространстве изучаемых параметров. В обозначении оценки NXYZ теперь постоянны позиции: Y – кого оценивают, Z – с чьей точки зрения оценивают, а варьируются позиции: X – кто оценивает. В результате имеем при оценивании каждой культуры четыре треугольника (один с прямыми и три с одинаковыми (по позиции Z) рефлексивными оценками). Для примера приведем рефлексивные оценки российской культуры с точки зрения американцев:

2RRA	2RRA	1ARA
1ARA	2IRA	2IRA
358	288	204



Представляются информативными также и результаты суммарных отличий рефлексивных оценок: суммирование по шесть чисел парных отличий оценок, прямых и рефлексивных, по четырем изучаемым параметрам (в приводимом выше примере  $S=1141$ ).

В результате имеем суммарные меры «вариативности» рефлексивности для пар XY, где X – кто оценивает, Y – кого оценивают (матрица 3x3). Далее можно оценить для матрицы суммарных мер «вариативности» (матрица 3x3) суммы по столбцам (вторая позиция) – кого оценивают (оцениваемые) и по строкам (первая позиция) – кто оценивает (оценщики).

В заключение рассмотрим выделенные 36 групп рефлексивных оценок через «уровни экстремальности» групп  $R_j$ , которые представляют собой суммы абсолютных значений элементов столбцов матрицы сравнительной весомости [2].

Величина  $R_j$  определяет количественно уровень доминирования в крайностях сравнительного проявления изучаемых параметров j-ой группы оценок над другими рассматриваемыми группами оценок на множестве одновременно всех изучаемых параметров.

36 прямых и рефлексивных оценок (обозначение NXYZ) полезно разделить на девять групп в зависимости от того какая культура (русская, американская, индийская) оценивается (Y) и кто ее оценивает (X). Ниже для примера приведены экстремальности (один из девяти блоков) оценивания россиянами своей культуры (см. таблицу 7).

Таблица 7

0RR* 234	1RRR 206	2RRA 195	2RRI 134
----------	----------	----------	----------

Представляют интерес (как и для суммарных отличий рефлексивных оценок) различные промежуточные результаты суммирования экстремальностей: кто оценивает и кого оценивают. В результате имеем суммарные меры «экстремальности» рефлексивных представлений для

пар  $XU$ , где  $X$  – кто оценивает,  $U$  – кого оценивают (матрица  $3 \times 3$ ). Далее можно оценить для матрицы суммарных мер «экстремальности» (матрица  $3 \times 3$ ) суммы по столбцам (вторая позиция) – кого оценивают (оцениваемые) и по строкам (первая позиция) – кто оценивает (оценщики).

В заключение отметим, что предлагаемых методы исследования этносов дадут более полную и структурированную картину представления разных культур друг о друге и о себе самих.

### Библиографический список

1. Басимов М.М. Методы множественного сравнения в психологических исследованиях // Методы исследования психологических структур и их динамики. Выпуск 3. / Под ред. Т.Н.Савченко и Г.М.Головиной. – М.: Изд-во ИП РАН, 2005. – С. 128-157.
2. Хромов А.Б. (Россия), Дюби, Бэнки (Индия), Мальхотра, Мамта (Индия), Моррисон, Дороти (США) Рефлексивные представления некоторых особенностей культуры России, Индии и США // Рефлексивные процессы и управление. Тез. IV Межд. симп. (7-9.10.2003). – М.: Изд-во «Институт психологии РАН», 2003. – С.153-154.

Бродский Ю. И., Москва

## Простейшие математические модели взаимодействия культур: толерантность, нетерпимость, идентичность, миграция<sup>1</sup>

---

### Аннотация

Межкультурное взаимодействие моделируется уравнениями конкуренции А. Лотки и В. Вольтера. Применение этой модели к социологическим процессам заметно отличается, например, от ее биологических приложений, применимостью в данной области принципа рациональных ожиданий — предположения, что все агенты модели вооружены знаниями, которые дает эта модель. Обсуждаются результаты моделирования.

**Ключевые слова:** толерантность, нетерпимость, идентичность, миграция, математическое моделирование, межкультурные взаимодействия

### О математическом и гуманитарном анализе сложных процессов и систем

Следуя терминологии предложенной академиком Н.Н. Моисеевым и член-корр. РАН Ю.Н. Павловским, будем называть те системы (процессы, явления), для которых существуют общепризнанные адекватные математические модели, «*простыми*». Те системы, для которых таковых не существует, однако, их адекватный прогноз могут осуществлять имеющие с ними дело специалисты-эксперты, будут называться «*сложными*». Методы (не являющиеся математическими), использующиеся такими специалистами для прогноза поведения сложных систем, будут называться «*гуманитарными*».

Обсуждению проблемы разработки и внедрения технологий, объединяющих математические и гуманитарные средства при исследовании сложных систем, посвящена работа [1]. В ней рассматривается один из способов такого объединения. Он состоит в составлении математической модели некоторой упрощенной системы из интересующего нас класса систем, выполнении прогноза и анализа свойств этой системы математи-

---

<sup>1</sup> Работа выполнена при поддержке РФФИ, грант №11-06-90409-Укр\_ф\_а.

ческими средствами, а затем в гуманитарном анализе того класса систем, к которому относится подвергшаяся математическому моделированию система. При гуманитарном анализе используются понятия и представления, возникшие в процессе математического моделирования. Это позволяет «выйти» за пределы возможностей математического моделирования, т. е. получить прогноз развития системы и ее свойств, который нельзя получить математическими средствами. Кроме того, класс систем, которые можно анализировать таким способом, расширяется. Однако выводы о значениях прогнозируемых характеристик и свойствах анализируемых систем становятся более «расплывчатыми», «приближенными», «ориентировочными».

Попробуем объяснить некоторые употребленные термины. Эти объяснения будут носить гуманитарный характер. Однако трактовки содержания объясняемых терминов будут более узкими, чем это имеет место в обычном словоупотреблении. В процессе объяснений возникнут новые термины, которые также будут поясняться. В этих пояснениях могут появляться возникшие ранее термины. Тем самым будет возникать определенная *языковая среда*.

Если кто-то правильно прогнозирует свойства, характеристики некоторой системы (процесса, явления) и этот факт достоверно установлен, то это означает, что в его распоряжении имеется (адекватная в определенном отношении) *модель* этой системы. Формы, в которой адекватные модели существуют (реализуются), могут быть различны. Математическое моделирование является одной из форм, позволяющей создавать адекватные модели.

Все математические модели, которые далее составляются и изучаются, являются совокупностью соотношений между характеристиками изучаемой системы (процесса, явления). Характеристики, присутствующие в модели (т. е. в ее соотношениях), разбиты на две части: внешние и внутренние. Внутренние характеристики – это те, которые желают узнать, обращаясь к средствам математического моделирования. Внешние характеристики – это те, от которых внутренние зависят, но обратной зависимости в пределах необходимой точности не имеет места. Модель обладает свойством давать прогноз внутренних характеристик, если при известных внешних характеристиках внутренне характеристики можно определить (вычислить) из соотношений модели. Такие модели будут называться «замкнутыми».

Предложенное выше деление систем на простые и сложные не является исчерпывающим. Имеются системы (процессы, явления), которые не поддаются адекватному прогнозу ни математическими, ни гуманитарными средствами. Кроме того, граница между «простыми» и «сложными» системами не является неподвижной: факт «вторжения» математических средств в историю, экономику, социологию, медицину очевиден: математическому моделированию, которое совершенствует свои инструменты, становится доступен все более точный прогноз все более сложных систем, т. е., в соответствии со смыслом понятий «простые» и «сложные» системы, определен-

ном выше, тех, которые ранее были ему недоступны. Можно сказать и так: развитие математического моделирования превращает «сложное» в «простое», конечно в том понимании этих терминов, которое определено выше.

Менее очевиден, в некотором смысле, «обратный» процесс: вторжение гуманитарных методов анализа и прогноза в математические. Чем сложнее система, тем лучше должно быть исходное представление о ней (т. е. ее «понимание») для того, чтобы составляемая упрощенная математическая модель была адекватна. Это «понимание», естественно, является «гуманитарным». Обратим внимание на то, что термин «гуманитарное» тем самым наделен еще одним оттенком содержания, не вполне совпадающим с тем его содержанием, которым оно было наделено выше (как способность давать правильный прогноз развития процессов без использования математических средств).

По-видимому, между математическими и гуманитарными средствами анализа и прогноза систем, процессов, явлений имеются отношения «двойственности». Другими словами, математические и гуманитарные средства анализа и прогноза систем, процессов, явлений друг без друга не существуют и друг друга обеспечивают.

### Моделирование межкультурного взаимодействия уравнениями конкуренции

Математика накопила огромный опыт создания и исследования моделей различных явлений, в основе которого лежит изучение количественных связей между различными величинами, характеризующими явление, и выявление законов изменения характеристик явления, на основе имеющих место связей между ними. Однако, для успешного применения методов математического моделирования, изучаемую предметную область приходится существенно упрощать, абстрагируясь от множества присущих ей деталей.

В данной работе из всех аспектов взаимодействия культур, предлагается выбрать лишь отношение к иной культуре, сравнивая его при этом, с отношением к собственной. В рамках предлагаемой абстракции, можно описывать взаимодействие культур уравнениями конкуренции. Заметим, что хотя исследования различных типов равновесия двумерных конкурентных систем известны со времен А. Лотки и В. Вольтерра (например, [2]), в социальных системах, в отличие, например, от биологических, к вопросу нахождения точек равновесия системы, неизбежно добавляется вопрос об устойчивости этого равновесия относительно рефлектирующих воздействий со стороны популяций. В соответствии с принципом рациональных ожиданий [4] (т. е., предположением, что все действующие агенты модели вооружены развиваемой теорией, и в соответствии с ней, поступают наилучшим для себя образом), естественно ожидать от них попыток улучшить свое положение, путем изменения имеющихся в их распоряжении параметров модели. Как будет видно в дальнейшем, применение принципа

рациональных ожиданий способно сделать неустойчивым устойчивый узел системы дифференциальных уравнений, и, наоборот, сделать точкой устойчивого равновесия точку фазовой плоскости, которая первоначально не была даже стационарной.

Итак, запишем уравнения конкуренции:

$$\begin{aligned} \frac{dN}{dt} &= \alpha N \left(1 - \frac{N}{N^*} - m \frac{M}{M^*}\right), \\ \frac{dM}{dt} &= \beta M \left(1 - \frac{M}{M^*} - n \frac{N}{N^*}\right). \end{aligned} \tag{1}$$

Здесь  $N$  и  $M$  численности популяций, соответственно  $N^*$  и  $M^*$  их емкости среды,  $\alpha$  и  $\beta$  - их коэффициенты прироста, и, наконец,  $n$  и  $m$  - их коэффициенты нетерпимости.

Емкости среды — это предельные численности популяций, которые способны выживать в данной среде обитания.

Коэффициенты нетерпимости показывают во сколько раз конкуренция между популяциями сильнее или наоборот, слабее конкуренции внутри самой популяции.

Диапазон изменения коэффициентов нетерпимости от  $-\infty$  до  $\infty$  будем делить на следующие области (см. таблицу 1).

Таблица 1

$(-\infty, 0)$	$[0, 1)$	1	$(1, \infty)$
сверхтолерантность	толерантность	отношение без предубеждений и предпочтений	нетерпимость

Поясним предложенные названия областей:

- Сверхтолерантность означает, что вместо конкуренции одна популяция помогает приросту другой (из-за отрицательности коэффициента соответствующий член добавляется к производной, а не вычитается).
- Толерантность означает, что с чужой популяцией конкуренция слабее, чем внутри своей. (Что вовсе не означает, что конкуренция с чужой популяцией слабее, чем конкуренция внутри самой чужой популяции, наоборот, она может быть, в том числе и гораздо сильнее.)
- Отношение без предубеждений и предпочтений означает одинаковую конкуренцию как внутри своей популяции, так и с чужой. (Отметим, что при этом сила конкуренции внутри рассматриваемых популяций может весьма существенно различаться.)



- Наконец, нетерпимость означает более сильную конкуренцию с чужой популяцией, нежели внутри своей. (Что опять же не означает, что конкуренция с чужой популяцией сильнее, чем конкуренция внутри самой чужой популяции, наоборот, она может быть, в том числе и гораздо слабее.)

Оказывается, что поведение системы (1) в значительной мере зависит от соотношения между собой коэффициентов нетерпимости. В работе [4] данная модель была подробно исследована, здесь мы приведем основные результаты этого исследования.

Модель (1) можно считать биматричной игрой с непротивоположными интересами, управлениями в которой являются коэффициенты нетерпимости, а выигрышем — предельные численности сторон. Поддержание своих коэффициентов нетерпимости в выделенных выше диапазонах можно считать стратегиями игры. Приведем результаты применения этих стратегий, стараясь если не избежать, то, по крайней мере, минимизировать употребление в данной работе математической символики.

Одним из основных гуманитарных выводов этого математического исследования является то, что даже в рамках такой простейшей модели, оказывается невозможным и неверным измерять и оценивать одну культуру мерками другой (например, сравнивать то, как относятся к коренным жителям «понаехавшие иноплеменники», с тем, как среди самих коренных жителей принято относиться друг к другу). Единственно верной мерой культуры оказывается она сама — плодотворным оказывается только сравнение силы конкуренции с «чужими» с силой конкуренции среди «своих», — учет коэффициентов нетерпимости. При этом точкой отсчета оказывается евангельская заповедь: «возлюби ближнего твоего, как самого себя» (Матф. 22, 39; Мар. 12, 31; Лук. 10, 27).

Качественно различающиеся типы поведения модели зависят в первую очередь от коэффициентов нетерпимости, т. е., от отношения силы конкуренции к чужой популяции с силой конкуренции внутри собственной. При этом совершенно несущественным для результата межкультурных отношений оказывается любое сравнение конкурентного давления «чужих» с внутрипопуляционной конкуренцией тех, на кого они давят. Например, можно с гораздо меньшей силой конкурировать с «чужими», чем они конкурируют между собою, и все равно быть нетерпимым (в нашей терминологии), по отношению к ним. Или же, наоборот, можно конкурировать с «чужими» гораздо жестче, чем это принято в их собственной среде, и все равно оставаться толерантным по отношению к ним, со всеми вытекающими из анализа модели последствиями этого.

Совместное развитие популяций в нашей модели может быть устойчивым относительно их стремления увеличивать свои предельные численности, лишь в условиях реального взаимообогащения ими друг друга (в нашей модели — это состояние обоюдной сверхтолерантности).

Один из выводов анализа модели, расходящийся с тезисом «толерантность спасет мир» — это то, что сочетание стратегий толерантность — нетерпимость, или даже толерантность — отношение без предубеждений

и предпочтений — непозволительная неосторожность со стороны толерантного партнера, ведущая в пределе к его полному исчезновению. Самое плохое, что может случиться с толерантной культурой — ее встреча с культурой нетерпимой!

Интересным оказывается вывод о том, что в условиях обоюдной нетерпимости, популяции не выгодно держаться за то, что когда-то она была «великой», т. е. за большую емкость среды, если ее текущая численность по сравнению с этой емкостью среды стала мала. Для выживания в условиях обоюдной нетерпимости достаточно стабилизировать свою численность, положив коэффициент прироста равным нулю (что, между прочим, противоречит расхожему мнению о возможности «демографической» победы в условиях обоюдной нетерпимости), и поддерживать достаточно высокий уровень внутривнутрипопуляционной конкуренции.

### Возникновение и развитие идентичности

Вслед за Дж. Свифтом, автор считает, что идентичность может возникнуть, в том числе и на вполне пустом месте, как это было у тупоконечников и остроконечников из «Путешествий Гулливера»:

«Поводом к войне послужили следующие обстоятельства. Всеми разделяется убеждение, что варёные яйца при употреблении их в пищу испокон веков разбивались с тупого конца; но дед нынешнего императора, будучи ребёнком, порезал себе палец за завтраком, разбивая яйцо означенным древним способом. Тогда император, отец ребенка, обнародовал указ, предписывающий всем его подданным под страхом строгого наказания разбивать яйца с острого конца. Этот закон до такой степени озлобил население, что, по словам наших летописей, был причиной шести восстаний, во время которых один император потерял жизнь, а другой — корону».

Посмотрим с точки зрения нашей модели, как может развиваться идентичность, однажды возникнув. Пусть, первоначально некая культура развивалась по логистическому закону. Выделим теперь из нее некую часть и предположим, что она, осознав свою общность, уменьшает конкуренцию внутри себя, оставляя прежней силу конкуренции с «внешними» по отношению к ней представителями рассматриваемой культуры. Исследование такой ситуации с помощью нашей модели показывает [2], что выделенная общность получает конкурентное преимущество за счет обретения идентичности.

Между прочим, этот факт вполне оправдывает всеобщую нелюбовь к различным тайным сообществам, особенно если их члены заметно лучше относятся друг к другу, чем ко всем внешним. К примеру, Тертуллиан в своей «Апологии» пишет об отношении римлян к христианам первых веков: «Любовь, которую они питали между собою, казалась опасным заговором».

Предположим теперь, что внутри рассматриваемой культуры стало известно о существовании выделившейся общности, и упомянутая выше нелюбовь к тайным обществам проявилась в соответствующем коэффициенте нетерпимости.

Такая система – пример обоюдной нетерпимости. При этом, у оставшегося вне выделенной группы большинства пространство для маневра шире, оно может беспредельно увеличивать свое давление на выделившуюся группу. Поэтому, скорее всего, «разоблаченное тайное общество» со временем будет вынуждено сойти со сцены.

Можно уравнивать шансы сторон, разрешив образовавшейся идентичности изменять знак коэффициента нетерпимости. В предметной области это означало бы замену внутривнутрипопуляционной конкуренции взаимопомощью. Возможно, именно так и было с христианами первых веков, испытывавшими суровые гонения со стороны Римской империи, и, тем не менее, пережившими эту империю. Во всяком случае, многие источники тех времен говорят о необыкновенной любви, связывавшей членов общины: «Все же верующие были вместе и имели все общее. И продавали имения и всякую собственность и разделяли всем, смотря по нужде каждого». «У множества же уверовавших было одно сердце и одна душа; и никто ничего из имения своего не называл своим, но все было у них общее...» «Не было между ними никого нуждающегося; ибо все, которые владели землями или домами, продавая их, приносили цену проданного и полагали к ногам Апостолов». «Верующих же более и более присоединялось к Господу, множество мужчин и женщин...» Деян. 2:44-45; 4:32,34; 5:14.

### Миграционные процессы

Попробуем посмотреть на процессы, связанные с миграцией населения, с позиций простейшей математической модели взаимодействия культур на основе уравнений конкуренции П. Ферхюльста, а также А. Лотки и В. Вольтерра, подробно описанной в [3].

Предположим, что начальное состояние нашей системы – это две страны, народонаселение которых без учета взаимодействия между этими странами, изменялось бы по закону конкуренции П. Ферхюльста.

Предположим, что конкурентное давление во второй стране выше, чем в первой. В предметной области нашей модели это неравенство можно трактовать как то, что в первой стране живется легче, чем во второй.

Предположим, что некоторая часть жителей второй страны осознает этот факт, и решает поискать лучших условий жизни в первой стране. Возможно при этом, мигранты сталкиваются не с тем отношением к себе, которое принято в среде коренных жителей первой страны, а с определенным их отношением к «чужим», которое корректируется соответствующим коэффициентом нетерпимости.

При этом мы тем не менее предполагаем, что конкурентное давление во второй стране все равно выше давления жителей первой страны на «чужаков», что и располагает к миграции. Естественно в первом приближении предположить, что количество мигрантов пропорционально с некоторым коэффициентом населению второй страны, а также разнице в условиях жизни мигрантов в первой стране, и условий внутри второй страны.

Далее, предположим, что некоторая часть мигрантов интегрируется в культуру первой страны, становясь, тем самым, полноправной с точки зрения нашей модели ее частью. Оставшаяся часть образует диаспору.

Сделанных предположений достаточно для того, чтобы записать динамику всех трех компонент нашей системы, что и было выполнено в [3].

Получилась трехмерная нелинейная система дифференциальных уравнений, достаточно сложная для аналитического исследования. Тем не менее, опыт исследования двумерной конкурентной модели может подсказать некоторые тенденции ее поведения, которые в дальнейшем проверялись с помощью численных экспериментов, результаты которых приводятся далее.

Например, если «принимающая» сторона толерантна к мигрантам, то она вполне может исчезнуть — произойдет полная замена исходной культуры на культуру мигрантов. Между прочим, история даже XX века знает подобные примеры.

Следующий ниже график иллюстрирует подобную ситуацию (см. рис. 1).

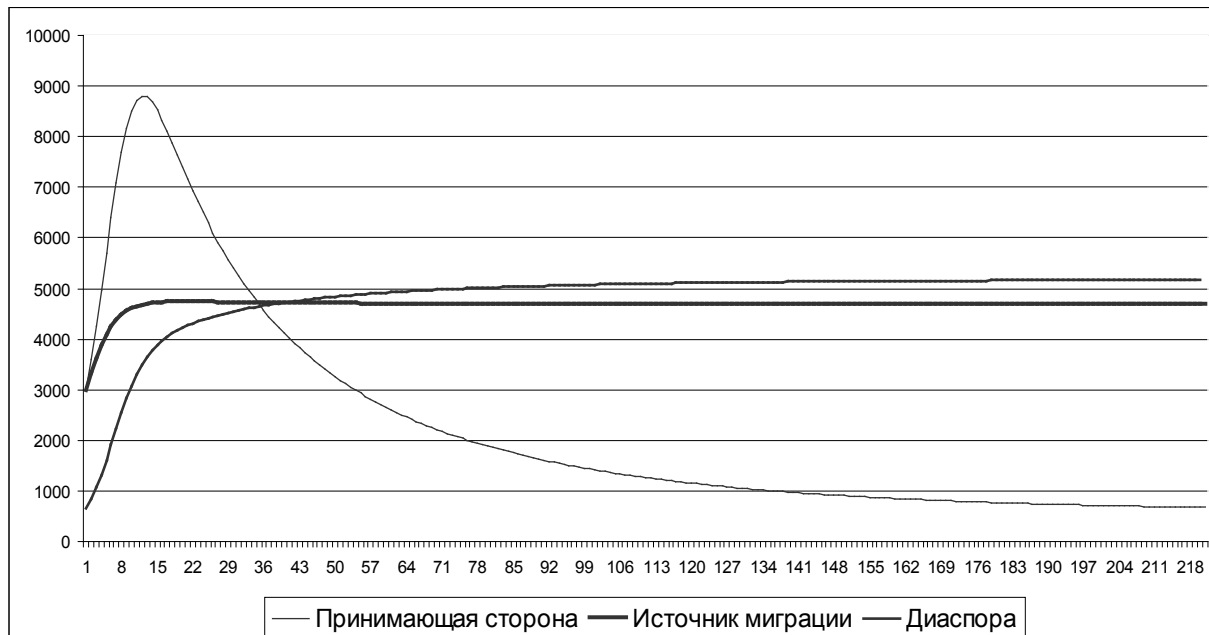


Рис. 1. Принимающая сторона толерантна к мигрантам

В нашей модели принимающая сторона может достаточно легко избежать описанной выше печальной участи, увеличив свой коэффициент нетерпимости, так, чтобы давление на мигрантов превосходило давление в их культуре. Такое неравенство гарантирует полное отсутствие миграции в нашей модели.

Рисунок 2 иллюстрирует сказанное.

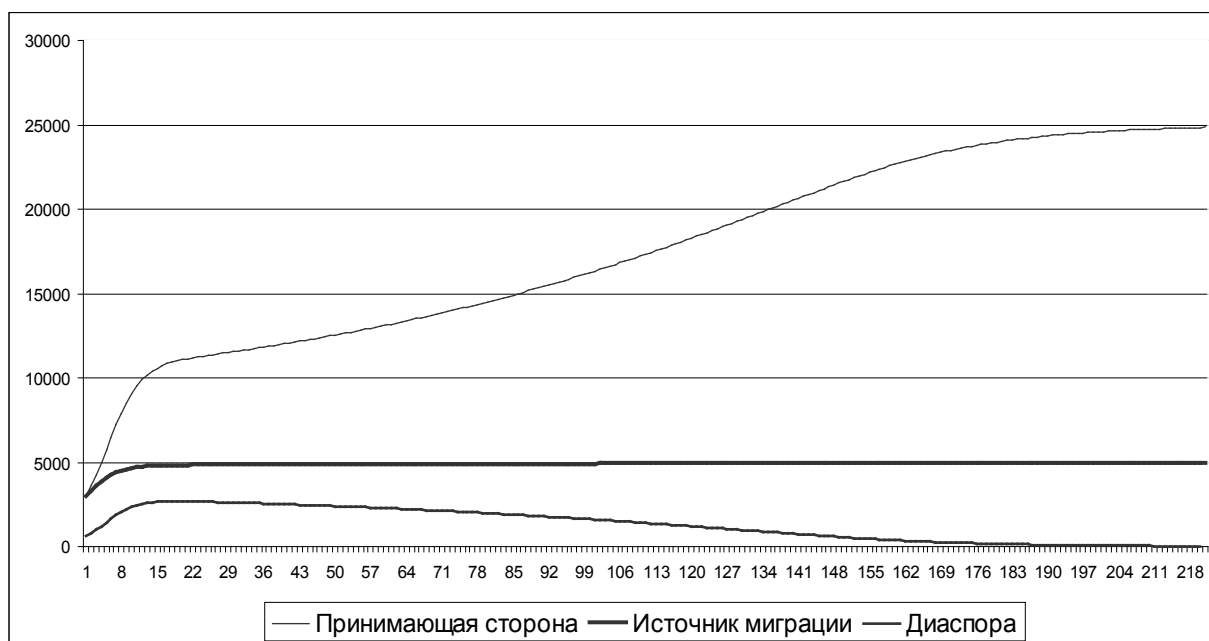


Рис. 2. Принимающая сторона нетерпима к мигрантам

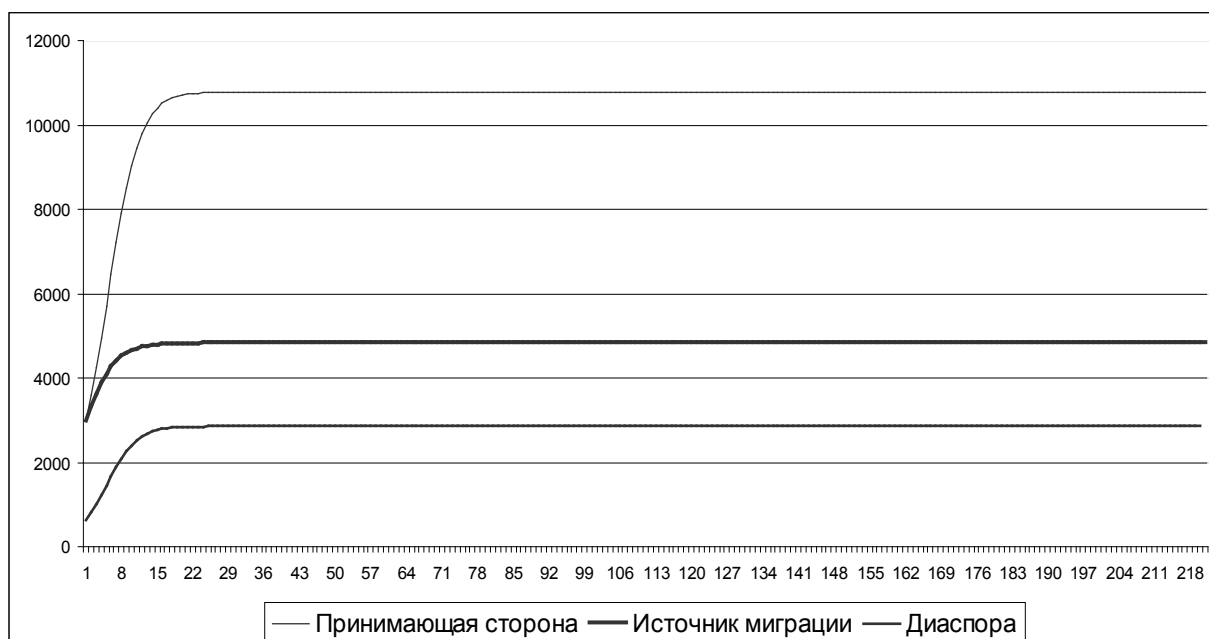


Рис. 3. Поддержание равновесия в системе с миграцией («ласковая нетерпимость»)

По-видимому, если по каким-то причинам миграция необходима принимающей стороне – оптимальной для нее будет тактика «ласковой нетерпимости», т. е. поддержание своего коэффициента нетерпимости  $n$ , на таком уровне, чтобы в системе оставалось необходимое число мигрантов. Приведенный на рисунке 3 результат численного эксперимента показывает, что последнее также вполне возможно в нашей системе.

## Библиографический список

1. Белотелов Н.В., Бродский Ю.И., Павловский Ю.Н. Сложность. Математическое моделирование. Гуманитарный анализ. М.: Книжный дом «ЛИБРОКОМ», 2009, 320 с.
2. Бродский Ю. И. ТОЛЕРАНТНОСТЬ, НЕТЕРПИМОСТЬ, ИДЕНТИЧНОСТЬ: простейшие математические модели взаимодействия культур – Saarbrücken: LAP Lambert Academic Publishing, 2011, 68 с.
3. Бродский Ю. И. Исследование процессов миграции с помощью математической модели взаимодействия культур // Сборник «Моделирование, декомпозиция и оптимизация сложных динамических процессов» / Отв. Ред. Ю.Н. Павловский. – М.: ВЦ РАН, 2011, С. 88-94.
4. Поспелов И.Г. Принцип рациональных ожиданий: обзор концепций и примеры моделей. М.: ВЦ РАН, 2008, 79 с.
5. Триандис Г.К. Культура и социальное поведение. М.: ФОРУМ, 2007, 384с.



Гаврилец Ю. Н., Москва

## Динамическая модель политического соперничества двух сил в обществе<sup>1</sup>

---

### Аннотация

В работе предложена математическая модель взаимодействия двух противоположных политических сил в обществе в условиях идеологической борьбы. Модель реализована на компьютере в пакете MATCAD на условных примерах.

**Ключевые слова:** социальные группы, установка, подражание, цепь Маркова, дифференциальные уравнения, вероятность перехода

Социально-политическая поляризация в России за последние годы не перестаёт ослабевать. Всё большая часть населения оказывается втянутой в активную борьбу за свою политическую позицию. СМИ и взаимные контакты между людьми формируют эти позиции, зачастую вопреки реальным интересам людей и общества в целом – в силу действия социально-психологических законов подражания, иррациональности и стадности поведения больших масс людей [1]. В связи с этим построение и анализ математических и компьютерных моделей указанных процессов представляет оправданный интерес [2].

Рассматривается условное общество [3], состоящее из пяти взаимодействующих групп переменной численности:

$y$  – численность малополитизированной (пассивной) группы;

$x$  – численность политизированной группы «зелёных»;

$z$  – численность политизированной группы «синих»;

$xx$  – численность активной группы «зелёных»;

$zz$  – численность активной группы «синих».

Общее количество участников в политической жизни фиксировано, так что выполняется баланс:

$$x+z +xx+zz+y=N.$$

Члены пассивной части населения имеют базовую социально-политическую установку, характеризующую склонность к поддержке «зелёных» (неподдержка «синих»), значения которой распределены на

---

<sup>1</sup> Работа выполнена при поддержке РФФИ, грант №10-06-00362.

действительной числовой оси согласно Гауссовому распределению с нулевым математическим ожиданием и дисперсией  $\sigma^2$  [1]. Изменение этой установки на величину  $\psi$  описывается дифференциальным уравнением

$$d\psi/dt = A(x-z) - a\psi,$$

где  $A$  – коэффициент подражания большинству, и коэффициент  $a$  характеризует скорость затухания приобретённого (положительного или отрицательного) «подражательного доверия»  $\psi$ . Величина  $\psi$  определяет вероятность перехода части пассивного населения в ту или иную политизированную группу. Предполагается, что для конкретного пассивного индивида с установкой  $u^+$   $\psi$  вероятность перейти в группу «зелёных» равна значению соответствующего интеграла вероятности с дисперсией  $k^2$ . Можно показать, что доля пассивных, переходящих в момент  $t$  в политизированную группу «зелёных», в этом случае будет равна:

$$\Psi_1 := \xi_1 \cdot \left[ \frac{1}{\sqrt{(k^2 + \sigma^2) \cdot 2 \cdot \pi}} \cdot \left[ \int_0^{\psi_t} \exp \left[ \frac{-x^2}{2 \cdot (k^2 + \sigma^2)} \right] dx \right] + 0.5 \right]$$

где все коэффициенты считаются известными. Доля переходящих в «синие» выражается похожим образом:

$$\Psi_2 := \xi_2 \cdot \left[ \frac{-1}{\sqrt{(k^2 + \sigma^2) \cdot 2 \cdot \pi}} \cdot \int_0^{\psi_t} \exp \left[ \frac{-x^2}{2 \cdot (k^2 + \sigma^2)} \right] dx + 0.5 \right]$$

Здесь коэффициенты  $\Psi_1$ ,  $\Psi_2$  показывают, какая доля из пассивных, но поддерживающих «зелёных» или «синих», переходят в эти группы ( $x$  или  $z$ ).

Члены политизированных групп имеют установки  $e_1$ ,  $e_2$ , которые определяют вероятности их перехода в группу активной борьбы. Эти установки определяются в зависимости от соотношения численностей групп  $x$ ,  $z$ , а также под влиянием внешнего стандарта (пропаганда, СМИ и т. д.). Их динамика описывается двумя дифференциальными уравнениями [4,5]. Таким образом, все три установки задаются уравнениями (в разностном виде):

$$\begin{pmatrix} e1_{t+1} \\ e1_{t+1} \\ \psi_{t+1} \end{pmatrix} := \begin{pmatrix} e1_t \\ e1_t \\ \psi_t \end{pmatrix} + h \begin{bmatrix} A1 \cdot \left[ \frac{x_t \cdot e1_t \cdot 1.5 + z_t \cdot (-e2_t)}{x_t + z_t} - e1_t \right] + B1 \cdot \left[ (E1 - e1_t) \cdot \exp \left[ - \left( \frac{E1 - e1_t}{12} \right)^2 \right] \right] \\ A2 \cdot \left( \frac{-x_t \cdot e1_t + z_t \cdot e2_t \cdot 1.7}{x_t + z_t} - e2_t \right) + B2 \cdot (E2 - e2_t) \cdot \exp \left[ - \left( \frac{E2 - e2_t}{13} \right)^2 \right] \\ A \cdot (z_t - x_t) - a \cdot \psi_t \end{bmatrix}$$

где параметры

A1, A2 суть коэффициенты подражания среднему мнению в среде политизированных групп;

B1, B2 – коэффициенты влияния внешней пропаганды E1, E2 на установку члена группы.

Численность политизированных групп меняется также вследствие непосредственных контактов между ними. Кроме того часть населения из активных «борцов» (xx, zz) переходит в более спокойные группы. В результате мы получаем динамический процесс, описываемый приведёнными дифференциальными уравнениями изменений установок, а также некоторой марковской цепью для пяти численностей групп. Матрица P переходных вероятностей равна:

$$\begin{bmatrix} 1 - \gamma1 - \alpha1 \cdot \frac{z_t}{x_t + z_t} - \varphi \cdot \left( \text{cnorm} \left( \frac{e1_t}{15.5} \right) \right) & \alpha2 \cdot \frac{x_t}{x_t + z_t} & \rho & 0 & \Psi1 \\ \alpha1 \cdot \frac{z_t}{x_t + z_t} & 1 - \gamma2 - \alpha2 \cdot \frac{x_t}{x_t + z_t} - \omega \cdot \left( \text{cnorm} \left( \frac{e2_t}{15} \right) \right) & 0 & \mu & \Psi2 \\ \varphi \cdot \left( \text{cnorm} \left( \frac{e1_t}{15.5} \right) \right) & 0 & -\rho & 0 & 0 \\ 0 & \omega \cdot \left( \text{cnorm} \left( \frac{e2_t}{15} \right) \right) & 0 & -\mu & 0 \\ \gamma1 & \gamma2 & 0 & 0 & -\Psi1 - \Psi2 \end{bmatrix}$$

а динамика описывается соотношениями:

$$\begin{pmatrix} x_{t+1} \\ z_{t+1} \\ xx_{t+1} \\ zz_{t+1} \\ y_{t+1} \end{pmatrix} := P \cdot \begin{pmatrix} x_t \\ z_t \\ xx_t \\ zz_t \\ y_t \end{pmatrix}$$

Параметры  $\alpha_1, \alpha_2, \gamma_1, \gamma_2, \delta, \omega, \rho, \mu, \varphi$  обозначают доли переходящих из одних групп в другие. Величины  $\text{sporm}(e/d)$  обозначают (из пакета MATHCAD) интеграл вероятности со среднеквадратическим отклонением  $d^2$ , а  $\text{fspnorm}(e/d)$  выражает вероятность соответствующего перехода в группу активных действий при установке  $e$ .

В пакете MATHCAD были произведены расчёты при разных комбинациях значений параметров и получены разные варианты сходимости к устойчивому стационарному состоянию общей системы. Как нам представляется, главной особенностью данного подхода явилось объединение дифференциальных уравнений с цепью Маркова в единую модель. По-видимому, при наличии необходимой статистической информации с помощью подобной модели можно не только прогнозировать ход политической борьбы, но и управлять этим процессом.

На рисунках ниже показаны отдельные траектории общего процесса (смысл обозначений был указан в тексте).

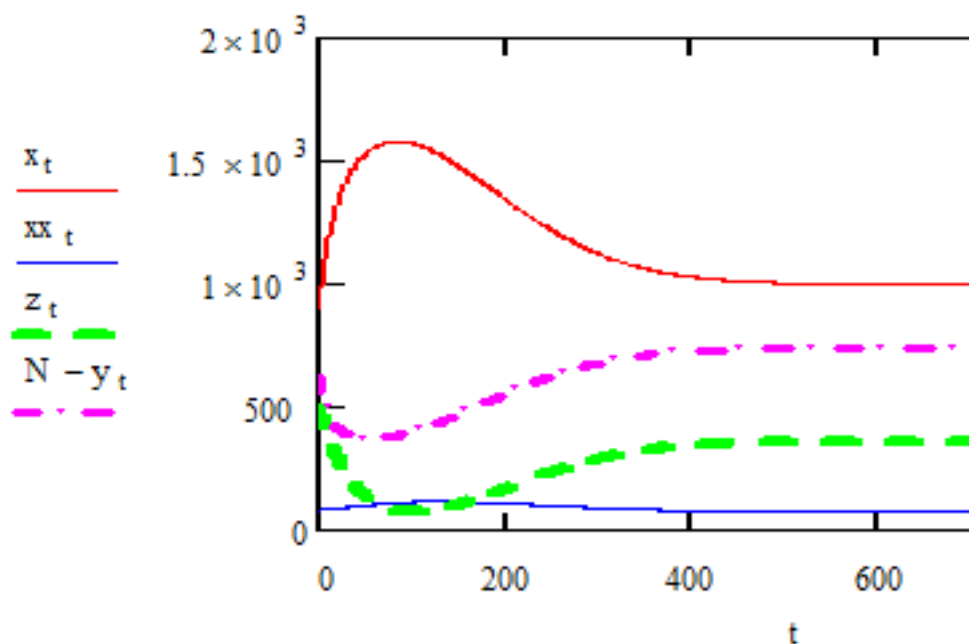


Рис. 1

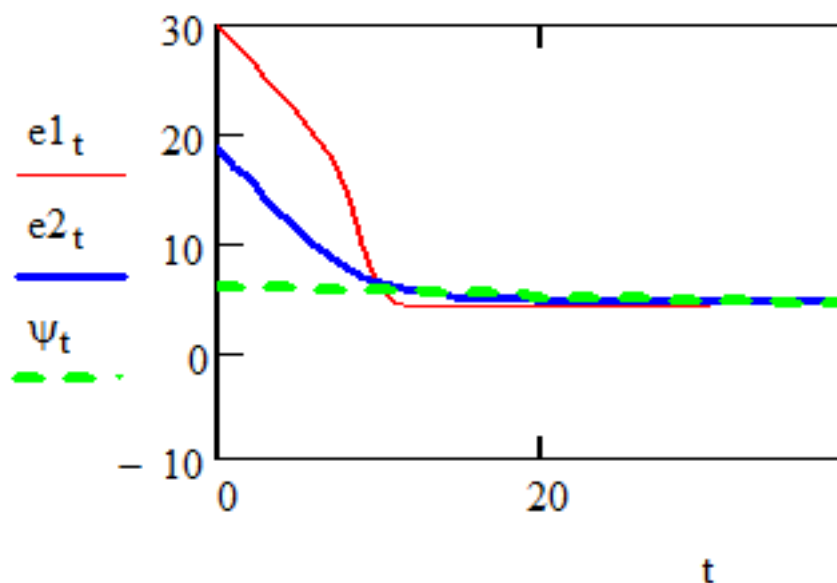


Рис. 2

На этих рисунках видно, что при заданных значениях параметров стационарные величины всех установок реализуются гораздо быстрее, чем значения численностей взаимодействующих групп.

## Библиографический список

1. Н.Рашевский, Две модели: подражательное поведение и распределение статуса, в сборнике «Математические методы современной буржуазной социологии», М., «Мир», 1966.
2. Ю.Н.Гаврилец, К синтезу теории систем и кибернетики в экономике, М., 2009, Международная академия организационных наук, ротапринт ЦЭМИ РАН.
3. В.Л.Макаров, Социальный кластеризм, М., Бизнес Атлас, 2010.
4. Ю.Н.Гаврилец, Стохастическое моделирование межгрупповых информационных взаимодействий, ЭММ, №2, 2003.
5. W.Weidlich, Sociodynamics, Taylor & Francis, 2002.

Галицкий Е. Б., Москва

## Категориальный метод главных компонент и нелинейная оптимизация в задаче прогнозирования итогов выборов

---

### Аннотация

В статье рассматриваются проблемы, возникающие при прогнозировании итогов выборов, и принципы преодоления этих проблем, на основе которых автор строит прогнозы, используемые в Фонд Общественное Мнение. В ходе анализа данных применяются категориальный метод главных компонент и нелинейный метод обобщенного понижающего градиента.

**Ключевые слова:** электоральные прогнозы, опросы населения, метод главных компонент, нелинейный метод обобщенного понижающего градиента

### Основные проблемы прогнозирования итогов выборов

Сегодня стало устойчивой традицией проведение опросов для предсказания итогов выборов. Мы не будем здесь обсуждать проблему репрезентативности таких опросов. Будем считать, что наши выборочные опросы не содержат систематических смещений по сравнению с тем, как если бы мы опросили всех без исключения избирателей. Рассмотрим лишь проблемы анализа данных. Мы не будем приводить принадлежащих Фонду Общественное Мнение конкретных методик прогнозирования, разработанных в результате этого анализа, обсудим лишь его основные принципы.

Со стороны задача прогнозирования выборов кажется простой. Надо спросить попавших в выборку избирателей, как они проголосуют, а все проголосуют примерно так же. Однако когда необходимо получить достаточно точный прогноз, возникают сложности.

Действительно, анкета предвыборного опроса всегда содержит прямой вопрос о том, как респондент планирует поступить в день выборов. В предлагаемом респонденту меню ответов наряду со списком возможных кандидатов<sup>1</sup>, есть варианты «испортил бы бюллетень» и «не пошёл бы на выборы». Если бы каждый респондент мог выбрать один из этих вариан-

---

<sup>1</sup> Здесь и далее имеются в виду, как кандидаты, так и партии.



тов, явку на выборы можно было бы оценить по доле давших любой ответ, кроме «не пойду на выборы», а распределение голосов избирателей – по доле выбравших соответствующие варианты ответа от числа учтённых при расчёте явки. К сожалению, многие респонденты (так называемые, неопределившиеся) не говорят, что сделают в день выборов.

Их доля даже за неделю до выборов бывает значительной: 20% и более. (Например, за неделю до выборов Президента РФ 4 марта 2012 года затруднились ответить на такой вопрос 18,0% респондентов, причём точно такая же доля (18,0%) была и за месяц до выборов). Когда нужно приблизительно оценить распределение сил между кандидатами, неопределившиеся респонденты обычно просто исключаются из рассмотрения, и расчёты выполняются без их учёта. Тем самым негласно предполагается, что голоса неопределившихся распределятся так же, как голоса остальных, определившихся. Предположение это, однако, совершенное безосновательно, и в этом состоит первая проблема, снижающая точность прогнозирования.

Вторая проблема состоит в том, что доля тех, кто не выбрал вариант «не пойду на выборы» при ответе на прямой вопрос анкеты, всегда оказывается намного выше, чем доля реально участвующих в выборах. Например, за неделю до выборов Президента РФ выбрали такой вариант лишь 8,2% респондентов, т. е. 10% от числа определившихся. Фактически же была зафиксирована явка 65,34%, то есть не пришли голосовать почти 35% избирателей.

Таким образом, одного «лобового» вопроса для прогноза результатов выборов недостаточно, нужен ещё хотя бы один детальный вопрос о склонности к участию в выборах. Приведем процентное распределение ответов на такой вопрос, заданный за неделю до Президентских выборов 2012 г.:

Таблица 1

Совершенно точно приму участие в выборах	54,1
Вероятнее всего приму участие в выборах	20,5
Скорее приму, чем не приму участие в выборах	10,2
Скорее не приму, чем приму участие в выборах	3,4
Вероятнее всего не приму участия в выборах	2,6
Совершенно точно не приму участия в выборах	5,5
Ещё не решил(-а), затрудняюсь ответить	3,7

Учитывая зафиксированную ЦИК явку, можно заключить, что в выборах участвовала какая-то (достаточно высокая) доля респондентов, общившихся, что совершенно точно придут на выборы, какая-то другая – что вероятнее всего придут на выборы и т. д. А если учесть, что электоральные предпочтения, а следовательно, и голоса респондентов, намеревавшихся совершенно точно прийти на выборы, могли распределиться не совсем так, как голоса, с более низкой готовностью участвовать в выборах, становится

понятно, что соотношение вероятностей прихода на выборы представителей каждой из вышеупомянутых семи групп влияет на прогноз не только явки, но общего распределения голосов.

Таким образом, для точного прогноза итогов выборов необходимо решить, по крайней мере, две проблемы: проблему распределения голосов тех, кто не сказал, что сделает в день выборов, и проблему толкования ответов респондентов о склонности принять участие в выборах.

Если бы анкета состояла только из приведённых выше двух вопросов и методика прогнозирования разрабатывалась бы для каждого выбора независимо, решить эти проблемы было бы нельзя. Согласно подходу ФОМ, о будущих действиях респондентов, затруднившиеся ответить на первый вопрос, можно судить по их ответам на другие, дополнительные вопросы об отношении к кандидатам, а о вероятности участия в выборах – путем сравнения распределения ответов на соответствующий вопрос с явкой, зафиксированной на ранее проведённых выборах.

### Принцип прогнозирования поведения неопределившихся респондентов

Как уже отмечалось, поведение на выборах неопределившихся респондентов прогнозируется по их ответам на ряд косвенных вопросов электоральной направленности. В частности, в уже упоминавшемся опросе респондентам задавались следующие вопросы:

1. Скажите, пожалуйста, к кому из перечисленных политиков Вы относитесь положительно?

2. А к кому из перечисленных политиков Вы относитесь отрицательно?

3. Если Вы примете участие в президентских выборах, то как, за кого из кандидатов Вы проголосуете? (Далее: «голосование»).

4. Скажите, пожалуйста, Ваше намерение голосовать за названного политика (испортить бюллетень, не участвовать в выборах) может или не может измениться? Вопрос задавался только определившимся респондентам. (Далее: «возможность изменений»).

5. Если Вы ещё не определились, то, может быть, Вы могли бы назвать двух-трёх кандидатов, из числа которых Вы, скорее всего, станете выбирать в марте 2012 года? Вопрос задавался только неопределившимся респондентам.

6. Скажите, пожалуйста, Вы допускаете или исключаете для себя возможность проголосовать за ...? Вопросы данной серии последовательно задавались о каждом кандидате. В качестве вариантов ответа предлагались варианты: «безусловно, допускаю», «скорее, допускаю», «скорее, исключая», «безусловно, исключая» и «затрудняюсь ответить». (Далее: «шкала»).

7. Скажите, пожалуйста, предвыборная агитация в пользу каких кандидатов Вам нравится?

8. А предвыборная агитация в пользу каких кандидатов Вам не нравится, вызывает раздражение?<sup>1</sup>

Благодаря разнообразию вопросов, подавляющее большинство (в данном случае 98,3%) респондентов хотя бы раз выказало какое-либо отношение к предстоящим выборам. Таким образом, дополнительные семь вопросов позволили сократить область полной неопределённости более чем на порядок: с 18,0% до 1,7%.

Однако, как по ответам на все эти вопросы распределить между кандидатами голоса неопределившихся респондентов? Например, кому приписать голос респондента, сказавшего, что он положительно относится сразу к двум кандидатам, если за первого из них он скорее допускает возможность проголосовать, но предвыборная агитация ему не нравится. Предвыборная же агитация второго кандидата респонденту нравится, но он затрудняется сказать, допускает ли возможность за него проголосовать. Заметим: это ещё простая ситуация, здесь ещё можно принять какое-либо решение из содержательных соображений. В реальности же встречаются более замысловатые сочетания ответа на перечисленные выше вопросы, и для приписывания таких респондентов к тому или иному кандидату необходимы чёткие количественные основания. Эти основания мы получаем с помощью метода главных компонент.

Его применение в данном случае базируется на гипотезе, что в основе электорального поведения лежит индикатор – скрытая (латентная) переменная, характеризующая отношение респондентов к каждому кандидату. Респондент голосует за того кандидата, индикатор отношения к которому у него выше. Если максимальные значения индикатора наблюдаются у двух и более кандидатов, голос этого респондента делится между ними поровну.

Поскольку мы планируем сравнивать значения индикатора у разных кандидатов, правило его построения по исходным переменным должно быть единым. Поэтому данные для анализа главных компонент должны быть подготовлены так, чтобы каждая переменная содержала ответы респондентов сначала о первом, затем о втором и т. д. кандидате. Только тогда, сопоставив ответы на разные вопросы, можно найти то общее, что лежит за ответами, т. е. количественно описать отношение респондентов к кандидату.

Заметим, в нашем случае ответы на все вопросы, кроме одного, напрямую относятся к какому-либо кандидату. Вопрос же «возможность изменений» нельзя напрямую включать в число исходных переменных: хотя ответы на него тоже говорят об отношении респондента к кандидату, но чтобы узнать, к какому именно, надо учесть ответ на вопрос «голосование». Ответ «моё решение, безусловно, не может измениться» может свидетельствовать как о высокой приверженности респондента к определённому кан-

---

<sup>1</sup> Когда речь идёт о партиях, задается ещё один вопрос: может ли респондент назвать себя сторонником какой-либо партии, и если да, то какой.

дидату (если ранее он выразил намерение за него проголосовать), так и об абсолютном отсутствии такой приверженности (если респондент назвал другого кандидата, выказал намерение испортить бюллетень или не идти на выборы). Поэтому на основе переменных «голосование» и «возможность изменений» предварительно нужно сформировать: «степень уверенности голосования» со следующими категориями: «голосует и, безусловно, исключает возможность изменить своё решение», «голосует и, скорее, исключает возможность изменить своё решение» и т. д. вплоть до «не голосует и, безусловно, исключает возможность изменить своё решение».

Итак, мы имеем таблицу данных, каждая клетка которой характеризует отношение определённого респондента к определённому кандидату. При этом клетки одной строки характеризуют это отношение с самых разных сторон. Если бы все переменные были бинарными измерялись в интервальной шкале, индикатор отношения к кандидатам можно было бы получить с помощью метода главных компонент, реализованного в SPSS в группе методов факторного анализа. Действительно, этот метод позволяет сформировать новую переменную, объясняющую наибольшую долю дисперсии исходных переменных. Далее метод позволяет построить вторую переменную, объясняющую наибольшую часть оставшейся дисперсии, третью и т. д. Для наших целей важна только первая из этих новых переменных, будем называть её первым фактором.

По нашему предположению в основе ответов респондентов на все перечисленные выше вопросы о каком-либо кандидате лежит скрытая, возможно даже от них самих (латентная) переменная, комплексно характеризующая их отношение к этому кандидату. Поскольку первый фактор объясняет больше всего дисперсии, он лучше, чем любая другая переменная, способен объяснить, почему на вопросы о том или ином кандидате респондент ответил именно так. Поэтому первый фактор можно считать искомым индикатором, комплексно описывающим скрытое (латентное) отношение респондентов к кандидатам. А так как фактор представляет собой линейную комбинацию исходных переменных, то с помощью регрессионного анализа легко найти его «формулу», правило, по которой он построен.

Метод анализа главных компонент применим, только когда все исходные переменные являются бинарными или измеряются в интервальной шкале. В нашем случае это совершенно не так. Например, переменная «шкала» имеет варианты ответов: «безусловно, допускаю», «скорее, допускаю», «скорее, исключаю», «безусловно, исключаю», «затрудняюсь ответить». Довольно часто исследователи перекодируют такие переменные, помещая позицию «затрудняюсь ответить» между «скорее, допускаю» и «скорее, исключаю», после чего трактуют получившуюся шкалу как интервальную, т. е. шкалу равных интервалов. Такой подход, конечно, допустим, но не в таких ответственных случаях, когда результаты выборов нужно предсказать как можно точнее. Здесь же приходится учитывать, что в действительности различие между позициями «безусловно, допускаю» и «скорее, допускаю» содержательно может быть несколько иным, чем, например, между «скорее, допускаю» и «затрудняюсь ответить» и т. д.



Отказ от трактовки такого рода переменных как интервальных влечёт необходимость использования не классического, а категориального метода главных компонент, реализованного в SPSS в блоке «Categories». Этот метод итеративно меняет количественную трактовку (квантификацию) каждой категории каждой переменной, начиная с тривиальной трактовки: категория с кодом 1 – число 1, категория с кодом 2 – число 2 и т. д. На каждом шаге алгоритма метод «пробует» выявить закономерности путем построения на квантифицированных переменных заданного числа факторов (в нашем случае – одного), а затем пробует модифицировать квантификацию. Пока квантификация остаётся недостаточно эффективной, факторы будут объяснять относительно небольшую часть дисперсии. Метод прекращает работу, когда удаётся оптимальным образом «истолковать» категории всех переменных и построить факторы.

Надо сказать, что квантификации переменных представляют отдельный интерес. Они позволяют измерить реальные смысловые различия между разными категориями исходных переменных. Иногда может оказаться, что между какими-то категориями различия значительны, а между какими-то их практически нет. Исходя из этого, иногда принимается решение о слиянии каких-то категорий исходных переменных. Именно так были слиты все категории, кроме «голосует и безусловно исключает возможность изменить своё решение» и «голосует и скорее исключает возможность изменить своё решение» синтетической переменной «степень уверенности голосования», о которой говорилось выше. После этого категориальный анализ главных компонент был реализован снова.

Чтобы выявить правило построения индикатора отношения к кандидатам, можно, например, применить линейный регрессионный анализ, взяв в качестве зависимой переменной фактор, а в качестве независимых – квантифицированные исходные переменные. Полученная формула, дополненная правилами квантификации, позволяет построить такой же индикатор на любом новом опросе. А поскольку в нашем случае абсолютное значение индикатора не играет роли, а важно лишь знать, при каком сочетании исходных переменных это значение выше, а при каком – ниже, мы обычно применяем к индикатору линейное преобразование, такое, чтобы наименьшее из в принципе возможных его значений равнялось нулю баллов, а наибольшее – ста баллам.

Так мы решаем первую проблему: толкование ответов неопределившихся респондентов относительно того, что они сделают в день выборов<sup>1</sup>. Заметим: как видно из приведённого выше изложения, индикатор строится по ответам всех, а не только неопределившихся респондентов. Иначе было бы невозможно проследить закономерности, связывающие намерение сделать то или иное в день голосования с ответами респондентов на другие вопросы анкеты.

---

<sup>1</sup> Как видно из приведённого выше изложения, индикатор строится по ответам всех, а не только неопределившихся респондентов. Это необходимо для того, чтобы проследить закономерности, связывающие намерение сделать то или иное в день голосования с ответами респондентов на другие вопросы анкеты.

При использовании приведённых выше вопросов слова респондента, что он намерен проголосовать за определённого кандидата, сразу обеспечивают ему более высокое значение индикатора отношения к этому кандидату, чем у любых респондентов, не высказавших такого намерения напрямую, вне зависимости от их ответов на остальные электоральные вопросы<sup>1</sup>. Поэтому и прогноз по индикаторам делается не только по неопределившимся, а по всем респондентам, и это не влияет на результат. Сказавшие, как проголосуют, распределяются по ответам на прямой вопрос, а неопределившиеся — в зависимости от значений индикатора.

### Принцип оценки вероятности явки респондента на выборы

Обсудим теперь принципы отыскания вероятности участия в выборах, которую следует приписать каждому респонденту в зависимости от его ответа на уже приводившийся выше вопрос: «Скажите, пожалуйста, Вы примете или не примете участие в президентских выборах 4 марта 2012 года?» Опишем сначала базовый принцип определения такой вероятности, а затем кратко остановимся на трёх приёмах, позволяющих повысить качество прогноза.

Поскольку ФОМ за последние год-два прогнозировал результаты многих региональных выборов, у нас есть возможность сопоставить высказывания респондентов об их намерении прийти на выборы с официально зафиксированной явкой. Для этого можно рассчитать доли респондентов, давших тот или иной ответ на приведенный выше вопрос («совершенно точно приму участие», «вероятнее всего приму участие в выборах» и т. д.), после чего приписать им какие-либо пробные вероятности участия и определить, исходя из них, расчётную явку, сопоставив её с официальной. Затем можно подправить приписываемые вероятности, стремясь уменьшить сумму квадратов расхождений между расчётной и официальной явками. Процесс такой оптимизации продолжается до тех пор, пока эта сумма квадратов не достигнет минимума. Для этого можно использовать, например, нелинейный метод обобщенного понижающего градиента (ОПГ), реализованный в надстройке «Поиск решения» пакета Microsoft Excel. Таков основной принцип решения данной проблемы. Отметим теперь три частных принципа, делающих применение общего принципа более эффективным.

Первый частный принцип полезен в связи с тем, что, как уже отмечалось, вопрос о намерении участвовать в выборах предусматривает семь вариантов ответа. Соответственно, при описанном выше подходе пришлось бы отыскивать слишком много — семь — параметров модели. Это число

---

<sup>1</sup> Кстати, попытки изменить данную ситуацию в ФОМ делались, но оказались неудачными. Таким образом, нам пока не удалось «разоблачить» путем задавания «хитрых» вопросов тех не слишком искренних респондентов, которые, например, в ответ на прямой вопрос говорят, что проголосуют за «кандидата власти», а в ответах на другие вопросы симпатии к этому кандидату не демонстрируют. Как будет показано ниже, проблему такого рода неискренних социально одобряемых ответов мы решаем иным способом.



сопоставимо с числом региональных выборов, по данным которых данная оптимизация должна была бы осуществляться, а следовательно, решение могло бы оказаться неустойчивым. Дело осложняется ещё и тем, что применяемый алгоритм оптимизации отыскивает не глобальный, а локальный минимум. Её результат может зависеть от начальных условий, и в семимерном пространстве шансы наткнуться на такую зависимость существенно выше. Для уменьшения числа параметров оптимизации с семи до двух принимается гипотеза, что искомая кривая вероятности участия в выборах является двухпараметрической S-образной, например, интегральной функцией нормального распределения<sup>1</sup>.

Второй частный принцип позволяет решить проблему социально одобряемых ответов, которая стоит у нас весьма остро. Для её решения строится не одна, а две кривые вероятностей участия в выборах. Одна для респондентов, хотя бы часть голоса которых согласно прогнозу отходит находящемуся у власти кандидату, вторая — для всех остальных респондентов. Опыт показывает, что вторая кривая проходит существенно выше первой.

И, наконец, третий принцип — одновременно минимизировать сумму квадратов отклонений процента явки и процента голосов за кандидата власти. Как показывает опыт, указанные три частных принципа заметно повышают точность прогнозирования.

## Заключение

Итак, мы рассмотрели основные принципы, согласно которым автор прогнозирует итоги выборов. Эти прогнозы сопоставляются с ещё одним прогнозом, иногда на основе информации качественного характера вносятся небольшие поправки, после чего формируются официальные прогнозы Фонда Общественное Мнение. Как показывает опыт, такая процедура обеспечивает весьма высокую точность прогнозирования.

---

<sup>1</sup> В качестве аргумента используются не номера категорий, а соответствующие им квантифицированные значения, которые находятся с помощью категориального анализа главных компонент по данным специального расширенного опроса, где задавались два дополнительных вопроса: ходит ли обычно респондент на выборы, и посещают ли обычно выборы его близкие и знакомые.

Зангиева И. К., Москва

## К вопросу о заполнении пропусков в социологических данных

---

### Аннотация

Статья посвящена различным аспектам заполнения пропусков в данных. Описываются основные разновидности неполной социологической информации: недостижимые и неполные наблюдения, созданные пропуски. Анализируется связь между причинами неответов на вопросы и степенью случайности порождаемых ими пропусков, определяющей допустимость их заполнения.

**Ключевые слова:** неполная информация, неполные наблюдения, отдельные пропуски, заполнение пропусков, неответы на вопросы, степень случайности, заполнение пропусков

### Виды неполной информации

Начать следует с фиксации объекта исследования-основного фокуса данной статьи. В ней речь идет, прежде всего, об отдельных пропусках в данных, соответствующих неполным наблюдениям. Отдельные пропуски в данных являются частным случаем неполной информации, наряду с недостижимыми наблюдениями и сознательно созданными пропусками. Кратко охарактеризуем каждый из них.

Цель полевого этапа любого эмпирического социологического исследования – собрать максимальное количество релевантных данных, то есть получить ответы на все вопросы (в ситуации опроса) от всех запланированных респондентов. Если респондент отвечает на все поставленные вопросы ему соответствует полное наблюдение (full response). Задача полевого этапа - максимизация количества полных наблюдений. В идеале, который практически недостижим, их должно быть 100%.

В реальности исследователь всегда имеет дело с неполной информацией. Всегда есть респонденты, которых не удалось опросить и респонденты, которые будучи опрошенными, не ответили на некоторые заданные им вопросы. В первом случае возникают недостижимые наблюдения (unit – nonresponse), во втором - неполные наблюдения (отдельные пропуски-item /partial nonresponse). Неполные наблюдения, для которых известна только

часть информации, являются промежуточным звеном между полными наблюдениями (известная вся информация) и недостижимыми наблюдениями (нет информации вообще).

### Недостижимые наблюдения (недостижимость респондентов)

В отечественной литературе в качестве синонима недостижимости часто используется понятие труднодоступности. Труднодоступными считаются респонденты, которые не могут (длительные командировки, болезни и т. д.), не хотят принять участие в опросе, которых трудно застать дома или невозможно опросить по причине того, что они проживают на отдаленных территориях [1; 4]. Таким образом, недостижимыми являются респонденты, которых не удалось опросить в принципе.

Недостижимость в массовых социологических опросах представляет собой серьезную проблему. Наличие труднодоступных единиц наблюдения является источником систематических ошибок, т. к. недостижимые респонденты могут существенно отличаться от тех, кто в итоге принял участие в исследовании и ответил на вопросы [3, с. 59-61].

Масштабы недостижимости часто используются в роли индикаторов качества проведенного исследования и, прежде всего, реализации полевого этапа: «Высокий процент откликов стал синонимом эффективной и высококачественной организации исследования» [8, с. 25].

### Неполные наблюдения (отдельные пропуски)

Далее понятия «неполные наблюдения» и «отдельные пропуски» будут использоваться как синонимы.

Отдельные пропуски в данных можно разделить на два вида: реальные пропуски и вынужденные пропуски.

Реальные отдельные пропуски возникают когда, несмотря на все усилия исследователя или интервьюера (анкетера), респондент не отвечает на некоторые вопросы.

Вынужденные отдельные пропуски возникают в результате чистки массива, осуществляемой по завершении сбора и ввода данных. При чистке массива удаляются нереалистичные, заведомо ложные, нарушающие логику варианты ответа. Последние имеют место, если на один из вопросов респондент дает ответ, противоположный другим ответам на взаимосвязанные вопросы, нарушая тем самым всю логику последовательности [10, с. 145]. В литературе вынужденные пропуски так же называют искусственными.

Можно привести следующие примеры заведомо ложных значений. Подросток в качестве уровня образования указывает «кандидат наук», человек без определенного места жительства указывает площадь квартиры,

в которой якобы проживает на данный момент. Чтобы заведомо не внести ложную информацию в данные, эти ответы из массива будут удалены, в результате чего возникнут искусственные, «артефактные» пропуски, или просто артефакты. [2].

Следует отметить, что наличие пропусков в данных наносит существенный урон качеству исследовательских результатов из-за:

- искажения распределений признаков (в некоторых случаях - возникновению систематических смещений);
- снижения статистической мощности результатов анализа данных в силу сокращения объема выборки;
- перехода порядковых шкал в частично упорядоченные;
- перехода непрерывных шкал в дискретные.

### Сознательно созданные пропуски

Ограниченность ресурсов накладывает ограничения на стоимость, время проведения проекта и на количество вопросов, и, следовательно, тем, которые включаются в инструментарий. В некоторой степени уменьшить влияние данных факторов удастся за счёт разбиения анкеты на несколько частей, предназначенных для различных групп респондентов, в рамках одного панельного исследования с чередующимися темами. Опрос в рамках одной волны нескольких групп респондентов по разным анкетам позволяет сэкономить временные, финансовые ресурсы и при этом ослабить нагрузку на респондентов.

Такой способ отдельного сбора данных с последующим объединением (слиянием) данных получил название data fusion.

Пропуски по вопросам блока анкеты, отсутствующему в данной волне некоторая группа респондентов не опрашивалась, затем заполняются с помощью стандартных алгоритмов заполнения отдельных пропусков [5; 11; 12].

Следует разделять пропуски, сознательно созданные исследователем еще на этапе планирования исследования, и незапланированные вынужденные пропуски, о которых было сказано выше. Первые представляют собой пропуски, полученные в ходе заранее спланированного экспериментального дизайна исследования, вторые же имеют вынужденный характер и изначально запланированы не были.

Решение проблемы недостижимости определенных респондентов и отдельные аспекты data fusion представляют собой крайне перспективные и актуальные направления для самостоятельных исследований и разработок, поэтому в данной работе мы не будем их далее рассматривать, а сосредоточимся только на работе с отдельными пропусками в данных.

Далее перейдем к работе с отдельными пропусками в данных. Существует 3 основных подхода к работе с ними уже после сбора данных: удаление неполных наблюдений, взвешивание имеющихся наблюдений для

достижения запланированного объема выборки и искусственное заполнение пропусков. В данной статье мы будем говорить только о заполнении пропусков, так как этот подход наиболее распространён в современной исследовательской практике и методической литературе и представляется наиболее перспективным.

### **Заполнение пропусков как центральный подход к работе с пропусками**

В пользу актуальности заполнения пропусков для современной исследовательской практики и методической литературы говорит следующее. Статьи, посвященные различным аспектам заполнения пропусков, появляются в таких журналах, как *Sociological Methods and Research* (издательство Sage), *Sociological Methodology* (издательство Wiley), *International Journal of Social Research Methodology* (издательство Taylor & Francis). Первый из этих журналов занимает 6 место в рейтинге влияния 132 социологических журналов (данные Thomson Reuters, 2011).

Следует говорить именно об искусственном заполнении пропусков, так оно происходит уже «постфактум», с помощью математических или, что встречается значительно реже, логических процедур. Искусственности заполнения, в упомянутом смысле, можно было бы избежать, повторно обращаясь к каждому не ответившему на определенный вопрос респонденту с просьбой все-таки дать ответ на вопрос.

Заполнение пропусков имеет четыре основных сравнительных преимущества относительно удаления неполных наблюдений или взвешивания полных.

Во-первых, в отличие от взвешивания полных наблюдений, заполнение пропусков позволяет реально сохранить объем выборки на запланированном уровне.

Во-вторых, при заполнении пропусков, наряду с приращением новой информации, сохраняется вся известная информация, которая могла быть утеряна при удалении наблюдений с пропусками или взвешивании имеющихся.

В-третьих, в отличие от взвешивания полных наблюдений, заполнение пропусков не вызывает смещений по другим переменным, значения которых известны или в данный момент не восстанавливаются.

В-четвертых, после заполнения пропусков запланированный анализ данных может осуществляться в обычном режиме. Не нужно вводить дополнительных поправок, как например при взвешивании. Массив данных воспринимается и анализируется, как будто изначально от всех респондентов были получены ответы на все вопросы, и пропусков в данных не было в принципе.

Наряду с названными преимуществами заполнение пропусков как способ решения проблемы недостающей информации имеет несколько недостатков, которые нельзя не учитывать:



1. Использование для предсказания пропусков имеющихся данных может исказить общую структуру данных, которая смещается в сторону структуры только полных наблюдений.

2. Искусственное заполнение вносит в массив определенную долю (равную доле пропусков, в том случае если заполнялись все пропуски) искусственных данных.

Можно встретить точку зрения о неэтичности математического заполнения пропусков. Критики заполнения пропусков говорят о его неэтичности, обусловленной «вменением» не ответившим на вопрос респондентам «искусственных», рассчитанных или подобранных математическими способами значений (ответов), которые затем выдаются за истинные.

Нам данное соображение кажется в корне ошибочным. При заполнении пропусков не стоит задачи точного «угадывания» сокрытого ответа каждого не ответившего респондента. Задача заключается в восстановлении общего распределения изучаемого признака, искаженного наличием пропущенных значений. Здесь важно понимать, что заполнение пропусков математическими методами применяется в первую очередь в массовых количественных исследованиях, основанных на опросе большого числа респондентов. В силу «количественности» при анализе данных важно получить выводы обо всей изучаемой совокупности, а не о каждом ее отдельном представителе. Поэтому ответ каждого отдельного респондента как таковой значения не имеет. Исходя из этого, при заполнении пропусков происходит восстановление максимально достоверной статистической картины всей совокупности, а не «угадывание» ответа каждого не ответившего респондента или вменение, приписывание ему искусственно определенных значений [9].

Точность «угадывания» пропущенных значений используется как показатель эффективности заполнения пропусков в специальных методических экспериментах со смоделированными пропусками. В реальных же исследованиях точность подстановки нельзя оценить, так как истинное значение неизвестно.

При заполнении пропусков «физическое» приписывание ответов (значений переменных) – подстановка некоторых чисел на место каждого пропуска в массиве данных с помощью статистического пакета происходит только для того чтобы сделать возможной обработку данных с помощью традиционных методов анализа данных, предполагающих работу только с полными наблюдениями. По итогам заполнения пропусков ни в коем случае не говорится, что конкретный не ответивший на вопрос респондент на самом деле ответил согласно значению, подставленному на место имеющегося у него пропуска. Данное высказывание, действительно было бы неэтичным.

При заполнении пропусков этика не нарушается. Необходимо соблюдать этику при презентации и публикации результатов исследования. Этические соображения здесь требуют от исследователя в отчете по



результатам исследования или в любой другой публикации результатов указания на то, что имело место заполнение указанного количества пропусков конкретным способом (алгоритмом).

Однако, даже с соблюдением всех этических норм и только для получения обобщенных результатов обо всей совокупности в целом, заполнение пропусков допустимо и правомочно далеко не всегда. Допустимость заполнения пропусков определяется их характером, а именно степенью случайности.

### Степень случайности пропусков как условие допустимости их заполнения

По степени случайности в литературе выделяют полностью случайные пропуски (missing completely at random – MCAR), случайные пропуски (missing at random – MAR) и неслучайные пропуски (not missing at random – NMAR).

Смысл каждого вида случайности можно пояснить на примере опроса следующим образом. Каждому вопросу в соответствие можно поставить случайную величину «ответ-неответ». Тогда степень случайности пропусков в ответах на конкретный вопрос определяется теми факторами, от которых зависит вероятность неответа респондентов на соответствующий вопрос (т. е. вид распределения дихотомической случайной величины «ответ-неответ», «привязанной» к каждому вопросу):

- при полной случайности пропусков вероятность неответа на вопрос не зависит ни от возможного ответа на данный вопрос, ни от ответов на другие вопросы. Распределение дихотомической величины «ответ-неответ» в данном случае одинаково при всех значениях данной переменной и при всех значениях остальных переменных.

- при случайности пропусков вероятность неответа не зависит от ответа на данный вопрос, но зависит от ответов на другие вопросы. Когда пропуски случайны распределение случайной величины «ответ-неответ» одинаково при всех значений рассматриваемого признака, но разное в группах, выделенных по значениям других рассматриваемых признаков.

- при неслучайности пропусков вероятность неответа на вопрос зависит от того, какой вариант ответа имеется в виду. Когда пропуски неслучайны и имеют систематический характер, распределение случайной величины «ответ-неответ» определить невозможно, так как оно разное для каждого значения рассматриваемой переменной [6, с.154-155].

При этом учитываются только факторы, отраженные в имеющейся у социолога информации о респондентах, т. е. в ответах на другие вопросы анкеты.

Степень случайности является математическим конструктом, оторванным от ситуации реального социологического исследования. Было бы полезно, помимо математического, найти и содержательное обоснование

допустимости заполнения пропусков. В качестве такого содержательного обоснования можно рассмотреть причины возникновения пропусков. Предположение о том, что пропуски каждой степени случайности порождаются определенными причинами, требует для своей проверки установления связи между причинами возникновения пропусков и степенью случайности порождаемых этими причинами пропусков.

В литературе выделяются три группы причин, по которым респонденты не отвечают на вопросы: психологические (различные характеристики личности респондента), социальные (особенности социальной ситуации и социального окружения в которых разворачивается ситуация опроса) и методические (различного рода ошибки, допущенные исследователем на этапе планирования исследования или интервьюером на этапе сбора данных) [7, с. 403-410].

Авторы, изучавшие основные причины возникновения пропусков (неответов респондентов на отдельные вопросы) не связывали причины неответов на вопросы со степенью случайности порождаемых ими пропусков. Аналогичное утверждение справедливо и для работ, посвященных изучению пропусков разной степени случайности: в этих работах практически не уделяется внимания причинам их возникновения. Другими словами, причины пропусков и их рассмотрение с точки зрения случайности в литературе рассматриваются отдельно. Это в определенном смысле естественно: первым аспектом фактически занимаются люди, решающие содержательные задачи (в нашем случае – социологи), вторым – математики. Попытаемся ликвидировать этот недостаток.

Говоря о выделенных выше типах причин возникновения пропусков нельзя установить жесткое соответствие между каждым типом причин и каждым типом пропусков по степени случайности.

В рамках каждой группы одни причины могут вызывать полностью случайные или случайные пропуски, а другие – не случайные.

На самом деле не всегда можно вычленить единственную причину, по которой респондент не ответил на вопрос. Процесс вопросно-ответной коммуникации иногда может быть подвержен влиянию нескольких причин одновременно. И, определение причин возникновения пропусков должно быть основано не столько на строгих доказательствах, сколько на опыте исследователя и его знаниях об особенностях темы и объекта исследования.

Выводы о связи между причинами возникновения пропусков, типами порождаемых ими пропусков по степени случайности и допустимыми способами работы с последними резюмируются в следующей таблице (см. таблицу 1).

Таблица 1

**Связь между причинами возникновения пропусков,  
их типами и допустимыми способами корректировки после сбора данных**

	Вероятность неответа на вопрос:	Степень случайности, пропусков	Допустимый способ корректировки после сбора данных
<b>Причины возникновения пропусков</b>	Не зависит от возможного ответа	Полностью случайные (MCAR)	Удаление  Взвешивание  Заполнение
		Случайные (MAR)*	
	Зависит от возможного ответа	Неслучайные (NMAR)	Не поддаются

\*Для случайных пропусков перед взвешиванием и заполнением необходимо разбиение выборки на части, внутри которых пропуски полностью случайны.

Выше было отмечено, что между общими типами причин и степенью случайности порождаемых ими пропусков установить однозначное соответствие нельзя. Но можно утверждать следующее. Для обоснования допустимости корректировки после сбора данных (удаления, взвешивания выборки или заполнения пропусков) пропусков в ответах на определенный вопрос необходимо определить возможные причины возникновения последних и проанализировать связь между этими причинами и вероятностью неответа. Неслучайные пропуски, исключающие возможность их ликвидации после сбора данных, возникают под влиянием социальных, психологических или методических причин, только если последние ставят вероятность неответа на вопрос в зависимость от самого возможного «истинного ответа» (значения характеристики, измеряемой данным вопросом, которое было бы получено в случае ответа).

Поэтому, если у исследователя действительно есть основания полагать, что может иметь место ситуация возникновения неслучайных пропусков, похожая на одну из описанных выше, корректировка пропусков должна заключаться в максимальном устранении причин, породивших эту неслучайность. Заполнять неслучайные пропуски даже с помощью самых сложных алгоритмов некорректно, так как при заполнении пропусков алгоритмы, так или иначе, используют имеющиеся данные. Но в случае неслучайных пропусков, респонденты, не ответившие на вопрос, отличаются от ответивших как по значениям рассматриваемого признака, так и по значениям других признаков. Поэтому, некорректно при заполнении неслучайных пропусков использовать имеющиеся данные с совершенно другим распределением.

Еще одна практическая рекомендация, которую можно сделать, основываясь на приведенных выше размышлениях и примерах, касается ситуации, когда исследователь имеет дело со случайными пропусками. При случайных пропусках, когда вероятность неответа зависит от значений другого признака, внутри групп выделенных по значениям этого «другого» признака присутствует свое распределение вероятности неответа на вопрос. Поэтому заполнением пропусков необходимо разбить совокупность на группы, внутри которых пропуски по данной переменной полностью случайны (случайная величина «ответ-неответ» имеет одинаковое распределение внутри группы), и заполнять пропуски внутри каждой группы в отдельности. При работе со случайными пропусками возникает проблема поиска признаков определяющих случайность пропусков. Определить эти признаки необходимо для того, чтобы разбив по их значениям выборку, добиться в каждой подвыборке полной случайности пропусков. В первую очередь, в качестве таковых признаков имеет смысл рассматривать объективные характеристики респондентов, смысл которых очевиден, понятен и слабо зависит от способа измерения. И, здесь важно понимать, что зафиксировать абсолютно все признаки, определяющие именно случайность пропусков по некоторой переменной нельзя, потому что круг имеющихся потенциальных признаков – факторов случайности ограничен только признаками, изучаемыми в данном исследовании. И, может сложиться ситуация, что относительно изучаемых в исследовании признаков пропуски могут быть полностью случайными, а относительно не рассматриваемых в нем признаков – случайными. То есть, выводы о полной случайности и случайности пропусков, и соответственно допустимости их заполнения, могут быть справедливы только с точностью до признаков, изучаемых в рамках данного конкретного исследования.

Когда пропуски полностью случайны и в совокупности имеет место одно распределение случайной величины «ответ-неответ» заполнение пропусков правомочно в полной мере, так как респонденты, не ответившие на вопрос, не отличаются от ответивших ни по значениям рассматриваемого признака ни по значениям других признаков, и использование при заполнении пропусков имеющихся данных при корректной реализации не внесет в структуру данных и результаты их анализа специфических смещений.

### Библиографический список

1. Бутенко И.А. «Нет ответа». Анализ методической ситуации на страницах журнала «Public Opinion Quarterly» // Социологические исследования. 1986. № 4. С.118-122.
2. Зангиева И.К. Проблема пропусков в социологических данных: смысл и подходы к решению // Социология: 4М (методология, методы, математические модели). 2011. № 33. С.28-56.

3. Хайкин С.Р., Павлов Э.П. Как помочь интервьюеру (из опыта методических исследований) // Социологические исследования. 1992. №4. С.48-64.
4. Чурилов Н.Н. Труднодоступные единицы исследования – источник систематических ошибок // Социологические исследования. 1986. № 1. С.64-78.
5. Adamek J. Fusion: Combining data from separate sources // Marketing Research: A Magazine of Management and Applications. 1994. Vol.6. No. 3. P.48-56.
6. D.de Leeuw E., Hox J., Huisman M. Prevention and Treatment of Item Nonresponse // Journal of Official Statistics. 2003. Vol. 19. No.2. P.155-156.
7. Ferber R. Item Nonresponse in a Consumer Survey // Public Opinion Quarterly. 1966. Vol. 30. No. 3. P. 399-415.
8. Ineke A.L.S. The Hunt for the Last Respondent. Nonresponse in sample surveys. Hague: Social and Cultural Planning Office of the Netherlands, 2005. P.18-35.
9. Rubin D.B. Multiple Imputation for Nonresponse in Surveys. New York: Willey, 1987. P. 64-69.
10. Sande I. Imputation in Surveys: Coping with Reality // The American Statistician. 1982. Vol.36. No.3. P.145-152.
11. Wagner K., Wedel M. Factor Analysis and Missing Data // Journal of Marketing Research. 2000. No.11. P. 490-498.
12. Wagner K., Wedel M. Statistical Data Fusion for Cross – Tabulation // Journal of Marketing Research. 1997. No.11. P. 485-497.

Лукичев П. Н., Ростов-на-Дону

## Критерий демократичности президентских выборов

---

### Аннотация

Принципы демократии – равное для всех избирательное право, свободное волеизъявление граждан, тайное голосование – обуславливают стохастический характер процесса выборов, что определяет возможность построения вероятностной модели его результатов.

**Ключевые слова:** закон биномиального распределения вероятностей; электоральные предпочтения; моделирование выборов; достоверность итогов голосования

На протяжении ряда лет мною публиковались статьи, связанные с отработкой методики прогнозирования президентских выборов и верификации легитимности их проведения [1; 2; 3; 4]. Вернуться к этой теме необходимо по причине уточнения нюансов модели и успешности ее апробирования за рубежом, в частности, во Франции на президентских выборах 2012 г., хотя нас, конечно, в первую очередь интересует свой собственный опыт, который и будет рассматриваться ниже.

Демократически осуществляемая процедура выборов является стохастическим процессом, в котором каждый из кандидатов, выражающий интересы и волю какой-либо социальной группы, имеет все основания рассчитывать на свою победу. Это позволяет расценивать шансы на успех каждого из претендентов как теоретически равные, что соответствует принципу равных прав граждан демократического государства избирать и быть избранным. С другой стороны, равное и прямое избирательное право при свободном и тайном голосовании означает, что каждый из избирателей при отсутствии иных побудительных мотивов с равной долей вероятности может отдать свой голос любому из кандидатов, а значит, исходные вероят-

ности мы должны принять как равные и считать  $p = \frac{1}{m}$  (где  $p$  – исходная вероятность,  $m$  – число кандидатов, зарегистрированных избирательной комиссией). Однако, что бы сами претенденты о себе и своих шансах ни думали, в действительности вероятность победы на выборах задана неко-



торой суммой факторов, которые к моменту объявления о регистрации кандидатов в президенты уже вполне проявили свое действие и определили их рейтинг в общественном мнении.

Этот рейтинг в ходе предварительного социологического опроса населения высвечивается с той степенью надежности, с которой такое исследование является репрезентативным, валидным и корректным. Следствием наличия рейтинговой оценки кандидатов является изменение исходно равных вероятностей на отличающиеся и весьма существенно. Данное отличие наиболее полно отражает закон биномиального распределения вероятностей:

$$P_j = C_{m-i}^{j-i} p^{j-i} q^{m-j}$$

где  $i$  – номер строки;  $j$  – номер столбца,  $m$  – число претендентов на президентское кресло.

Матрица вероятностей перехода каждого из претендентов в президентское кресло будет иметь вид:

$$\|P_j\| = \begin{vmatrix} P_0 & P_0 & P_0 & \dots \\ 0 & P_1 & P_2 & \dots \\ 0 & 0 & P_2 & \dots \end{vmatrix}$$

Причем, по сути, в действительности нас интересуют только две первые строки матрицы, из которых первая характеризует вероятности рассматриваемого события с участием действующего президента, который переизбирается на следующий срок, т. е. пытается сохранить президентское кресло за собой (вероятность  $P_{00}$ ), а вторая – вероятности при участии в выборах претендентов, не занимавших никогда раньше пост президента страны.

Например. Первый этап президентских выборов 1996 года изначально предполагал, что будет и второй этап, так как ни один из претендентов не мог бы сразу набрать 50% + 1 голос. Считая, что кандидатов было 11, включая кандидата «против всех» и действующего президента, который именно в силу данного обстоятельства имел возможность занимать любую из последующих «нулевой» рейтинговых позиций на правах участника в выборах в качестве кандидата, получаем исходную вероятность  $p \approx 0.091$ , и искомые переходные вероятности:

$$P_{00} \approx 0.3505, P_{01} \approx 0.3855, P_{02} \approx 0.1928, P_{03} \approx 0.0578, P_{04} \approx 0.0116, \\ P_{05} \approx 0.0016, P_{06} \approx P_{07} \approx P_{08} \approx P_{09} \approx P_{010} \approx 0.$$

Это практически совпадает с реальным распределением голосов, поскольку процентная ( $P_{ij} \times 100\%$ ) оценка вероятности есть в то же время прогноз результатов выборов в процентном же выражении вне зависимости от численности электората и количества участвовавших в выборах.

Данный пример, кстати, показывает, что претенденты, находящиеся за пределом «пятой» рейтинговой позиции, имеют шансы перемещения в президентское кресло, близкие к нулю, что согласуется со сделанным ранее выводом о поглощающем состоянии при общем числе позиций  $m = 5.56$  [5, с.190-207].

Имея эмпирические данные и предполагая в соответствии с высказанными основаниями, что закон распределения вероятностей задан расчетными значениями  $p_i$  (4-ый столбец Таблицы 1), мы можем определить дисперсию и среднее квадратическое отклонение  $\sigma$ . При этом будем исходить из того, что имеем дело с дискретными, случайными и взаимно независимыми величинами и используем классическую формулу для расчета дисперсии через математическое ожидание, отталкиваясь от эмпирических величин результатов голосования ( $x_i$ ):

$$D(X) = M(X^2) - [M(X)]^2 = \sum x_i^2 p_i - [\sum x_i p_i]^2;$$

$$\sigma = 100\% \times \sqrt{D(x)} = 9,51\%$$

Но можем подойти и с другой стороны. Поскольку речь идет о математической модели, призванной дать прогноз результатов выборов, то допустим, что эмпирическое значение случайных величин – число голосов избирателей, поданных за каждого из кандидатов ( $x_i$ ) – нам неизвестно, и будем исходить только из расчетных величин, считая, что  $x_i = p_i$ . В этом случае

$$D(X) = M(X^2) - [M(X)]^2 = \sum x_i^2 p_i - [\sum x_i p_i]^2 = \sum p_i^3 - (\sum p_i^2)^2;$$

$$\sigma = 100\% \times \sqrt{D(X)} = 10,15\%.$$

Как видим, разница между двумя вариантами расчета дисперсии не столь существенна, и теоретическое значение среднего квадратического отклонения смело можно использовать для прогностического моделирования, получая практически 100%-ную надежность. Однако с увеличением надежности точность прогноза в этом случае страдает, так как разброс ожидаемых значений величин оказывается очень большим, но не большим их эмпирических значений. В связи с этим примем без доказательств наличие интервала наиболее вероятных значений эмпирических величин  $x_i$

в пределах  $x_i = p_i \pm \frac{\sigma}{\sqrt{m}}$ . В последнем случае мы, конечно, сужаем интервал, увеличивая точность прогноза в ущерб его надежности.

Таблица 1

Эмпирические данные и прогноз результатов президентских выборов 1996 г.

Рейтинг	Кандидаты	Результаты голосования в (x <sub>i</sub> ) %	Прогноз в (p <sub>i</sub> ) %	Отклонение $\sigma = \pm 10,15\%$	Наиболее вероятное отклонение $\frac{\sigma}{\sqrt{5}} = \pm 4,54\%$
0-позиция	Ельцин	35,28	35,05	от 24,90 до 45,20	$P_{00}+P_{05} \pm \frac{\sigma}{\sqrt{m}} = 35,21 \pm 3,06\%$ от 30,67 до 39,75
1	Зюганов	32,03	38,55	от 28,40 до 48,70	от 34,01 до 43,09
2	Лебедь	14,52	19,28	от 9,13 до 29,43	от 14,74 до 23,82
3	Явлинский	7,34	5,78	от 0 до 15,93	от 1,24 до 10,32
4	Жириновский	5,7	1,16	от 0 до 11,31	от 0 до 5,70
5	Ельцин		0,16		
6	Против всех	1,54	0,016	от 0 до 10,31	от 0 до 4,56
7	Федоров	0,92	» 0	от 0 до 10,15	от 0 до 4,54
8	Горбачев	0,51	» 0		
9	Шаккум	0,37	» 0		
10	Власов	0,20	» 0		
11	Брынцалов	0,16	» 0		

Эмпирические данные множества избирательных кампаний, не только президентских и не только в России, позволяют утверждать, что эта величина является наиболее часто встречающимся отклонением эмпирических данных от расчетного значения числа голосов в их процентном представлении, поданных за того или другого кандидата. Причем, как правило, она характерна для  $x_i$ , соответствующего рейтинговой позиции ниже 2-ой. С другой стороны, именно в силу малой вероятности победы на выборах в отношении кандидатов, находящихся на данных рейтинговых позициях, в наименьшей степени вероятно осуществление подтасовок и искажений результатов голосования. Это позволяет утверждать, что в случае отклонения значения эмпирической величины  $x_i$  от расчетного значения  $p_i$  (обе величины представляются или в процентах или в долях от «1») более чем

на  $\frac{\sigma}{\sqrt{5}}$ , мы вправе ставить вопрос о сомнении в достоверности результатов

голосования в отношении данного кандидата. Если величина  $x_i > p_i + \frac{\sigma}{\sqrt{5}}$ , мы можем подозревать наличие нарушений в пользу данного кандидата.

Если величина  $x_i < p_i - \frac{\sigma}{\sqrt{5}}$ , мы можем ставить вопрос о нарушениях, совершенных против данного кандидата. Так, в рассмотренном случае выборов 1996 года, когда отмечались массовые нарушения в ходе избирательной кампании, они, очевидно, были направлены против двух кандидатов, занимавших первую и вторую рейтинговые позиции — Зюганова и Лебедева.

Характерно, что на втором этапе выборов Б.Н.Ельцин прибавил к своему «административному ресурсу» большую часть электората второй, третьей, четвертой и пятой позиций, голосовавшую не «за», а «против» самой возможности коммунистической реставрации, получив окончательно 53,82% при минимально возможных  $P_{00} + P_{02} + P_{03} + P_{04} + P_{05} - \sigma = 51,11\%$  (или при другом варианте без голосов яблочников:  $P_{00} + P_{02} + P_{04} + P_{05} -$

$\frac{\sigma}{\sqrt{5}} = 51,11\%$ ). Его главный конкурент, Г.А.Зюганов, сохранив первое место в рейтинге, проиграл выборы (40,31%).

Впрочем, по большому счету, выход за интервал наиболее вероятных значений процента голосов, поданных за того или иного кандидата, означает, что в действие демократических норм, процедур и принципов осуществляется вмешательство сторонних факторов, отнюдь не обязательно в виде откровенного искажения воли избирателей. Сторонние факторы могут влиять на свободный выбор даже не только посредством прямых угроз, давления и запугивания электората. Это может быть весьма тонкая и расчетливая психологическая игра общественным мнением, манипулирование сознанием людей, искажающее их собственную волю и подменяющее их истинные ценности и интересы скорректированными установками. Формально демократические процедуры могут и не нарушаться, но по существу вести к фальсификации самой природы демократии.

На президентских выборах 2000 года было двенадцать кандидатов, включая кандидата «против всех». В.В.Путин к этому времени уже имел имидж известного политика, поскольку занимал пост премьер-министра, являлся официальным преемником Б.Н.Ельцина и исполнял обязанности президента страны, впрочем, в общественном мнении он занимал тогда вторую рейтинговую позицию, уступая Г.А.Зюганову в популярности. Поэтому прогноз тогда строился по «нулевой» строке матрицы с объединением  $P_{00} + P_{02}$ .

Таблица 2

**Эмпирические данные и прогноз результатов президентских выборов  
2000 г.**

Рейтинг	Кандидаты	Результаты голосования (xi) в %	Прогноз (pi) в %	Отклонение $\sigma = \pm 10,16\%$	Наиболее вероятное отклонение $\frac{\sigma}{\sqrt{5}} = \pm 4,54\%$
0-позиция	Путин	52,99	35,2	от 44,24 до 64,56	$P_{00} + P_{02} \pm \frac{\sigma}{\sqrt{m}} =$ <b>54,40 ± 4,54%</b> от 49,86 до 58,94
1	Зюганов	29,24	38,40	от 28,24 до 48,56	от 33,86 до 42,94
2	Путин		19,20		
3	Явлинский	5,8	5,82	от 0 до 15,98	от 1,28 до 10,36
4	Тулеев	2,98	1,19	от 0 до 11,35	от 0 до 5,73
5	Жириновский	2,7	0,16	от 0 до 10,32	от 0 до 4,7
6	Против всех	1,88	0,017	от 0 до 10,18	от 0 до 4,56
7	Титов	1,47	» 0	от 0 до 10,16	от 0 до 4,54
8	Памфилова	1,01	» 0		
9	Говорухин	0,44	» 0		
10	Скуратов	0,43	» 0		
11	Подберезкин	0,13	» 0		
12	Джбраилов	0,10	» 0		

В одном случае в данной таблице  $x_i$  лежит не в интервале  $p_i \pm \frac{\sigma}{\sqrt{5}}$  — это первая рейтинговая позиция, в отношении к которой мы можем выразить сомнение о достоверности результатов, однако мы не можем сомневаться в самих результатах президентских выборов.

На президентских выборах 2004 года В.В.Путин переизбирался на второй срок, и прогноз результатов снова строился по «нулевой» строке матрицы. Кандидатов тогда было семь, если считать также и кандидата «против всех».

Таблица 3

**Эмпирические данные и прогноз результатов президентских выборов  
2004 г.**

Рейтинг	Кандидаты	Результаты голосования (xi) в %	Прогноз (pi) в %	Отклонение $\sigma = \pm 10,22\%$	Наиболее вероятное отклонение $\frac{\sigma}{\sqrt{5}} = \pm 4,57\%$
0-позиция	Путин	71,31	33,99	от 63,43 до 83,87 $P_{00}+P_{01} = 73,65$	$P_{00}+P_{01} \pm \frac{\sigma}{\sqrt{m}} =$ $73,65 \pm 4,57\%$ от 69,08 до 78,22
1	Путин		39,66		
2	Харитонов	13,69	19,83	от 9,61 до 30,5	от 15,26 до 24,4
3	Глазьев	4,1	5,51	от 0 до 15,73	от 0,94 до 10,08
4	Хакамада	3,84	0,92	от 0 до 11,14	от 0 до 5,49
5	Против всех	3,45	0,09	от 0 до 10,31	от 0 до 4,66
6	Малышкин	2,02	0,018	от 0 до 10,24	от 0 до 4,59
7	Миронов	0,75	» 0	от 0 до 10,22	от 0 до 4,57

Процентное количество голосов, полученных стоящим в общественном мнении на второй рейтинговой позиции Харитоновым, находится в пределах теоретического  $\sigma$ -отклонения, но за границами интервала  $p_i \pm$

$\frac{\sigma}{\sqrt{5}}$ , что позволяет подозревать наличие нарушений в ходе избирательной кампании, направленных против второго кандидата. Однако в целом на результатах выборов это никоим образом не отразилось, и, откровенно говоря, если нарушения и имели место, будучи направленными против Харитонova, то смысла в них не было никакого, поскольку результат был предопределен.

В этом плане теоретический интерес представляют президентские выборы 2008 года, когда Д.А.Медведев должен был идти в качестве кандидата, первый раз участвующего в избирательной кампании, а значит, прогноз ее результатов необходимо было бы проводить по «первой» строке матрицы. Однако манипулирование общественным мнением было осуществлено чрезвычайно искусно. Медведев был преподнесен в качестве *alter ego* Путина, и занял не только его «первую» рейтинговую позицию в общественном мнении, но и стал преемником его «административного ресурса», что перевело модель выборов на «нулевую» строку матрицы.



Таблица 4

**Эмпирические данные и прогноз результатов президентских выборов  
2008 г.**

Рейтинг	Кандидаты	Результаты голосования (xi) в %	Прогноз в (pi) %	Отклонение $\sigma = \pm 10,37\%$	Наиболее вероятное отклонение $\frac{\sigma}{\sqrt{4}} = \pm 5,19\%$
0	Медведев	70,28	31,64	от 63,46 до 84,2 $P_{00}+P_{01} = 73,83$	$P_{00}+P_{01} \pm \frac{\sigma}{\sqrt{m}} =$ $73,83 \pm 5,19\%$ от 68,64 до 79,02
1	Медведев		42,19		
2	Зюганов	17,7	21,09	от 7,33 до 28,07	от 15,9 до 26,28
3	Жириновский	9,35	4,69	от 0 до 15,06	от 0 до 9,88
4	Богданов	1,3	0,39	от 0 до 10,76	от 0 до 5,58

Как видим, сами по себе выборы были проведены исключительно «чисто», и их результаты полностью попадают в интервалы, заданные наиболее вероятным отклонением от расчетных значений.

Таблица 5

**Эмпирические данные и прогноз результатов президентских выборов  
2012 г.**

Рейтинг	Кандидаты	Результаты голосования (xi) в %	Прогноз (pi) в %	Отклонение $\sigma = \pm 10,19\%$	Наиболее вероятное отклонение $\frac{\sigma}{\sqrt{5}} = \pm 4,56\%$
0	Путин	63,6	32,77	от 63,54 до 83,92 $P_{00}+P_{01} = 73,73$	$P_{00}+P_{01} \pm \frac{\sigma}{\sqrt{m}}$ $=73,73 \pm 4,56\%$ от 69,17 до 78,29
1	Путин		40,96		$P_{00}+P_{01} \pm \frac{\sigma}{\sqrt{m}}$ $=73,73 \pm 4,56\%$
2	Зюганов	17,18	20,48	от 10,29 до 30,67	от 15,92 до 25,04
3 (5)	Прохоров	7,98	5,12	от 0 до 15,31	от 0,56 до 9,68
4 (3)	Жириновский	6,22	0,64	от 0 до 10,83	от 0 до 5,2
5 (4)	Миронов	3,85	0,032	от 0 до 10,22	от 0 до 4,59

Президентские выборы 2012 года в РФ прогнозировались по сценарию «нулевой» строки. Это объясняется тем, что В.В.Путин обладал имиджем известного политика, уже занимавшего пост президента, а по сути, и не отходил от власти, занимая должность премьер-министра, поэтому «административный ресурс», безусловно, присутствовал и работал на вероятный результат избирательной кампании.

В двух случаях очевиден выход за интервал наиболее вероятного результата голосования:

$$P_{00} + P_{01} - \frac{\sigma}{\sqrt{5}} = 0,6917 > 0,6360 \text{ (63,6\%)}$$

и

$$P_{04} + \frac{\sigma}{\sqrt{5}} = 0,052 < 0,0622 \text{ (6,22\%)}$$

Трудно предположить, чтобы представители «Единой России» на избирательных участках и при работе в избирательных комиссиях осуществляли нарушения «против» своего кандидата. С другой стороны, по предварительным социологическим исследованиям, за исключением первых двух кандидатов, рейтинг остальных в общественном мнении был иным, и Прохоров занимал последнюю рейтинговую позицию, а Жириновский и Миронов делили между собой соответственно третью и четвертую позиции. При строгом следовании рейтинговой схеме социологических исследований прогноз в отношении числа голосов, которые могли быть получены кандидатами от ЛДПР и СП, полностью совпадает с эмпирическим

результатом, но тогда из модели выпадает  $P_{05} + \frac{\sigma}{\sqrt{5}} < 0,0798 \text{ (7,98\%)}$

Таким образом, если в ходе избирательной кампании искажения имели место, то, во всяком случае, они не были направлены в пользу избранного президента, а если и повлияли на результаты голосования, то только в сторону занижения числа полученных им голосов. Но для того, чтобы так изменилась – с пятой на третью – позиция последнего в рейтинговом списке кандидата, необходимо было воздействие достаточно мощного фактора, повлиявшего на значение результирующего фактора общественного мнения. Если расчетное значение числа голосов (в процентах), которое должен был получить М.Прохоров, вычесть из полученных им на выборах 7,98% (7,98 – 0,032 = 7,948%), то приходим именно к тому количеству, которое не добрал В.В.Путин и которое оказалось бы в интервале наиболее вероятных значений (69,17% < 63,6 + 7,948 = 71,548% < 78,29%). Однако процент голосов на грани интервала, заданного средним квадратическим отклонением (63,54% < 63,6%), заставляет задуматься. Вывод же состоит в том, что, хотя на основе биномиальной модели, легитимность победы В.В.Путина не вызывает сомнений, но это Пиррова победа, равная поражению. Это – личная победа В.В.Путина, и поражение «Единой России». Это – отрицание обществом внутренней политики, проводившейся в последние четыре года и не давшей реальных результатов: и тех, которые показывает официальная статистика, и тех, которые благозвучно, но совершенно напрасно, назывались реформами. Голосование на выборах, особенно в сравнении с предыдущими, показало надежду, которую еще питает большая часть электората, на возвращение к взвешенному курсу, имевшему место в предшествующее президентство В.В.Путина.

## Библиографический список

1. Лукичев П.Н. Технология выборов на этапе постфакторной определенности // Четвертые межрегиональные научные чтения по актуальным проблемам социальной истории и социальной работы: тез. докл. и сообщ., Новочеркасск, 28-29 апр. 2003 г. – Ростов н/Д: Пегас, 2003. – С.155-156.
2. Лукичев П.Н. Онтологическое единство динамических и статистических закономерностей // Философия и будущее цивилизации: тез. докл. и выступлений IV Российского философского конгресса (Москва, 24-28 мая 2005 г.): В 5 т. Т.1 – М.: Современные тетради, 2005. – С.564-565.
3. Лукичев П.Н. Фундаментальный алгоритм наследования власти // Лосевские чтения: труды Междунар. науч.-теорет. конф., г. Новочеркасск, май 2006 г. – Новочеркасск: УПЦ «Набла» ЮРГТУ (НПИ), 2006. – С.260-269.
4. Лукичев П.Н. Верификация демократичности выборов // Методология, теория и история социологии: сб. науч. статей. Материалы межрегион. науч. конф. «Методология, теория и история социологии», Ростов-на-Дону, 10 ноября 2010. – Ростов-на-Дону: СКНЦВШ, 2011. – С.281-285.
5. Лукичев П.Н. Общая теория социальной динамики: Основания и начала анализа. – Ростов–на–Дону: СКИАПП, 2002.

Милек О. В., Шмерлинг Д. С., Москва

## Некоторые проблемы математического моделирования неравенства и распределения доходов<sup>1</sup>

---

Тема «распределения доходов населения» сопряжена с проблематиками из сферы социологии и экономики. Говоря о неравенстве распределения доходов, мы имеем в виду и гносеологический аспект справедливости социально-экономического положения граждан, и реальную политику в сфере благосостояния населения конкретной страны. Цель достоверного определения уровня бедности и пути эффективной борьбы со столь злободневным социальным недугом также формулируется социологами-эмпириками. Необходимость определения среднего класса, выявление корреляции неравенства населения в обществе с различными экономическими показателями страны, разработка эффективной налоговой политики — это лишь некоторые содержательные задачи, которые стоят на повестке дня у исследователей.

Экономический рост как основной интегральный показатель улучшения качества жизни населения используется повсеместно. И в измерении эффективности социально-экономической политики, и в планировании мер по внедрению основных механизмов реализации этой политики. При обсуждении вопроса обеспечения экономического роста и повышения благосостояния населения страны зачастую используются понятия «ликвидация бедности», сокращение «дифференциации доходов», формирование устойчивого «среднего класса». Во всех этих разговорах понятие «экономический рост» является ключевой категорией, которая обуславливает остальные параметры социально-экономической структуры. Существует и иной методологический подход к проблеме, который в принципиально ином ракурсе рассматривает причинно-следственную связь «экономического роста» и «социально-экономической дифференциации населения». В данном подходе «расслоение» является основополагающим критерием принципиальной возможности развития экономики страны. Что влечет за

---

<sup>1</sup> Статья подготовлена по результатам научно-исследовательской работы «Системная модернизация экономики России», выполняемой в рамках Тематического плана прикладных исследований, реализуемых ФГОБУ ВПО «Финансовый университет при Правительстве Российской Федерации» в 2012 году в рамках бюджетного финансирования.

собой понимание не просто причинно-следственной связи явлений, а анализ структурной организации и взаимодействия всех процессов социально-экономической жизни в рамках парадигмы системного подхода<sup>1</sup>.

Одним из направлений в рамках данной научно-познавательной проблемы является вопрос математического моделирования неравенства распределения доходов. Огромное количество методов анализа данных в социологии были изобретены для другого рода задач (не социологической направленности). Несомненно, качество и степень эффективности данных методов в социальных науках не оспорима, однако необходимо помнить об ограничениях и рамках возможности анализа. Например, говоря об изучении факторов, влияющих на распределение трудовых доходов, исследователь берет во внимание не только тип распределения этих доходов, но и все ограничения использования регрессий для социальных данных.

Множество проблем всех уровней анализа (микро-, мезо-, макро-) делают эту проблему актуальной как для теоретических изысканий, так и для практической реализации социально-политических и налоговых мер экономической политики.

Системный подход в своей методологической парадигме не предполагает изучать объект с разных точек зрения путем синтеза и анализа отдельных ее частей. В данной методологии предполагается изучать целое как единый организм. Как отмечает один из идеологов системного анализа в России И.Н. Дрогобыцкий «увидеть целое можно только при одновременном понимании структуры, функции и процесса»<sup>2</sup>. Тем самым необходимо осуществить качественный переход к анализу проблемы от кластеризации участников и отраслей, к таким понятиям как структура процесса, функция данного явления, реализация самого процесса, и среда, которая присутствует на данном историческом этапе.

Проблема неравенства в доходах хорошо известна по крайней мере с работы Макса Лоренца (Lorenz M.O., 1905)<sup>3</sup> о кривых рассеяния, см. например Кендалл, Стьюарт (1966), пар. 2.25<sup>4</sup>.

Существует огромная литература по проблеме дифференциации населения, территории, предприятий и т. п. по доходам, богатству, имуществу.

Еще в прошлом веке В. Парето в своих трудах установил тип распределения доходов населения путем обширного анализа статистического материала о доходах по многим странам. Данное распределение именуется «распределением Парето», закон «80:20» был сформулирован как продолжение выявленного классиком правила «20% самых богатых людей концентрируют в своих руках 80% общественного благосостояния». Можно охарактеризовать подход Парето к измерению неравенства доходов как

---

<sup>1</sup> Существует фундаментальное пособие по данной проблематике: Atkinson A.B., Bourguignon F. Eds. Handbook of income distribution, v.1. – Amsterdam e.a.: Elsevier North Holland, 2007. – xix, 918, 38 pp.

<sup>2</sup> Дрогобыцкий И.Н. Системный анализ в экономике.-М.: Финансы и статистика, 2007.-508 с.

<sup>3</sup> Биография Макса Лоренца есть у Маршала, Олкина (1983), см. также Lorenz\_curve.Wikipedia.html

<sup>4</sup> Впрочем, неравенство было описано и в работе В.И. Ленина «Развитие капитализма в России» (1899).



системный подход к проблеме. В рамках системной парадигмы необходимо не просто производить оценку доли населения с величиной ниже прожиточного минимума (абсолютная бедность), или считать долю населения, доход которой ниже среднего дохода по РФ (относительная бедность). А развивать системность измерения перераспределения доходов (лишь элементом которой является прогрессивная шкала налогообложения) и учитывать функции стратификации, искать объективные показатели расслоения общества (например, уровень коэффициентов стратификации: фондов, уровней, Джини, Аткинсона), когда уже не возможно стабильное экономическое развитие.

Профессор А.Ю. Шевяков совместно с А.Я. Кирутой в своих трудах формулируют мысль, которая является ключевой в понимании социально-экономической политики на современном этапе ее развития: обеспечение экономического роста напрямую зависит от уровня социально-экономического неравенства в обществе, которая в свою очередь не может быть эффективно решена никакими другими методами, кроме как эффективными перераспределительными механизмами<sup>1</sup>.

Возможность связать т. н. «избыточную» дифференциацию с уменьшением человеческого потенциала за счет ухудшения всякого рода социальных лифтов весьма плодотворна. Большое исследование в этом направлении предпринял А.Я. Кирута<sup>2</sup>. Дело в том, что ещё не изжит миф о полезности значительной дифференциации, как стимулирующего развитие фактора. На самом деле стимулировать может лишь не слишком большая дифференциация доходов — возможно не выше «линейной», см. ниже.

Исследование и разъяснение этого вопроса широким кругам представителей бизнеса, органам государственной власти представляет собой важнейшую задачу современного научного сообщества. События последнего времени под лозунгом «Захвати Wall Street» (осень 2011) подтверждает мысль о том, что большое неравенство не одобряют не только теоретики-экономисты, социологи, но и широкие круги населения.

И, несмотря на тот факт, что в США проблематика неравенства доходов стоит на повестке дня уже не первый год, мы можем констатировать очевидную злободневность этой проблемы и в обществе свобод. На примере данных исследования доходов домохозяйств мы продемонстрируем проблемы математического характера изучения распределения доходов.

Впервые построив гистограмму распределения домохозяйств США по доходу мы обнаружили для себя загадку. Распределение доходов, что и предполагалось, визуально более соответствует предположению Парето

<sup>1</sup> Шевяков А.Ю. Экономическое неравенство: тормоз демографического роста // Журнал новой экономической ассоциации. – М. 2011. № 9. С. 197 – 201. Шевяков А.Ю., Кирута А.Я. Неравенство, экономический рост и демография: неисследованные взаимосвязи. – М.: М-студия, 2009.

<sup>2</sup> Кирута А.Я. Неравенство, бедность и социально-гуманитарные факторы модернизации. Август 2011. Рукопись; Кирута А.Я. Влияние неравенства на качество человеческого потенциала в России // Вестник Института социологии РАН, 2011, №3, с.67-87.



(Парето на «правом хвосте» и логлогистическому в средней части распределения), нежели статистическому распределению Гаусса. Эмпирическое распределение имеет явный вид «двугорбого», что неправдоподобно для распределения доходов. Почему так случилось – загадка.

Дело в том, что “математическая” загадка была на самом деле социального характера. В промежутке от \$150 000 до \$200 000 сконцентрировалось 3 595 домохозяйств, а в соседних интервалах значительно меньше. Данный феномен полностью объясняется налоговым законодательством страны, где пропорциональная система стимулирует концентрировать доходы ниже границы в \$200 000. Несколько выше \$179 000 (эта граница плавающая, и в 2012 году составила уже \$198 000) налог на домохозяйство повышается с 28% до 33%<sup>1</sup>. Это факт демонстрирует нам, что даже в системе жесткого налогового контроля, где проблема неравенства позиционируется как первоочередная цель достижения общества благосостояния, наблюдаются проблемы сокрытия доходов.

Содержательно при математической экстраполяции доходов правого хвоста необходимо учитывать данную проблему «скрытия доходов». Одним из вариантов является равномерное разнесение этой группы на следующие интервалы доходов (или «сглаживание правого хвоста» распределением Парето, а середину логлогистическим).

Таким образом, даже на данных генеральной совокупности о доходах населения мы можем констатировать сомнительность достоверного изучения «правого хвоста» данных, а значит и оценки дифференциации доходов населения.

Здесь мы сталкиваемся с проблемой измерения неравенства и интерпретацией полученных результатов измерения. Что касается измерения неравенства, то существуют традиционные меры Gini, Pietra и др.<sup>2</sup> Однако, распространённой интерпретации общепринятых коэффициентов дифференциации доходов нам не известны. В качестве варианта таковой, предлагаем следующую модель, описывающую распределение доходов в небольшом коллективе. Эта модель даёт представление о возможных внутренних механизмах income distribution.

Весьма распространён метод измерения неравенства с помощью коэффициента (индекса) Джини (Corrado Gini, 1912). Пусть,

$$\Delta_1 = \frac{1}{N(N-1)} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} |x_j - x_k| f(x_j) f(x_k) \quad (1)$$

средняя разность Джини (для дискретного случая),  
где  $x_1, x_2, \dots$  - величина доходов,  $f(x_1), f(x_2)$  – вероятность (или частота по выборке) людей с доходами  $x_1, x_2, \dots$  соответственно.

<sup>1</sup> URL: [http://en.wikipedia.org/wiki/Income\\_tax\\_in\\_the\\_United\\_States#Year\\_2012\\_income\\_brackets\\_and\\_tax\\_rates](http://en.wikipedia.org/wiki/Income_tax_in_the_United_States#Year_2012_income_brackets_and_tax_rates).

<sup>2</sup> См. Kotz, Kleiber (2003) §2.1, §2.2 P. 35-39.

Величину  $\Delta 1$  обычно нормируют так, чтобы  $\Delta 1^* = \Delta 1 / \Delta 1^{\max} \in [0,1]$ , при этом чем больше (и ближе к 1), тем значительней неравенство населения.

Кстати, площадь над кривой рассеяния (Лоренца) и под диагональю (рис. 2.2 у Кендалла, Стьюарта, 1966, с.75) равна  $\Delta 1/4 \mu 1$ , где  $\mu 1 =$

$$\int_{-\infty}^{\infty} (x-a)f(x)dx,$$

$f(x)$  – плотность распределения, обычно организуют  $a=0$ , (пар.2,3 Кендалл, Стьюарт, 1966).

Собственно кривая рассеяния есть

$$\phi(x) = \frac{1}{\mu_1} \int_{-\infty}^x f(x)dx \quad (2)$$

т. е. «неполный первый момент распределения» (там же, с. 75-77, (2.31)).

Вычисление оценки средней разности (там же, пар.2.26, (2.33), (2.34) и пример).

Коэффициент Джини вычисляется и публикуется для большинства стран уже десятки лет, о методах оценивания, в т. ч. с группированными данными см. Gastwirth (1972), Modorres, Gastwirth (2006)<sup>1</sup>.

Некоторую сводку значений нормированного индекса Джини (по данным ЦРУ), к примеру в Норвегии 0,25 (2008 год), во Франции 0,32 (2008 год), в России 0,423 (2008 год), в Нигерии 0,437 (2003 год), в США 0,45 (2007 год), в Мексике 0,482 (2008 год), на Гаити 0,538 (2001 год), в Сьерра-Леоне 0,629 (1989 год), в Южной Африке 0,65 (2005 год), в Намибии 0,707 (2003 год)<sup>2</sup>, а позже до 0,75.

Читатель-экономист, может быть, уже привык к таким данным, но насколько обществом понят смысл значений коэффициента Джини?

Существует обширная литература о вреде высокого ( $> 0,3$ ) коэффициента Джини, см., например, Справочник по распределению дохода (Atkinson, Bourguignon, 2000). Граница 0,3 или 1/3 (0,33) достаточно произвольна и представляет что-то около среднеевропейского среднего коэффициента Джини.

Было бы полезно поискать за значениями нормированного коэффициента  $G'$ ,  $0 \leq G' \leq 1$ , какой-нибудь «физический» смысл (то бишь, экономический!).

Рассмотрим следующую модель. Пусть  $x_i$  – доход лиц, относящихся к  $i$ -му уровню иерархии применительно к компании, населению территории и т. п.,  $i = 1, \dots, n$ . Модель Р (“тарифная сетка”) такова.

<sup>1</sup> См. также Morgan J. The Anatomy of Income Distribution // Rev.Econ.Statist. (MITpress), 1962, v.44, N 3, p. 270-283

<sup>2</sup> В Боливии индекс Джини 0,592, а децильное отношение ужасно велико – 168,1 (!)

Доход на  $i$ -м уровне ( $i = 1$  – лица с наименьшим, а  $i = n$  – с наибольшими доходами) равен  $x_i = ki^m$ ,  $m = 1, 2, 3, \dots, k > 0$ .

Теорема. Коэффициент (индекс) Джини  $G'_m(n)$  для модели  $P$  равен асимптотически при  $n \rightarrow \infty$

$$G'_m(n) = \frac{m}{m+2} \quad (1a)$$

Набросок доказательства

$$G'(n) = \frac{\sigma(n)}{\max_x \sigma(n)}, \quad (2a)$$

где максимум берется по всем возможным  $\{x_{(1)}, x_{(2)}, \dots, x_{(n)}\}$ , таким, что

$$\sum_{1 \leq i \leq n} x_{(i)} = C(n) \quad (3)$$

$$\sigma(n) = \frac{2\sqrt{\pi}}{n(n-1)} \sum_{1 \leq i \leq n} \left(i - \frac{n+1}{n}\right) x_{(i)}, \quad (4)$$

$X_{(i)}$  –  $i$ -ая порядковая статистика<sup>1</sup>

$$G(n) = \frac{1}{n(n-1)} \sum_{1 \leq i, j \leq n} |x_i - x_j|. \quad (6)$$

Используя другую форму  $\sigma(n) = \sigma$

$$\sigma = \frac{\tilde{k} 2^* \sqrt{\pi}}{n(n-1)} \left\{ \sum_1^n i x_{(i)} - \frac{n+1}{n} \sum_1^n x_{(i)} \right\} \quad (7)$$

И ВЫЧИСЛЯЯ

<sup>1</sup> См. Дэйвид (1978), (7.4.1.) и весь §7.4., с. 187–189 упражнение 7.4.1., с. 214, §9.6, (9.6.1.), где обсуж-

дается асимптотическая нормальность “ $\sigma$ ”. При этом  $E\sigma = 2 \sqrt{\pi} \int_{-\infty}^{\infty} x \left[ P(x) - \frac{1}{2} \right] dP(x)$ , “ $\sigma$ ” – несмещенная оценка для  $\sigma$  в случае нормальных выборок,  $P(x)$  – функция распределения.

$$\max_x \sigma = \frac{\sqrt{\pi} * \tilde{K}}{n} S_m(n) \quad (8)$$

где<sup>1</sup>

$$S_m(n) = \sum_{k=0}^{n-1} k^m$$

сумма целых чисел в степени  $m=1, 2, 3, \dots, k=0, 1, 2, \dots, n-1$ , можно получить выражение для (3).

Именно, при  $\tilde{S}_m(n) = S_m(n) + n^m$

$$\sigma = \frac{2\tilde{K}\sqrt{\pi}}{n(n-1)} \left\{ \tilde{S}_{m+1}(n) - \frac{n+1}{n} \tilde{S}_m(n) \right\} / \frac{\tilde{K}\sqrt{\pi}}{n} \tilde{S}_m(n) \quad (9)$$

Откуда (здесь  $m$  напоминает о степени многочлена)

$$G'_m(n) = \frac{2}{n-1} \left\{ \frac{\tilde{S}_{m+1}(n)}{\tilde{S}_m(n)} - \frac{n+1}{2} \right\} \quad (10)$$

Теперь нам понадобятся выражения  $Si(n)$  в удобной форме из великопейной книги Грехэм, Кнут, Паташник (2009), (61.78),

$$\tilde{S}_m(n) = \frac{1}{m+1} \sum_{0 \leq k \leq n} \binom{m+1}{k} B_k n^{m+1-k} \quad (11)$$

где  $B_k, k=0, 1, 2, \dots$ , числа Бернулли (Якова), именно

Таблица 1

k	1	2	3	4	5	6	7	8	9	10	11	12	...
$B_k$	1	$-\frac{1}{2}$	0	$-\frac{1}{6}$	0	$\frac{1}{42}$	0	$-\frac{1}{42}$	0	$\frac{5}{66}$	0	$-\frac{691}{2730}$	...

<sup>1</sup> Так у Грехэма, Кнута, Паташника (2009).

Формулы  $S_m(n)$ ,  $m = 0, 1, 2, \dots, 10$  см. в упомянутой книге с. 314.

Из (10) легко получается при  $n \rightarrow \infty$

$$G'_m(n) \cong \frac{2}{n-1} \left\{ \frac{m+1}{m+2} n - \frac{n+1}{2} \right\} \approx \frac{m}{m+2}, QED.$$

Приведем таблицу 2 для  $G'_m(n)$ <sup>1</sup>

Таблица 2

M	1	2	3	4	5	6	7	8	9	10	...
$G'_m(n)$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{3}{5}$	$\frac{2}{3}$	$\frac{5}{7}$	$\frac{3}{4}$	$\frac{7}{9}$	$\frac{4}{5}$	$\frac{9}{1}$	$\frac{5}{6}$	...

Вернемся к началу изложения. Если в модели Р брать  $k = 1$ , то  $G' = 0,25$  (Норвегия) означает  $m < 1$  (дробное число),  $G' = 0,327$  (Франция) означает  $m \approx 1$ ,  $G' = 0,423$  (Россия) означает, что  $1 < m < 2$ ,  $G' = 0,482$  (Мексика) дает  $m \approx 2$ ,  $G' = 0,538$  (Гаити) –  $2 < m < 3$ , Сьерра-Леоне с  $0,629$  приводит к  $3 < m < 4$ , ЮАР с  $0,65$  дает  $m$  несколько больше, а Намибия, где по разным данным  $0,707 < G' < 0,75$ , может претендовать на  $5 < m < 6$ .

Попробуем интерпретировать наши результаты.

1. Компании, сообщества, регионы, страны можно (условно) делить на линейные ( $m=1$ ), квадратичные ( $m=2$ ), кубические ( $m=3$ ), «тетричные» ( $m=4$ ), «пентальные» ( $m=5$ ), «гексальные» ( $m=6$ ) и т. д. Например, для Москвы, где  $G'$  в разные годы доходил до  $0,62$  можно предполагать кубический тип распределения доходов. При этом надо иметь ввиду, что реальные значения  $G'$  в Москве могут быть, по мнению многих, если не большинства специалистов,  $0,60-0,70$ .

2. Поскольку мы использовали свойства сумм степеней целых чисел, то можно вспомнить Леопольда Кронекера (1823-1891), который не зря говорил, что целые числа придумал Бог, а остальное – люди.

3. Что касается российских компаний, то данные отчетов ряда крупных компаний, которые регулярно обсуждаются газетой «Ведомости», показывают следующее<sup>2</sup>. Вознаграждения (бонусы) членов правления (не считая заработной платы) в 1-3 млн. долларов в год при зарплате на уровне  $i=1$  в 1.5-3.0 тыс. долларов и (условно) указывают на величину  $2 < m < 3$ . Возможно, что  $m$  достигает 4, т. к. за 2010г. зарплата Генерального директора одной из самых крупных компаний России м.б. равна \$25 млн. в год, см. [http://www.rfcor.ru/print/news\\_rfc\\_987htm](http://www.rfcor.ru/print/news_rfc_987htm).

<sup>1</sup> При  $m=1$  выражении для индекса Джини точное, при всех  $n = 1, 2, 3, \dots$

<sup>2</sup> См. магистерскую диссертацию: Милек О. Изучение распределения дохода с помощью распределения с тяжелыми хвостами. – М., ГУ-ВШЭ, 2010.

Степень  $m$  “градуирует” (термин из физики и инженерного дела) “шкалу Джини”.

Здесь возникает вопрос о разумности механизмов распределения доходов при различных  $G^*$ , т. е. при соответствующих величинах  $m$ , см. Atkinson, Bourguignon (2000), Chacravarty (1990), Foster, Sen (1997) и т. п.

А.Я. Кирута в своём любезно написанном 8-страничном комментарии, полученном автором 03 мая 2011 г. в 14:34 предпринял глубокий анализ рассмотренной здесь задачи.

Заметим, что можно рассматривать и вещественные (по интерполяции) значения

$$m=2G^*/(1-G^*).$$

Здесь  $G^*$  -нормированное значение индекса Джини.

Для Москвы за 2009 г.  $G^*=0.521$ , что даёт величину  $m=2.742$ .

Для Норвегии (2008)  $G^*=0.25$  и  $m=0.667$ .

Здесь требуется обсуждение.

Можно увязать обсуждаемую модель с традиционными статистическими распределениями. Для распределения Парето с таким же  $G^*$ , как в нашей модели

$P$ , лишь степень  $m < 1$  обеспечивает конечную дисперсию, а для лог-логистического распределения для того же требуется  $m < 2$ . Что касается лог-нормального распределения, то дисперсия не стремится к бесконечности, а лишь медленно растёт при росте  $m$ . Заметим, что распределения Парето и лог-логистическое хорошо описывают правый (верхний) хвост распределения доходов, а лог-нормальное хорошо описывает не слишком большие доходы, но плохо описывает правый хвост.

Интерпретация может быть следующей: при высокой степени неравенства в модели  $P$  малая (богатая) часть общества стремится увеличить свои доходы, так что верхние хвосты распределения утяжеляются и дисперсия стремится к бесконечности. В тоже время средняя по доходам часть общества медленно реагирует на рост степени модели  $m$ .

Недостижимым образцом для подражания авторам служил доклад Владимира Игоревича Арнольда “Жесткие” и “мягкие” математические модели, сделанный им 25 сентября 1997 г. в Администрации Президента Российской Федерации на семинаре “Аналитика в государственных учреждениях”. Руководили семинаром А.Н.Райков, Г.А.Сатаров, Д.С.Шмерлинг.

Авторы благодарят В.И.Арнольда, А.Я.Кируту, Я.Ю.Никитина, О.А.Оберемко, А.И.Орлова, Ю.Н.Толстову, Ю.Н. Тюрину, В.В.Ульянова за содействие и обсуждение.



## Библиографический список

1. Грэхем Р., Кнут Д., Паташник О. Конкретная математика. Основание информатики. 3-е изд.: Пер. с англ. — М.: МИР, Бином. Лабор. Знания, 2009. — 703 с.
2. Дейвид Г. Порядковые статистики: Пер. с англ. — М.: Наука, 1978. — 336 с., §§7.4, с. 187, упр. 7.4.1, с. 214, §9.6, (9.6.1.).
3. Дрогобыцкий И.Н. Системный анализ в экономике.-М.: Финансы и статистика,2007.-508 с.
4. Кендалл М Дж., Стьюарт А. Теория распределения: Пер. с англ. — М.: Наука, 1966. — 588 с., §2.21-2.26
5. Кирута А.Я. Неравенство, бедность и социально-гуманитарные факторы модернизации. Август 2011. Рукопись; Кирута А.Я. Влияние неравенства на качество человеческого потенциала в России //Вестник Института социологии РАН,2011,№3, с.67-87
6. Маршалл А., Олкин И. Неравенства. Теория мажоризации и ее приложения: Пер. с англ. — М.: Мир, 1983. — 576 с.
7. Шевяков А.Ю. Экономическое неравенство: тормоз демографического роста // Журнал новой экономической ассоциации. — М. 2011. № 9. С. 197 — 201.
8. Шевяков А.Ю., Кирута А.Я. Неравенство, экономический рост и демография: неисследованные взаимосвязи. — М.: М-студия, 2009.
9. Atkinson A.B., Bourguignon F., ed. Handbook of Income Distribution, vol. I.- S.l.: Elsevier, 2000.
10. Chakravarty S.R. Ethical Social Index Numbers. — N.Y.:Springer-Verlag, 1990.
11. David H.A. Gini's Mean Difference Rediscovered.// Biometrika, 1968, v. 55, p. 573-575.
12. Dikhanov Y. Decomposition of Inequality Based on Incomplete Information. — Wasington, D.C.: The World Bank, 1996. — 26 pp.(A contrib. paper to the LARIW 24<sup>th</sup> Gen.Conf. Lillehammer, Norway, aug.18-24, 1996. Sitereseoucesworldbank.org/ .../13279\_Decomposition\_of-inequality based.
13. Foster J.F., Sen Amartya. On Economic Inequality, exp. ed. With Substantial Annex. Oxford: Oxford Univ. Press, 1997. — 280 pp.
14. Gastwirth Y.L. The Estimation of the Lorenz Curve and Gini Index // Rev. Econ. Statistics, 1972, v. 52, #3, p. 306-316. doi: 10.2307/1937992.
15. Gini\_coefficient.wikipedia.htm.

16. Gini, Corrado. Variabilità e mutabilità, contributo allo studio delle distribuzioni e relazioni statistiche // Studi Economico-Guirdici della R. Università di Cagliari, Repr. In Memorie di Metodologia Statistica / Piretti E, Salvemini T. ed. – Rome: Libreria Eredi Virgilio Veschi, 1955.(Воспроизведена знаменитая работа 1912 года).
17. Lorenz M.D. Methods of Measuring the Concentration of Wealth // Publ. Amer. Statist. Ass., 1905, v. 9, #70, p. 209-219.
18. Moderres R., Gastwirth J.L. A Cautionary Note on Estimating the Standard Error of the Gini Index of Inequality // Oxford Bull. Econ. Statist., 2006, v. 68, p. 385-390. doi: 10.1111/j.1468-0084.2006.00167 [http://en.wikipedia.org/wiki/list\\_of\\_countries\\_by\\_income\\_equality](http://en.wikipedia.org/wiki/list_of_countries_by_income_equality)
19. Kleiber C., Kotz S. Statistical Size Distributions in Economics and Actuarial Sciences. – Hoboken, NJ: Wiley, 2003. – 332p.
20. Dagum C. Income inequality measures. – In: Encyclopedia of Statistical Sciences. 16 vols. N.L. Johnson, S. Kotz ed-in-chiefs. – N.Y. e.a.: Wiley, 2006. – 9686pp., p. 3387-3405.

Никитин С. А., Москва

## Имитационное моделирование поведения членов первобытного общества с учетом альтруизма<sup>1</sup>

---

### Аннотация

В работе изложены первые результаты использования имитационной модели взаимоотношения между членами первобытного племени в разнообразных условиях внешней среды. Установлено, что в рамках модели выживаемость племени определяется, в частности, внешней средой и альтруизмом его членов.

**Ключевые слова:** альтруизм, имитационная модель, внешняя среда, взаимодействие индивидов, выживаемость племени

### Введение

В лаборатории математической социологии ЦЭМИ РАН ведутся исследования по моделированию социально-этических аспектов в экономических системах. Одним из направлений исследований является имитационное моделирование эволюции общества на ранних стадиях его развития с учетом альтруизма. В частности используются эволюционный подход Г. Спенсера и генетический подход В. Эфроимсона, которые объединены в данной работе.

Использованная в работе программа была разработана С. Лушиным в программной среде Borland C++ Builder 6.0.

Целью имитационной модели был анализ выживаемости первобытного племени в зависимости от различных внешних условий и внутренних характеристик каждого члена племени, в частности, альтруизма.

---

<sup>1</sup> Работа выполнена при поддержке РФФИ, грант №10-06-00362.

### Описание внешней среды

Племя располагается на территории, представляющей собой квадрат, перемещаясь внутри которого каждый житель ищет пропитание. Функция распределения пищи по территории задается формулой  $z = a - b(x - c)^2 - d(y - e)^2$ , где  $x, y$  – координаты точки,  $z$  – количество пищи в данной точке. Параметры  $a, b, c, d, e$  – могут быть различными и устанавливаются пользователем до запуска очередного цикла программы. Также в данной местности случайным образом без участия пользователя выделяются четыре клетки, играющие роль опасностей с тремя степенями вредоносности также задаваемых случайно, попадая в которые человек теряет часть своего здоровья, которая устанавливается пользователем.

### Описание племени

Племя может быть охарактеризовано следующими параметрами, являющихся переменными, значение которых устанавливается до запуска программы:

- начальное число людей;
- продолжительность жизни (является равномерно распределенной случайной величиной с задаваемым пользователем средним значением и отклонением от него)
- доля женщин;
- доля «гена альтруизма» (обуславливает предрасположенность членов племени к альтруистическому поведению, например, помощи другому при его столкновении с опасностью);
- доля «гена агрессии» (обуславливает предрасположенность эгоистического поведения членов племени, например, при борьбе встретившихся неженатых мужчин племени за женщину или еду);
- радиус взаимодействия (максимальное расстояние между двумя членами племени, когда между ними происходит взаимодействие – конфликт между мужчинами или создание семьи между холостым мужчиной и незамужней женщиной);
- число тактов невзаимодействия (время, когда между членами племени, находящимися на расстоянии, не превосходящем радиус взаимодействия, ничего не происходит);
- скорость перемещения (количество клеток, которые проходит житель за единицу времени);

- радиус слышимости (максимальное расстояние, с которого другие члены племени услышат попавшего в опасность соплеменника);
- радиус видимости (максимальное расстояние, с которого другие члены племени увидят попавшего в опасность соплеменника);
- коэффициент подражания плохому (сила влияния, с которой член племени становится эгоистичнее и агрессивнее, если встретит соплеменника с более низким уровнем альтруизма);
- коэффициент подражания хорошему (сила влияния, с которой член племени становится альтруистичнее и добрее, если встретит соплеменника с более высоким уровнем альтруизма);
- коэффициент подражания Идеалу (степень желания каждого члена племени подражать соплеменнику с уровнем альтруизма, считаемым в данном племени наилучшим (идеальным));
- альтруизм Идеала;
- прожиточный минимум (минимальное количество пищи для сохранения собственной силы члена племени, в противном случае сила уменьшается на 10 процентов);
- уровень интеллекта (обуславливает способность найти место с наибольшим количеством пищи с определенной вероятностью, задаваемой пользователем);
- максимальное количество детей в семье;
- вероятность появления ребенка;
- возраст выхода из семьи;
- вероятность передачи «гена альтруизма»

Также в племени присутствует вождь, в задачи которого входит наказание холостых мужчин племени за драку друг с другом, а также распределение пищи из общего «котла» между всеми членами поровну. Каждый свободный член племени складывает в общий котел половину добытого пропитания.

### Поведение членов племени

Если встречаются двое холостых мужчин, то между ними происходит конфликт из-за еды. Поскольку сила каждого члена племени, его жестокость и доброта являются равномерными случайными величинами, то исходы конфликтов могут быть разными. А именно, если один из участников «плохой» и одновременно сильнее, то драка закончится убийством слабого, и сильный заберет всю добытую им еду. Если же более сильный окажется «хорошим», то убийства не произойдет, но еду у слабого он все равно отберет.

Поведение человека, попавшего в опасное место («опасную» клетку) может быть охарактеризовано двояко: если он видит, что у него хватает сил справиться с опасностью, то он действует, не призывая никого на помощь, и при этом теряет часть здоровья, в противном случае он зовет соплеменников на помощь. Значения т. н. функции преодоления, т. е. необходимого количества силы для того, чтобы справиться с опасностью, устанавливается пользователем. Если никого не оказывается в зоне слышимости или видимости, то соплеменник, столкнувшийся с опасностью, погибает.

Выходными данными программы являются следующие:

- число побед (фактически число драк за весь период существования племени, закончившихся убийством одного из участников);
- конечное число людей в племени (примечание: промежуток времени, на котором рассматривается жизнь племени, может быть установлен пользователем произвольно также как и длина одного такта, выраженного в миллисекундах);
- родившиеся дети (общее количество детей, родившихся за период жизни племени);
- общий котёл (количество пищи в котле, собранное за весь рассматриваемый период);
- наказанные (общее число наказанных вождем соплеменников за драки в течение заданного периода жизни племени).

### Описание интерфейса

Интерфейс состоит из нескольких окон: одного главного и шести вспомогательных. В главном окне находятся кнопки вызова вспомогательных окон, в которые вводятся необходимые начальные условия, кнопки запуска и приостановки программы. Кроме этого, в главном окне находится графическое представление территории, на котором живет племя — квадрат, с перемещающимися внутри него кружками разного цвета, символизирующие членов племени. Красные кружки символизируют незамужних женщин, синие — холостых мужчин, зеленые — членов образовавшихся семей (родителей и детей), один лиловый кружок символизирует вождя племени. Также выводятся графики средней силы, среднего интеллекта племени, а также график влияния общества на индивида (фактически график среднего альтруизма). Дети рождаются с силой и интеллектом, равными соответствующим средним значениям этих характеристик у его родителей. Уровень силы и интеллекта изначально являются случайными величинами, задаваемыми в интервале от 0 до 100.



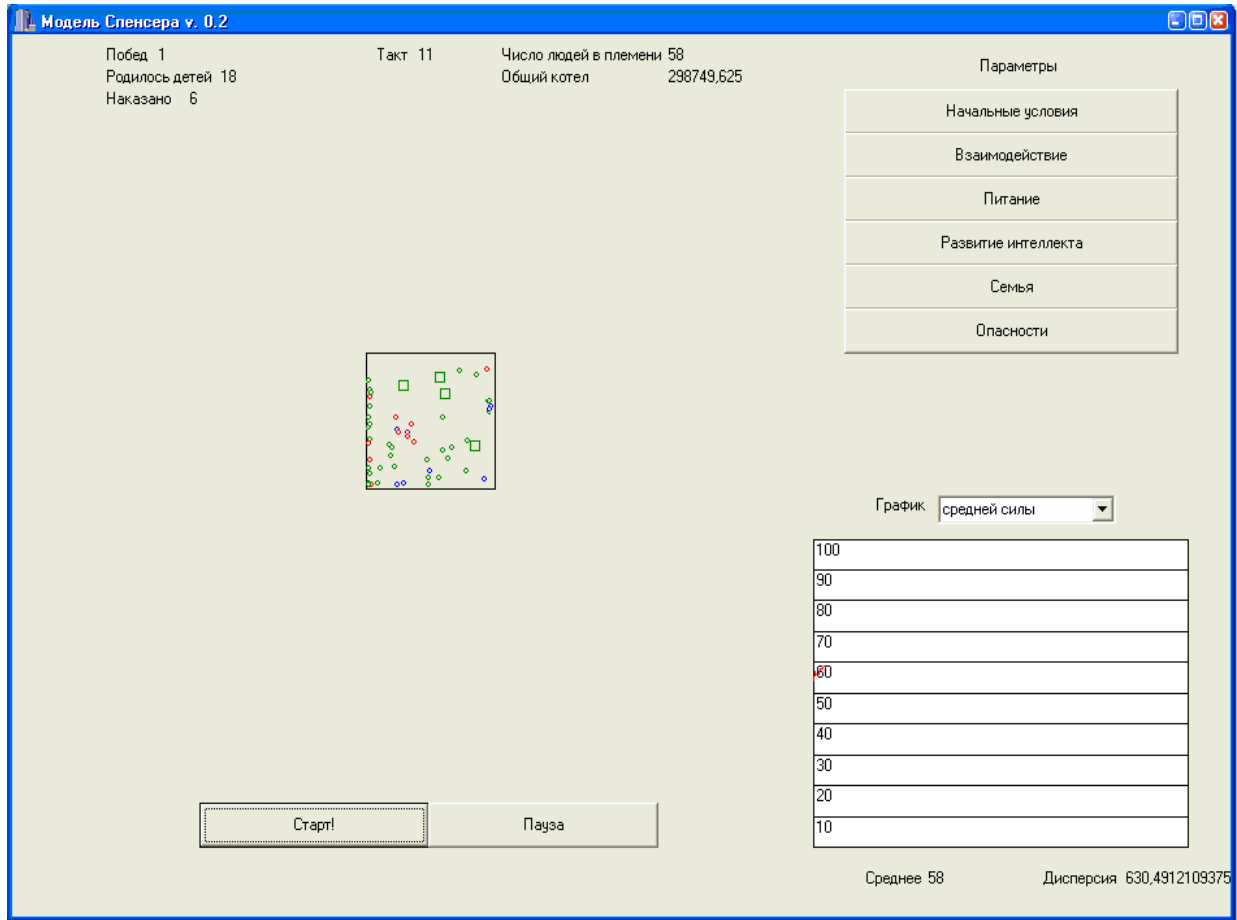


Рис. 1. Вид главного окна



Рис. 2. График средней силы



Рис. 3. График среднего интеллекта

Начальные условия

Число тактов	250
Длительность такта	100
Начальное число людей в племени	60
Доля женщин в племени	0,5
Доля "гена альтруизма" в племени	0,5
Доля "гена агрессии" в племени	0,1
Продолжительность жизни	40 +/- 10 лет

Рис. 4. Окно для ввода начальных условий

Взаимодействие между соплеменник...

Радиус взаимодействия	12
Число тактов невзаимодействия	2
Число тактов драки	1
Скорость перемещения	2
Радиус слышимости	10
Радиус видимости	10
Коэффициент подражания плохому	0,1
Коэффициент подражания хорошему	0,5
Коэффициент подражания Идеалу	0,5
Альтруизм Идеала	5

Рис.5. Окно для ввода условий взаимодействия между соплеменниками

Питание в племени

Прожиточный минимум

Функция распределения пищи

$$100000 + (-0,01)(x + -200)^2 + (-0,01)(y + -200)^2$$

Рис. 6. Окно для ввода параметров питания

Интеллект	Вероятность выбора наилучшей точки
< <input type="text" value="20"/>	<input type="text" value="0,25"/>
<input type="text" value="20"/> < < <input type="text" value="40"/>	<input type="text" value="0,4"/>
<input type="text" value="40"/> < < <input type="text" value="65"/>	<input type="text" value="0,6"/>
> <input type="text" value="65"/>	<input type="text" value="1"/>

Рис. 7. Окна для ввода параметров интеллекта

Семья

Максимальное число детей в семье

Вероятность появления ребенка

Возраст выхода из семьи

Вероятность передачи "гена альтруизма"

Рис. 8. Окно для ввода характеристик семьи

Класс опасности	Значение функции преодоления	Штраф здоровья
Зеленый	<input type="text" value="20"/>	<input type="text" value="20"/>
Желтый	<input type="text" value="30"/>	<input type="text" value="30"/>
Красный	<input type="text" value="40"/>	<input type="text" value="40"/>

Рис. 9. Окно для ввода параметров опасностей

## Расчеты и выводы

Целью проведенных расчетов являлось выявление влияния альтруизма членов племени на его выживаемость.

Начальные условия представлены выше.

Изменяющимся параметром была выбрана вероятность передачи «гена альтруизма».

Всего было проведено 500 запусков программы с одновременным изменением через каждые 50 запусков следующих параметров: доля гена альтруизма, коэффициент подражания хорошему, альтруизм Идеала, вероятность передачи «гена альтруизма». Результаты представлены в таблице ниже.

Таблица 1

Результаты имитационного моделирования

Начальные условия					Средние значения результирующих показателей для 10 серий запусков программы (по 50 запусков в каждой серии с одинаковыми начальными условиями)							
0,05	0,05	0,1	1	0,05	18,46	259,28	79,64	38,56	212035,52	0,01	-10,60	-12,26
0,15	0,15	0,1	2	0,15	16,22	253,88	64,10	33,16	191366,44	0,06	-15,92	-16,60
0,25	0,25	0,1	3	0,25	16,02	274,48	52,90	41,74	237158,52	0,13	-17,82	-18,88
0,35	0,35	0,1	4	0,35	0,94	313,62	39,98	44,78	302169,42	0,50	-9,80	-11,24
0,45	0,45	0,1	5	0,45	0,22	337,58	35,96	51,90	366640,62	1,23	-3,92	-6,56
0,55	0,55	0,1	6	0,55	0,34	348,30	24,16	59,00	455893,20	2,26	-0,20	-3,46
0,65	0,65	0,1	7	0,65	0,18	368,94	18,00	65,48	566372,74	4,08	-1,00	-2,48
0,75	0,75	0,1	8	0,75	0,18	374,12	10,18	63,60	619188,28	5,74	-3,20	-3,14
0,85	0,85	0,1	9	0,85	0,06	399,34	13,94	76,84	767630,56	8,31	-3,40	-1,54
0,9	0,9	0,1	9	0,9	0,14	392,54	5,06	75,50	775953,96	9,33	-1,14	-5,46

Таблица 2

Пояснения к таблице 1

Номер столбца таблицы 1	Название столбца таблицы 1
1	Доля гена альтруизма в племени
2	Коэффициент подражания хорошему
3	Коэффициент подражания плохому
4	Альтруизм Идеала
5	Вероятность передачи гена альтруизма
6	Число побед в драках за весь период существования
7	Число детей за весь период существования племени
8	Количество наказанных вождem за драку за весь период
9	Численность племени на конец периода существования
10	Количество пищи в котле на конец периода
11	Влияние общества на индивида в конце периода
12	Прирост силы на конец периода по сравнению с началом
13	Прирост интеллекта к концу периода в сравнении с началом

Альтруизм («хорошесть») каждого члена племени изменяется во времени по следующему правилу:

$$G_{t+1} = gA + DG_t^* + k(I - G_t), \text{ где}$$

$g, k$ , – коэффициенты,

$A$  – альтруизм члена племени в начальный момент времени,

$DG_t^*$  – средний прирост альтруизма во всем племени,

$I$  – альтруизм Идеала (внешний стандарт).

Начальное значение альтруизма является случайной величиной и задается программой без участия пользователя.

Из таблицы 1 видно, что с ростом показателей, характеризующих альтруизм, увеличивается численность племени, количество детей, количество собранной пищи, и уменьшается количество драк и убийств в племени.

Результаты имитаций, часть которых приведена в таблице, представлены ниже в графическом виде.

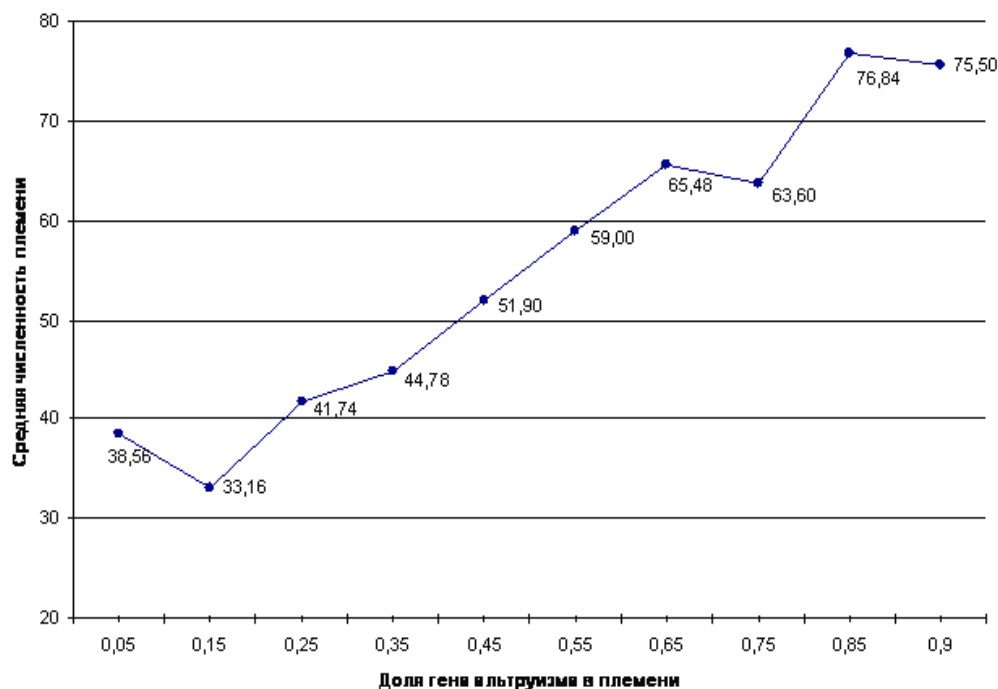


Рис. 10. Зависимость численности племени на конец периода от альтруизма

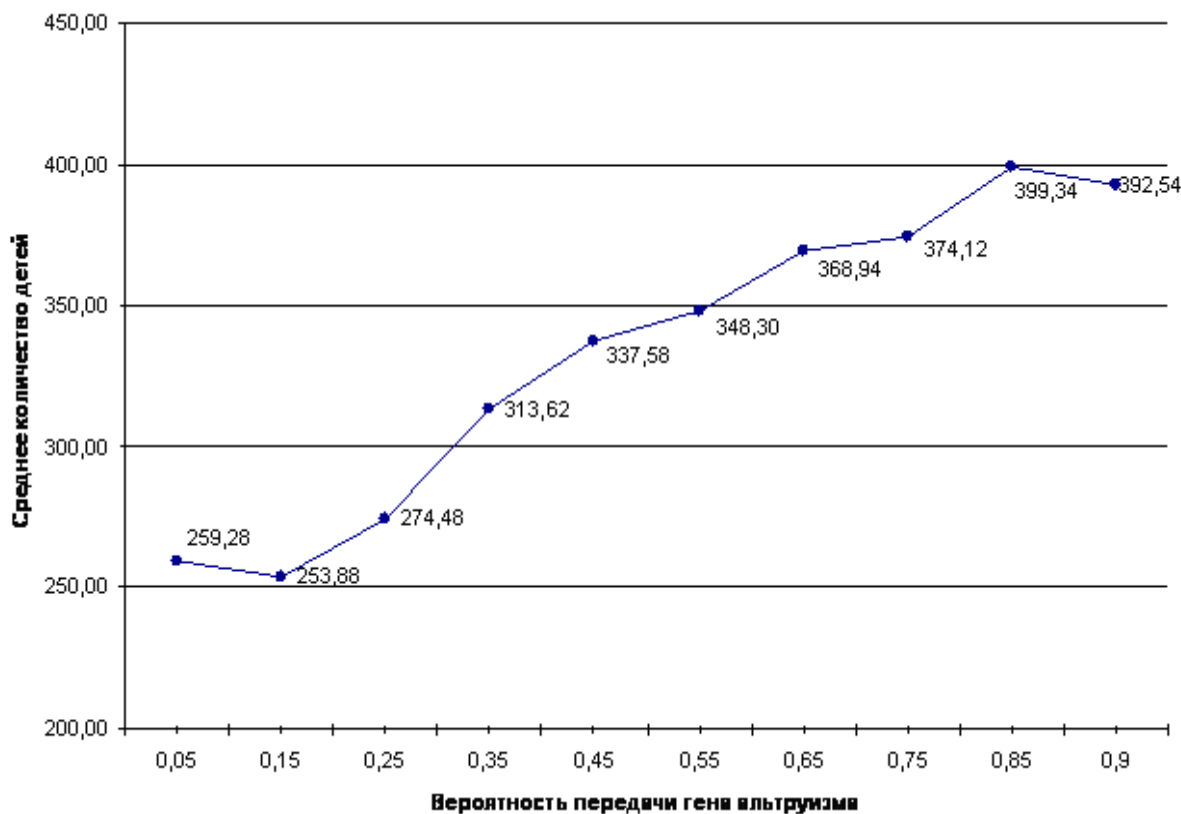


Рис. 11. Зависимость количества рожденных детей за весь период от альтруизма в племени

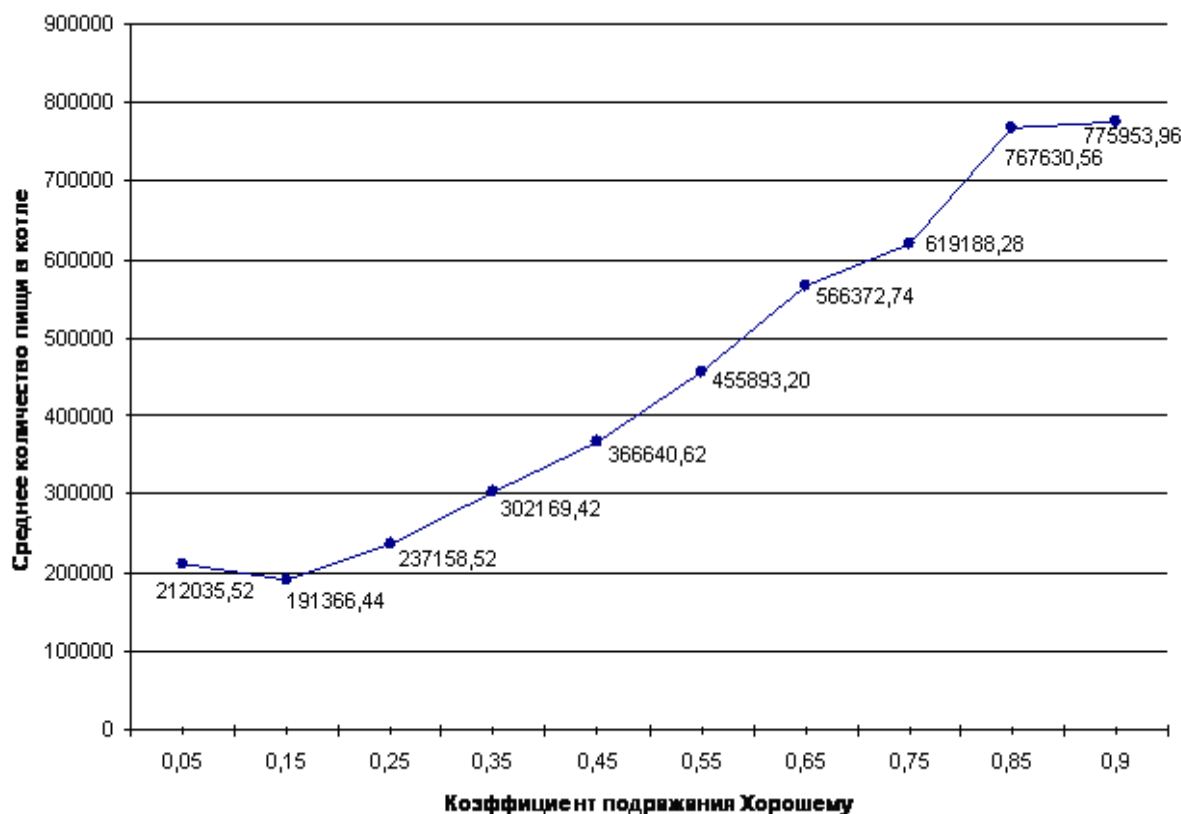


Рис. 12. Зависимость количества пищи в котле от альтруизма в племени



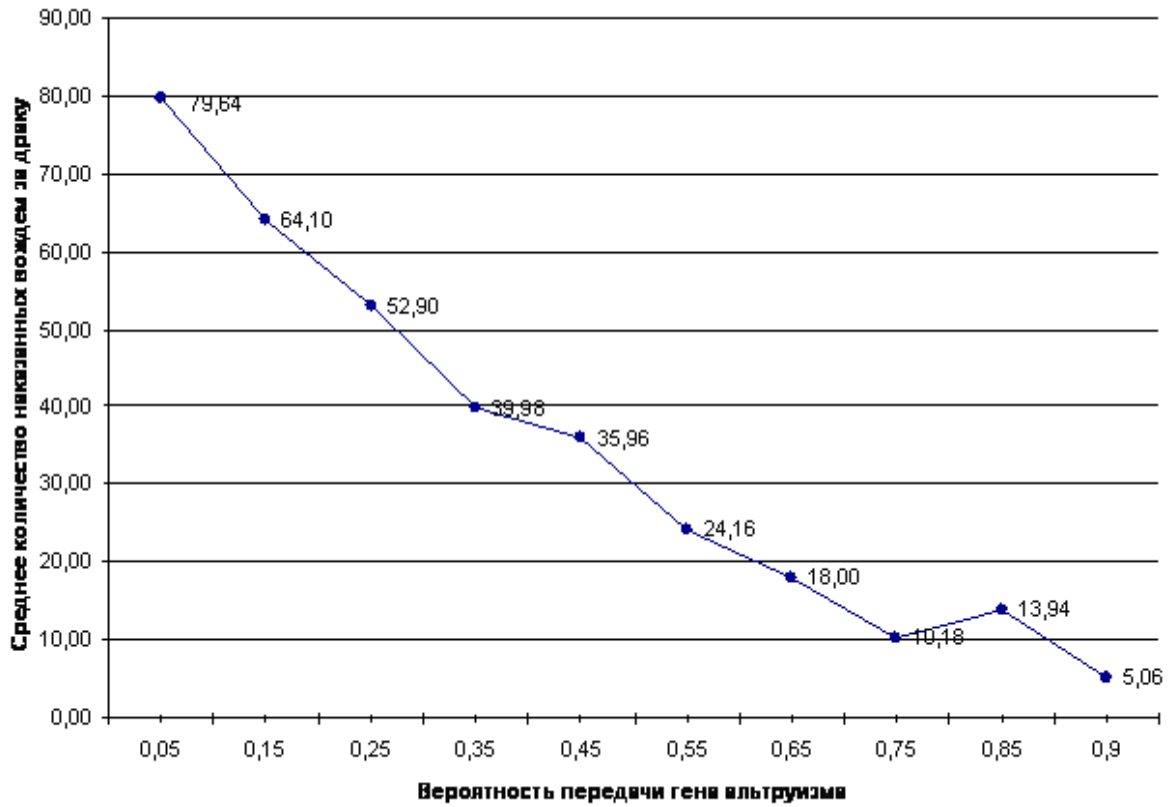


Рис. 13. Количество наказанных за драку в зависимости от альтруизма

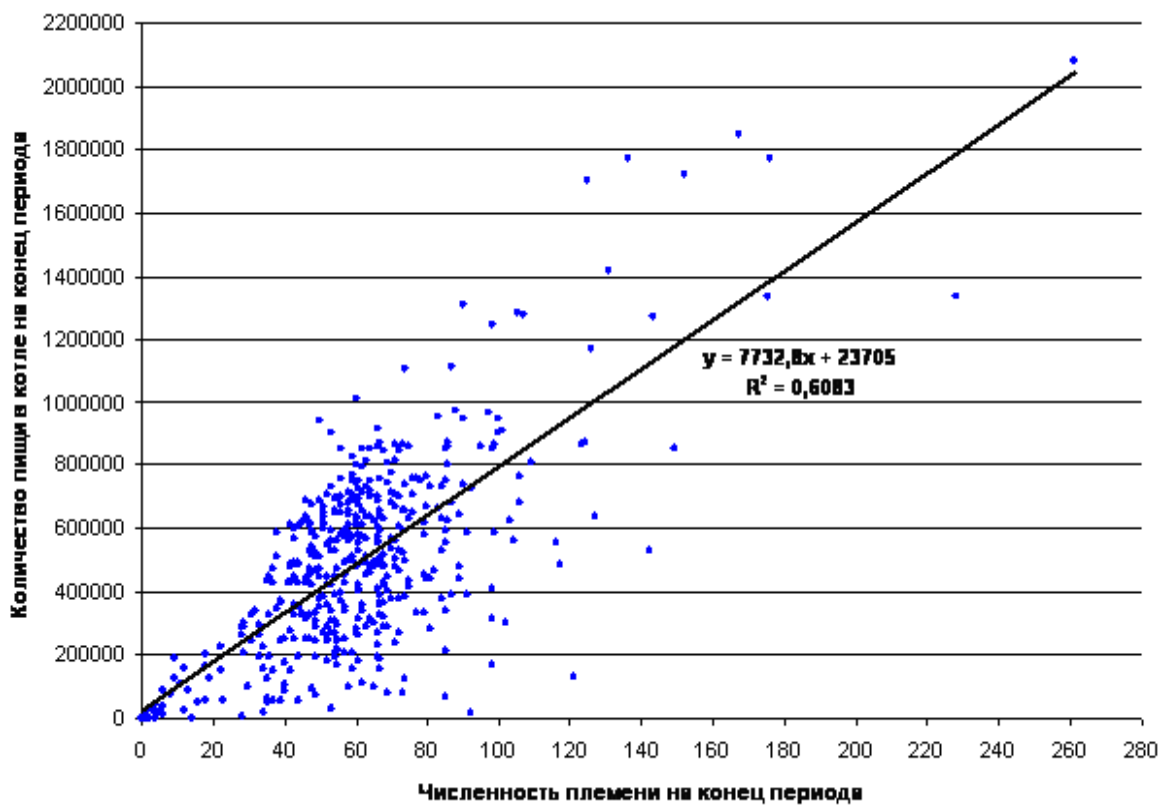


Рис 14. Зависимость между численностью племени и количеством пищи, принесенным в общий котел

В результате было установлено, что альтруизм оказывает заметное влияние на выживаемость племени и рост его численности. Существенное влияние оказывает также влияние внешней среды, т. е. наличие пищи и количества «опасностей» на территории проживания.

### Библиографический список

1. А.Р. Бахтизин. Агент-ориентированные модели экономики. М.: Экономика, 2008.
2. Ю.М. Бородай. К вопросу о социально-психологических аспектах происхождения и первобытнородовой общины. Принцип историзма в познании социальных явлений. М.: «Наука», 1972.
3. В.Л. Макаров, А.Р. Бахтизин, С.С. Сулакшин. Применение вычислимых моделей в государственном управлении. М.,: Научный эксперт, 2007.
4. Г. Спенсер. Научные основания нравственности. М.: УРСС, 2007 (факсимиле с издания 1892 г.).
5. В. П. Эфроимсон. Родословная альтруизма. Новый мир, №10, 1971.

Орлов А. И., Москва

## Теория измерений и методы анализа<sup>1</sup>

---

Предварительно обосновав необходимость развития научной специальности «Математические и инструментальные методы в социологии», рассматриваем роль теории измерений при выборе методов сбора и анализа социологической информации.

### Математические и инструментальные методы в социологии

Статья относится к достаточно самостоятельной области — математическим методам анализа социологических данных. Основной интерес в ней — к математическим вопросам, социологические постановки служат для постановки математических задач. Эта область относится к математической социологии — научной дисциплине, аналогичной математической экономике, математической физике и др.

Классификация наук закреплена формальными решениями. Например, в нашей стране утвержден список специальностей научных работников. Однако формальные решения могут быть модернизированы. Время от времени это происходит. Например, около 20 лет назад появились новые группы специальностей — социологические и политологические. Однако недостатки действующей системы очевидны. Приведем четыре примера.

*Пример 1.* Продолжает использоваться термин «физико-математические науки», хотя его нелепость ясна всем специалистам. Математика относится к формальным наукам, изучает конструкции, созданные мыслью, т. е. находящиеся не в реальном мире, а в идеальном (по Платону). Математика может быть применена в любой сфере деятельности, в любой отрасли народного хозяйства. Например, широко распространен термин «экономико-математические методы и модели», очевидно, относящийся к применению математики в экономике. В то же время физика — одна из областей естествознания, наука, изучающая наиболее общие и фундаментальные закономерности, определяющие структуру и эволюцию материаль-

---

<sup>1</sup> Работа выполнена при поддержке Министерства образования и науки РФ в рамках Постановления Правительства РФ № 218.

ного мира. Термин «физико-математические науки» не более обоснован, чем, например, термины «химико-математические науки» или даже «ветеринарно-социологические науки».

*Пример 2.* Как известно, статистические методы применялись на практике (и, следовательно, были теоретически разработаны) с древних времен. В Библии Ветхий завет начинается с Пятикнижия Моисеева, и четвертая книга Пятикнижия называется «Числа». Она начинается с описания проведенной под руководством Моисея переписи военнообязанных. Со времен библейского Моисея статистика получила значительное развитие. В США число статистических кафедр в университетах превышает число математических, соответственно и число статистиков больше числа математиков (примерно вдвое) [1]. Следовательно, в США статистика воспринимается одной из «больших» наук: математика, физика, статистика, химия, биология... Совсем не так в нашей стране. В официальной структуре науки статистика упоминается дважды, и оба раза на вторых ролях. Во-первых, как одна из экономических наук (специальность 08.00.12 «Бухгалтерский учет, статистика», присуждаются ученые степени по экономическим наукам). Во-вторых, в названии математической специальности 01.01.05 «теория вероятностей и математическая дисциплина» (присуждаются ученые степени по физико-математическим наукам). Все остальные применения статистических методов, в частности, в социологических исследованиях, остаются вне официальной структуры науки.

*Пример 3.* На знамени научного прогресса второй половины XX в. начертано: «Кибернетика». Однако нет в нашей стране докторов и кандидатов кибернетических наук (есть, правда, математическая специальность «Дискретная математика и математическая кибернетика», при защите присуждается ученая степень по физико-математическим наукам).

*Пример 4.* Очевидно, что менеджмент (управление людьми) — более широкая сфера деятельности, чем экономика. Управленческие решения необходимо принимать на основе все совокупности социальных, технологических, экологических, экономических, политических факторов [2]. Между тем в действующей официальной номенклатуре специальностей научных работников (в редакции Приказа Минобрнауки РФ от 11.08.2009 N 294) менеджмент находится внутри экономической специальности 08.00.05 «Экономика и управление народным хозяйством». При этом есть целый ряд технических специальностей, включающих в себя термин «управление», среди которых выделяется специальность 05.13.10 «Управление в социальных и экономических системах» (присуждаются ученые степени по техническим (!) наукам).

Приведенные примеры показывают, что действующая официальная номенклатура специальностей научных работников нуждается в модернизации.

К социологическим наукам близки экономические. Вплоть до того, что на включение в свою сферу маркетинга (изучения предпочтений потребителей) претендуют и те, и другие. Однако у экономистов есть специальность 08.00.13 «Математические и инструментальные методы в экономике», а у социологов нет аналогичной специальности, математическая социология не выделена среди социологических наук.

К чему это приводит? В частности, к отсутствию должного внимания к развитию математических методов в социологии, к их вытеснению из перечней секций социологических конференций и конгрессов. В результате падает квалификационный уровень работ. На заседании секции «Измерение в социологии» VI научно-практической конференции памяти первого декана факультета социологии Александра Олеговича Крыштановского «Современная социология — современной России» (1–3 февраля 2012 года) пришлось урезонивать воинствующего невежду, который пытался навязать докладчику свое неправильное понимание проверки значимости при проверке статистических гипотез. Впрочем, и докладчик продемонстрировал непонимание необходимости обязательной проверки значимости различия долей тех или иных значений признаков при сравнения совокупностей, сказавши: «В журнале «Социология-4М» нас *заставили* проверить значимость различия долей». К необходимости повышения качества математической составляющей социологических исследований мы старались привлечь внимание в работе [3].

Мы считаем необходимым усилить внимание к проблемам развития и применения математических методов анализа социологических данных, математического моделирования социальных процессов, короче — к математической социологии. Целесообразно в рамках социологической науки создать специальность «Математические и инструментальные методы в социологии», аналогичную экономической специальности «Математические и инструментальные методы в экономике».

К математическим методам в социологии относим не только методы анализа числовых и нечисловых социологических данных, но и методы математического моделирования социальных процессов [4, 5].

Под инструментальными методами понимаем прежде всего методы, нацеленные на развитие и применение информационных технологий, включая сетевые (в том числе модели распространения нововведений в сфере информационных и телекоммуникационных технологий [6] и онлайн исследования [7]).

### О развитии математической социологии в нашей стране

Много интересных работ, относящихся к математической социологии, было выполнено в нашей стране в 70–80-е годы XX в. Назовем только некоторые из них. В 1977 г. Институт социологических исследований выпустил сборники [8, 9]. На основе материалов Всесоюзной научной конферен-

ции «Проблемы применения математических методов в социологическом исследовании» издательство «Наука» опубликовала солидный сборник [10]. Хотя прошло уже 30 лет, материалы этих сборников по-прежнему актуальны. Квалифицированные работы не устаревают. И даже необходимо отметить методологическую несостоятельность современных публикаций Росстата по переписям населения по сравнению с книгой «Числа» Ветхого Завета, в которой рассказано о переписи военнообязанных, проведенной под руководством Моисея.

По сей день наиболее многоплановой публикаций по методам анализа нечисловых данных является сборник [11], подготовленный совместно академическим Институтом социологии и комиссией «Статистика объектов нечисловой природы» Научного Совета АН СССР по комплексной проблеме «Кибернетика». В настоящее время анализу нечисловых данных посвящены обширные разделы в учебниках по прикладной статистике [12], есть и специальные учебники по нечисловой статистике [13], но сборник [11] по-прежнему актуален и необходим тем, кто хочет разобраться в методах анализа нечисловой (т. е. качественной) информации в социологических исследованиях. Отметим, что именно практические запросы социологов (и специалистов по экспертным оценкам) послужили стимулом для разработки нечисловой статистики [13].

В дальнейшем появились адресованные студентам-социологам учебники и учебные пособия, в частности, книги Ю.Н. Толстовой [14-16] и Г.Г. Татаровой [17, 18]. С 1991 г. выпускается журнал «Социология: методология, методы, математическое моделирование» (сокращенно «Социология-4М»). Развитию математических и статистических методов в российской социологии посвящены обзорные работы [19, 20].

Казалось бы, все хорошо в области математической социологии. Однако всё заметнее проявляются отрицательные тенденции. Большинство социологов остаются невежественными в области методов анализа данных. Проявляется это, например, в преклонении перед давно устаревшим западным статистическим пакетом SPSS (анализу статистических пакетов посвящена статья [21]). Полученные еще в 70-е годы XX в. научные результаты остаются неизвестными, а потому, естественно, не применяются. Научный инструментарий социолога зачастую соответствует уровню XIX в. В последнее время даже номинальное признание важности математической социологии в виде организации отдельных секций на социологических конгрессах и конференциях постепенно сходит на нет. Подробнее эти мысли развиты в нашем выступлении [22] в «Дискуссии о социологии» на сайте Российского общества социологов.

### О теории измерений

Согласно теории измерений все реальные данные измерены в той или иной шкале [12-14]. Обычно выделяют шесть основных шкал — наименований (номинальную), порядковую, интервальную, отношений, раз-



ностей, абсолютную. Первые две – шкалы качественных признаков, остальные четыре – шкалы количественных признаков. Только абсолютная шкала не накладывает никаких ограничений на математические методы анализа данных.

Группы допустимых преобразований – вот что интересует нас в шкалах измерения, поскольку именно они задают ограничения на методы обработки данных, измеренных в соответствующих шкалах.

На основе анализа реальной ситуации выясняем шкалу измерения интересующих нас данных, т. е. ее группу допустимых преобразований. *Методы анализа данных должны быть инвариантны относительно этой группы.* Возникает целый ряд задач:

- является ли инвариантным конкретный метод анализа данных;
- найти хотя бы один конкретный метод анализа данных, решающий поставленную задачу;
- найти все инвариантные методы из некоторого заранее определенного класса;
- для каких шкал инвариантен заданный метод;
- найти все шкалы, относительно которых инвариантен определенный метод;
- найти все методы, выводы с помощью которых инвариантны с вероятностью, стремящейся к 1 (здесь своя достаточно обширная система постановок исследовательских задач), и т. д.

Перейдем к более подробному изложению. Выяснение типов используемых шкал необходимо для адекватного выбора методов анализа данных. Основопологающим требованием является независимость выводов от того, какой именно шкалой измерения воспользовался исследователь (среди всех шкал, переходящих друг в друга при допустимых преобразованиях). Например, если речь о длинах, то выводы не должны зависеть от того, измерены ли длины в метрах, аршинах, сажнях, футах или дюймах. Другими словами, выводы должны быть инвариантны относительно группы допустимых преобразований шкалы измерения. Только тогда их можно назвать адекватными, т. е. избавленными от субъективизма исследователя, выбирающего определенную шкалу из множества шкал заданного типа, связанных допустимыми преобразованиями.

Требование инвариантности выводов накладывает ограничения на множество возможных алгоритмов анализа данных. В качестве примера рассмотрим порядковую шкалу. Одни алгоритмы анализа данных позволяют получать адекватные выводы, другие – нет. Например, в задаче проверки однородности двух независимых выборок алгоритмы ранговой статистики (т. е. использующие только ранги результатов измерений) дают адекватные выводы, а статистики Крамера-Уэлча и Стьюдента – нет. Значит, для обработки данных, измеренных в порядковой шкале, критерии Смирнова и Вилкоксона можно использовать, а критерии Крамера-Уэлча и Стьюдента – нет.

## Выбор вида средних величин на основе условия устойчивости результата сравнения средних

Оказывается, требование инвариантности является достаточно сильным. Из многих алгоритмов анализа статистических данных ему удовлетворяют лишь некоторые. Покажем это на примере сравнения средних величин.

Пусть  $X_1, X_2, \dots, X_n$  - выборка объема  $n$ . Наиболее общее понятие средней величины введено французским математиком первой половины XIX в. академиком О. Коши. Средней величиной (по Коши) является любая функция  $f(X_1, X_2, \dots, X_n)$  такая, что при всех возможных значениях аргументов значение этой функции не меньше, чем минимальное из чисел  $X_1, X_2, \dots, X_n$ , и не больше, чем максимальное из этих чисел. Средними по Коши являются среднее арифметическое, медиана, мода, среднее геометрическое, среднее гармоническое, среднее квадратическое.

Средние величины используются обычно для того, чтобы заменить совокупность чисел (выборку) одним числом, а затем сравнивать совокупности с помощью средних. Пусть, например,  $Y_1, Y_2, \dots, Y_n$  - совокупность оценок экспертов, «выставленных» одному объекту экспертизы,  $Z_1, Z_2, \dots, Z_n$  - второму. Как сравнивать эти совокупности? Очевидно, самый простой способ - по средним значениям.

При допустимом преобразовании шкалы значение средней величины, очевидно, меняется. Но выводы о том, для какой совокупности среднее больше, а для какой - меньше, не должны меняться (в соответствии с требованием инвариантности выводов, принятом как основное требование в теории измерений). Сформулируем соответствующую математическую задачу поиска вида средних величин, результат сравнения которых устойчив относительно допустимых преобразований шкалы.

Пусть  $f(X_1, X_2, \dots, X_n)$  - среднее по Коши. Пусть среднее по первой совокупности меньше среднего по второй совокупности:

$$f(Y_1, Y_2, \dots, Y_n) < f(Z_1, Z_2, \dots, Z_n).$$

Тогда согласно теории измерений для устойчивости результата сравнения средних необходимо, чтобы для любого допустимого преобразования  $g$  (из группы допустимых преобразований в соответствующей шкале) было справедливо также неравенство

$$f(g(Y_1), g(Y_2), \dots, g(Y_n)) < f(g(Z_1), g(Z_2), \dots, g(Z_n)),$$

т. е. среднее преобразованных значений из первой совокупности также было меньше среднего преобразованных значений для второй совокупности. Причем сформулированное условие должно быть выполнено для любых двух совокупностей  $Y_1, Y_2, \dots, Y_n$  и  $Z_1, Z_2, \dots, Z_n$ . И, напомним, для любого допустимого преобразования. Средние величины, удовлетворяющие сформулированному условию, назовем *допустимыми* (в соответствую-

ющей шкале). Согласно теории измерений только допустимыми средними величинами можно пользоваться при анализе мнений экспертов и иных данных, измеренных в рассматриваемой шкале.

С помощью математической теории, развитой в монографии [23], удастся описать вид допустимых средних величин в основных шкалах. Рассмотрим обработку, для определенности, мнений респондентов или экспертов, измеренных в порядковой шкале. Справедливо следующее утверждение.

**Теорема 1.** *Из всех средних по Коши допустимыми средними в порядковой шкале являются только члены вариационного ряда (порядковые статистики).*

Теорема 1 справедлива при условии, что среднее  $f(X_1, X_2, \dots, X_n)$  является непрерывной (по совокупности переменных) и симметрической функцией. Последнее означает, что при перестановке аргументов значение функции  $f(X_1, X_2, \dots, X_n)$  не меняется. Это условие является вполне естественным, ибо среднюю величину находим для *совокупности (множества)* чисел, а не для *последовательности*. Множество не меняется в зависимости от того, в какой последовательности мы перечисляем его элементы.

Согласно теореме 1 в качестве среднего для данных, измеренных в порядковой шкале, можно использовать, в частности, медиану (при нечетном объеме выборки). При четном же объеме следует применять один из двух центральных членов вариационного ряда - как их иногда называют, левую медиану или правую медиану. Моду тоже можно использовать - она всегда является членом вариационного ряда. Можно применять выборочные квартили, минимум и максимум, децили и т. п. Но никогда нельзя рассчитывать среднее арифметическое, среднее геометрическое и т. д.

Естественная система аксиом (требований к средним величинам) приводит к так называемым ассоциативным средним. Их общий вид нашел в 1930 г. А.Н.Колмогоров [24]. Теперь их называют «средними по Колмогорову». Для чисел  $X_1, X_2, \dots, X_n$  средним по Колмогорову является

$$G\{(F(X_1) + F(X_2) + \dots + F(X_n))/n\},$$

где  $F$  - строго монотонная функция (т. е. строго возрастающая или строго убывающая),  $G$  - функция, обратная к  $F$ . Среди средних по Колмогорову - много хорошо известных персонажей. Так, если  $F(x) = x$ , то среднее по Колмогорову - это среднее арифметическое, если  $F(x) = \ln x$ , то среднее геометрическое, если  $F(x) = 1/x$ , то среднее гармоническое, если  $F(x) = x^2$ , то среднее квадратическое, и т. д. (в последних трех случаях усредняются положительные величины).

Среднее по Колмогорову - частный случай среднего по Коши. С другой стороны, такие популярные средние, как медиана и мода, нельзя представить в виде средних по Колмогорову. Справедливы следующие утверждения.

**Теорема 2.** *В шкале интервалов из всех средних по Колмогорову допустимым является только среднее арифметическое.*

Таким образом, среднее геометрическое или среднее квадратическое температур (в шкале Цельсия), потенциальных энергий или координат точек не имеют смысла. В качестве среднего надо применять среднее арифметическое. А также можно использовать медиану или моду.

**Теорема 3.** *В шкале отношений из всех средних по Колмогорову допустимыми являются только степенные средние с  $F(x) = x^c$ ,  $c \neq 0$ , и среднее геометрическое.*

Есть ли средние по Колмогорову, которыми нельзя пользоваться в шкале отношений? Конечно, есть. Например, с  $F(x) = e^x$ . Среднее геометрическое является пределом степенных средних при  $\tilde{n} \rightarrow 0$ . Теоремы 2 и 3 справедливы при выполнении некоторых внутриматематических условий регулярности.

На наш взгляд, теоремы 1-3 должны быть известны всем студентам-социологам. (Как и все специалисты, я не могу претендовать на полное знание литературы. Буду благодарен за указание учебников для социологов, в которых приведены теоремы 1-3.)

Аналогично средним величинам могут быть изучены и другие статистические характеристики - показатели разброса, связи, расстояния и др. Нетрудно показать, например, что коэффициент корреляции не меняется при любом допустимом преобразовании в шкале интервалов, как и отношение дисперсий. Дисперсия не меняется в шкале разностей, коэффициент вариации - в шкале отношений, и т. д.

К сожалению, достаточно систематическому изучению подверглись лишь средние величины (см. выше) и расстояния ( см. [14, 25] и другие работы Ю.Н. Толстовой). Отметим, что исходные работы 1970-х годов по средним величинам породили достаточно обширное множество следующих работ (обзор дан в [26]), к сожалению, ничего существенно не добавивших к полученному вначале.

По нашему мнению [23, 27], необходимо развивать теорию на стыке математической статистики и теории измерений. Это — призыв к математикам. А социологов надо призвать к использованию полученных результатов. Например, для *усреднения порядковых данных использовать не среднее арифметическое, а медиану.*

Как показали многочисленные опыты, человек более правильно (и с меньшими затруднениями) отвечает на вопросы качественного, например, - сравнительного характера, чем количественного [28]. В очередной раз мы убедились в этом в 2011-2012 гг., проводя опросы летного состава авиакомпании «Волга-Днепр» в ходе разработки автоматизированной системы прогнозирования и предотвращения авиационных происшествий. Пилоты достаточно уверенно отвечали на вопросы о том, какое из предшествующих событий сильнее влияет на последующее, без труда ранжировали предшествующие события. Измерения в порядковой шкале не представляли для них сложности. При этом на вопросы типа «В скольких случаях на 1000 полетов осуществится определенное событие» отвечали с трудом или

вообще отказывались отвечать. Поэтому пришлось отказаться от измерений в количественных шкалах и ограничиться порядковыми, с соответствующими ограничениями на методы обработки данных.

### Библиографический список

1. Налимов В.В. О преподавании математики экспериментаторам // О преподавании математической статистики экспериментаторам. Препринт Межфакультетской лаборатории статистических методов №17. – М.: Изд-во Московского университета им. М.В. Ломоносова, 1971. – С.5-39.
2. Орлов А.И. Менеджмент: организационно-экономическое моделирование. Учебное пособие для вузов. Ростов-на-Дону: Феникс, 2009.
3. Орлов А.И. Об оценке качества процедур анализа данных // Социологические методы в современной исследовательской практике: Сборник статей, посвященный памяти первого декана факультета социологии НИУ ВШЭ А.О. Крыштановского / Отв. ред. и вступит. ст. О.А. Оберемко; НИУ ВШЭ, ИС РАН, РОС. М.: НИУ ВШЭ, 2011. – С.7-13.
4. Математическое моделирование социальных процессов. Вып. 10: сб. ст. / Под ред. А.П. Михайлова. М.: КДУ, 2009.
5. Шведовский В.А. Особенности социолого-математического моделирования в исследовании социальных процессов. – М.: АПК и ППРО, 2009.
6. Делицын Л.Л. Количественные модели распространения нововведений в сфере информационных и телекоммуникационных технологий. М.: МГУКИ, 2009.
7. Онлайн исследования в России 2.0 / Под ред. Шашкина А.В., Девятко И.Ф., Давыдова С.Г. – М.: РИЦ «Северо-Восток», 2010.
8. Методы современной математики и логики в социологических исследованиях. / Под ред. Э.П. Андреева. М.: Институт социологических исследований АН СССР, 1977.
9. Математические методы и модели в социологии. / Под ред. В.Н. Варыгина. М.: Институт социологических исследований АН СССР, 1977.
10. Математические методы в социологическом исследовании. / Под ред. Т.В. Рябушкина и др. М.: Наука, 1981.
11. Анализ нечисловой информации в социологических исследованиях. / Под ред. В.Г. Андреевкова, А.И. Орлова, Ю.Н. Толстой. М.: Наука, 1985.
12. Орлов А.И. Прикладная статистика. М.: Экзамен, 2006.



13. Орлов А.И. Организационно-экономическое моделирование: учебник: в 3 ч. Часть 1: Нечисловая статистика. М.: Изд-во МГТУ им. Н.Э. Баумана. 2009.
14. Толстова Ю.Н. Измерение в социологии. М.: Инфра-М, 1998.
15. Толстова Ю.Н. Анализ социологических данных: методология, дескриптивная статистика, изучение связей между номинальными признаками. М.: Научный мир, 2000.
16. Толстова Ю.Н. Основы многомерного шкалирования. Учебное пособие для вузов. М.: Издательство КДУ, 2006.
17. Татарова Г.Г. Методология анализа данных в социологии (введение). Учебник для вузов. М.: NOTA BENE, 1999.
18. Татарова Г.Г. Основы типологического анализа в социологических исследованиях. М.: Издательский Дом «Высшее образование и наука», 2007.
19. Толстова Ю.Н. Математические методы в социологии. / Социология в России. Под ред. В.А. Ядова. - 2-е изд., перераб. и дополн. - М.: Издательство Института социологии РАН, 1998. С.83-89, 98-103.
20. Орлов А.И. Статистические методы в российской социологии (тридцать лет спустя). - Журнал «Социология: методология, методы, математические модели». 2005. №.20. С.32-53.
21. Орлов А.И. Статистические пакеты – инструменты исследователя. - Журнал «Заводская лаборатория». 2008. Т.74. №.5. С.76-78.
22. Орлов А.И. Черная дыра отечественной социологии. - Выступление 09-01-2011 в «Дискуссии о социологии» на сайте Российского общества социологов URL: [http://www.ssa-rss.ru/index.php?page\\_id=19&id=456](http://www.ssa-rss.ru/index.php?page_id=19&id=456).
23. Орлов А.И. Устойчивость в социально-экономических моделях. М.: Наука, 1979.
24. Колмогоров А.Н. Избранные труды. Математика и механика. М.: Наука, 1985. С. 136–138.
25. Толстова Ю.Н. Адекватность функции расстояния в алгоритмах автоматической классификации. – В сб.: Исследования по вероятностно-статистическому моделированию реальных систем. М.: ЦЭМИ АН СССР, 1977. С.168-173.
26. Барский Б. В., Соколов М. В. Средние величины, инвариантные относительно допустимых преобразований шкалы измерения. – Журнал «Заводская лаборатория». 2006. №1. С.59-.67.
27. Орлов А.И. Связь между средними величинами и допустимыми преобразованиями шкалы. – Журнал «Математические заметки». 1981. Т. 30. №.4. С. 561-568.
28. Тюрин Ю.Н., Литвак Б.Г., Орлов А.И., Сатаров Г.А., Шмерлинг Д.С. Анализ нечисловой информации (препринт). М.: Научный Совет АН СССР по комплексной проблеме «Кибернетика», 1981.



Сороко Е. Л., Москва

## Новые функции Базы данных демографических показателей по регионам России и странам мира

---

### Аннотация

В статье рассматриваются некоторые новые функции Базы данных демографических показателей по регионам России и странам мира. Предложена постановка задачи гармонизации единиц измерения показателя в различных информационных массивах, рассмотрены возможные подходы к ее решению.

**Ключевые слова:** база данных, демографический показатель, единица измерения, информационный массив, регионы России, страны мира

### Постановка задачи

В настоящее время в Национальном исследовательском университете Высшая школа экономики по проекту «Учитель-ученики» Научного Фонда НИУ ВШЭ №11-04-0039 разрабатывается База данных демографических показателей по регионам России и странам мира [1,2]. База данных содержит несколько сот информационных массивов, при работе с которыми возник ряд новых достаточно интересных задач, требующих решения. В данной статье предполагается рассмотреть следующие две. Первая связана с единицей измерения демографического показателя. Опыт работы с Базой данных показал, что число различных единиц измерения, используемых в информационных массивах демографических показателей из различных источников, превысил уже полсотни. Возникает вопрос о том, каким образом и на какой стадии работы в базе данных должны быть учтены различия в единицах измерения одного и того же показателя в информационных массивах, полученных из разных источников. Вторая проблема связана с необходимостью логического контроля значений демографического показателя, загружаемого в Базу данных. Здесь альтернативами может служить два подхода: забраковать ошибочные значения на этапе формирования входных информационных массивов, либо отсекал их «на лету» на этапе формирования результата запроса к Базе данных. Предполагается обсудить достоинства и недостатки каждого из подходов.

## Единицы измерения демографических показателей

Важнейшей компонентой описания информационного массива демографического показателя является единица измерения. Это связано с тем, что разные показатели могут иметь различную размерность и поэтому измеряться в различных единицах, а один и тот же показатель в массивах из разных источников может измеряться различными единицами, отличающимися масштабом. Простейшим примером может служить единица измерения такого показателя как численность населения. В ряде источников она может представляться числом лиц (людей), в других даваться в тысячах или миллионах человек. Демографические коэффициенты могут быть измерены в расчете на 1000 населения, но в других случаях используется единица измерения %. Коэффициенты смертности могут быть измерены в расчете на тысячу жителей, а в других случаях – на 100.000 населения. Если рассмотреть один конкретный показатель, то для его корректного хранения в базе данных требуется решение задачи, которую можно было бы назвать задачей гармонизации или задачей стандартизации единицы измерения. То есть, во всех информационных массивах, содержащих значения данного показателя должна быть использована одна и та же единица измерения.

Как может быть решена эта задача? На первый взгляд может показаться, что существует только один единственно правильный способ ее решения. Он состоит в том, что для данного показателя выбирается (утверждается) одна «правильная», «стандартная» единица измерения. Все информационные массивы, в которых единица измерения совпадает со «стандартной», значения показателя сохраняются в неизменном виде. В тех же информационных массивах, где она отличается от стандарта, все значения в них пересчитываются с использованием коэффициента, равного соотношению единицы измерения в исходном массиве и стандарте.

Однако это не совсем так, поскольку может быть реализован и альтернативный подход. Он состоит в том, что все исходные информационные массивы, содержащий данный демографический показатель, остаются неизменными независимо от использованных в них единицах измерения. Пересчет же на «стандартную» единицу измерения производится «на лету» только в процессе формирования запроса к Базе данных. Данный метод может оказаться более подходящим для новых показателей, у которых еще нет устоявшегося стандарта единицы измерения, в случае неожиданной смены «стандарта», различных стандартах для разных стран мира. Реализация данной функции предполагается в разрабатываемой базе данных.

## Контроль значений демографических показателей

Важнейшим критерием качества базы данных является хранение в ней логически корректных, правильных значений показателя, которые не противоречат данным статистики и его содержательному смыслу. Возникает естественный вопрос, как обеспечить высокое качество базы данных по данному критерию. То есть избавиться ее от значений, которые противоречат мировой демографической статистике и науке. Например, было бы невозможно пользоваться такими значениями показателей, как нулевая смертность, общий коэффициент брачности населения в 500 промилле в год или коэффициент суммарной рождаемости в 48 детей на 1 женщину.

Подобно первой проблеме, здесь также на первый взгляд может показаться, что существует единственно правильный способ ее решения. И состоит в том, что значения данного показателя должны быть проверены перед загрузкой в базу данных на соответствие существующей науке и статистике. Для всех значений, которые им не соответствуют, должен быть установлен барьер для загрузки в базу данных, они не должны в нее попасть. Значения же, которые не противоречат им, могут быть загружены.

Но и для решения данной задачи может быть также реализован альтернативный подход. Он состоит в том, что все исходные информационные массивы, оставляются неизменными и загружаются в базу независимо от их значений. Проверка же на допустимость их значений производится «на лету» только в процессе формирования запроса к Базе данных. Такой метод может оказаться более подходящим для показателей, у которых имеется значительный тренд в динамике для развитых стран мира или при появлении новых научных знаний. В этом случае решение о том, является ли значение показателя корректным, может быть пересмотрено с использованием «гибкой» коррекции границ допустимых значений конкретного показателя. Реализация данной функции также предполагается при разработке данной базы.

## Библиографический список

1. Сороко Е.Л. База данных демографических показателей по регионам России и странам мира: принципы построения // «Инновационное развитие экономики России: институциональная среда»: Четвертая Международная научная конференция, Москва, МГУ, Экономический факультет; 20–22 апреля 2011 г. Сборник статей: Том 4 / Под ред. В.П. Колесова, Л.А. Тутова. – М., МАКС Пресс, 2011, с. 1077-1085. (URL: [http://db.demoscope.ru/doc/soroko\\_lomonosov\\_2011.doc](http://db.demoscope.ru/doc/soroko_lomonosov_2011.doc)).
2. Сороко Е.Л. База данных демографических показателей по регионам России и странам мира: основные компоненты / XIII Апрельская международная научная конференция по проблемам развития экономики и общества НИУ ВШЭ, 3–5 апреля 2012 г. (URL: <http://regconf.hse.ru/uploads/604e4a2edd2abe0d64d5dd50de29e1c49c0b1007.doc> - текст доклада).

Сушко Е. Д., Москва

## Имитация поведения человека-работника и его взаимодействия с внешней средой в мультиагентной региональной модели

---

### Аннотация

В статье представлена агент-ориентированная модель, имитирующая социально-экономическое состояние региона на основе реконструкции его внутренней структуры и деятельности включенных в него экономических агентов разных типов. Основное внимание уделено модели личности агентов-людей и поведению таких агентов в качестве участников производства.

**Ключевые слова:** агент-ориентированное моделирование, структура личности, трудовой потенциал, поведенческая экономика, апробация региональной политики

### Постановка задачи

Целью настоящей работы являлось создание достаточно реалистичной компьютерной модели, имитирующей поведение региона на основе реконструкции его внутренней структуры, а также структуры и поведения действующих на его территории самостоятельных экономических субъектов.

В последнее время для моделирования подобных активных систем [1] все шире применяется особый вид имитационного моделирования, а именно – агент-ориентированное [2, 3]. Агент-ориентированная (мультиагентная) модель представляет собой искусственное общество, состоящее из взаимодействующих между собой самостоятельных агентов, каждый из которых обладает заданным набором личностных характеристик («ресурсов»), целевой функцией («интересами») и подчиняется правилам поведения, определяющим его реакцию в различных ситуациях, затрагивающих сферу его интересов. Важной особенностью агент-ориентированных моделей является то, что конкретные значения характеристик агентов могут различаться. Именно эта особенность позволяет использовать мультиагентные модели для воссоздания внутренней структуры больших социально-экономических систем, и тогда полученное искусственное общество

состоит не из неких усредненных «типичных представителей» сообщества моделируемого, а из множества самых разнообразных его представителей, встречающихся действительности. Реальное общество неоднородно, и изменения макроэкономических параметров, так же, как и изменения социальной и/или экономической политики властей, по-разному влияет на состояние и поведение различных экономических субъектов, а также их групп, действующих на территории региона. И это представляется существенным при моделировании поведения региона как сложного экономического субъекта в нестабильной среде, а также при разработке и апробации методов регулирования экономики с учетом человеческого фактора.

Очевидно, что самым важным из таких экономических субъектов является человек, живущий на территории региона, участвующий в производстве, производящий и потребляющий общественные блага. Важным — потому, что, с одной стороны, улучшение качества жизни и повышение удовлетворенности ею людей является критерием успешности региональной политики, а с другой стороны, именно трудовая деятельность людей является «ресурсом» для экономических объектов других уровней и служит средством реализации любой политики.

### Конструкция модели

В разработанной региональной модели [4] присутствуют агенты, соответствующие основным типам экономических акторов, которые образуют следующую иерархию: агенты-люди → агенты-организации, в которых люди трудятся, → агенты-муниципальные районы, на территории которых находятся организации и проживают люди, → регион-субъект Российской Федерации, к которому относятся муниципальные районы.

Каждый тип агентов обладает определенным набором характеристик, обуславливающих его ценность с точки зрения предназначенной ему в модели роли в общественном производстве, причем часть из этих характеристик сохраняют постоянное значение, а часть — изменяются во времени и/или вследствие каких-либо процессов, происходящих в модели. Агрегирование частных характеристик позволяет оценить общую «полезность» каждого отдельного агента, которая в модели для агентов-людей соответствует их трудовому потенциалу, а для агентов следующих уровней является функцией от суммарного трудового потенциала включенных в них агентов-людей. Объемы производства агентов-предприятий реального сектора облагаются налогами и служат источником доходов регионального бюджета и бюджетов муниципальных районов, так же, как и доходы работающих агентов-людей. Далее поступившие в бюджет средства распределяются в соответствии с заданными нормативами между государственным управлением, образованием, здравоохранением, наукой и культурой, а оставшиеся средства могут быть направлены на развитие других отраслей. Таким образом, деятельность агентов-людей трудоспособного возраста, реализующих свой трудовой потенциал, служит в модели основой всей экономической жизни региона.



Кроме того, каждый тип агентов обладает «системой ценностей» (предпочтений) и способен оценивать доступную ему информацию о характеристиках внешней среды и своем положении в ней, на основе чего имитируется активность агентов, их «поведение» как целенаправленное изменение собственных характеристик для максимизации своей функции полезности. На каждом шаге работы модели (один шаг соответствует одному году в реальной действительности) агенты, занятые в производстве, анализируют доступную им информацию о характеристиках среды и собственном положении в ней, и в соответствии со своей системой ценностей переходят в то или иное состояние: «нормальное» (стартовое), «недовольство» и «критическое» (необходимость срочных действий). Если состояние какого-либо агента отлично от нормального, то это побуждает его к реакции, причем, эта реакция будет зависеть не только от значений собственных характеристик агента, но и от возможностей, которые предоставляет ему внешняя среда. Таким образом, поведение агентов является функцией структуры их личности и характеристик внешней среды, что для агентов-людей согласуется с теорией личности Р. Кеттела [5].

Роль внешней для агентов среды играют агенты вышестоящих уровней иерархии (агенты одного уровня в модели напрямую не контактируют друг с другом). Так, для агентов-людей ими являются агенты-организации, агенты-муниципальные районы и регион в целом, то есть, в соответствии с общепринятой терминологией [6], описываемая модель относится к классу  $SNP^n$ -моделей (*Social Norm Promotion with n Influential Groups*) — агент-ориентированных моделей поощрения социальных норм с  $n$  группами влияния, где  $n=3$ .

### Модель человека

Рассмотрим конструкцию трудового потенциала, реализованную в модели, а также внешние условия, определяющие успех функционирования агентов-людей. Агенты—жители региона обладают следующим набором характеристик, обуславливающих их «ценность» с точки зрения участия в общественном производстве:

- Психофизиологические: возраст, пол, состояние здоровья, уровень способностей, темперамент. Так как «темперамент — это характеристика индивида со стороны его динамических особенностей: интенсивности, скорости, темпа, ритма психических процессов и состояний» [7], то и в модели наделение агента определенным типом темперамента (флегматик, меланхолик, холерик или сангвиник) определяет такие его свойства, как активность, скорость реакции и способность к адаптации в изменяющихся условиях внешней среды.

- Профессиональные: образование, сфера деятельности, опыт работы.

- Личностные: отношение к труду, личностная направленность агента на внешний или же на внутренний мир (экстраверсия – интроверсия).

Такие характеристики как здоровье, способности, активность, отношение к труду измеряются по шкале уровней, где 1 означает низкий уровень, 2 – нормальный, а 3 – высокий.

Интегральной характеристикой агента является Трудовой потенциал, который принимает значение от 0 до 1 и определяется двумя группами характеристик агента:

1. *Характеристики энергетического потенциала*, к которым относятся здоровье, способности, способность к творчеству и активность.

2. *Характеристики социального потенциала*, к которым относятся квалификация (состоящая из образования и трудового стажа) и отношение к труду.

Каждый показатель группы переводится в шкалу от 0 до 1:

$$x' = \frac{x - \min_{x \in X}(x)}{\max_{x \in X}(x) - \min_{x \in X}(x)}$$

где  $x$  – исходное значение показателя у агента,  $x'$  – нормированное, а  $X$  – множество всех возможных значений данного показателя.

По каждой из групп показателей определяется соответствующий потенциал как среднее арифметическое нормированных значений показателей группы, а интегральное значение трудового потенциала определяется как среднее геометрическое из двух полученных групповых потенциалов. Данная процедура развивает подход, предложенный в [8].

**Система ценностей:** приоритетность основных жизненных устремлений и соответствующих критериев оценки удовлетворенности своим положением. В модели предусмотрены такие критерии как уровень дохода, уверенность в завтрашнем дне (стабильность), карьерный рост и самореализация, а система ценностей представляет собой массив четырех чисел, каждое из которых определяет значимость (вес) соответствующего критерия (в сумме веса составляют единицу).

**Условия внешней среды:**

1. *Зависящие от места жительства* – рабочее место, налоговая нагрузка, инфраструктура, системы здравоохранения и образования. Работа системы здравоохранения в модели влияет на состояние здоровья агентов, которое является одним из элементов трудового потенциала, а система образования обеспечивает необходимый уровень квалификации работников, то есть, также влияет на величину их трудового потенциала. Наличие рабочих мест на рынке труда определяет спрос, а налоговая нагрузка и развитость инфраструктуры влияют на уровень удовлетворенности агентов.

2. *Зависящие от места работы* – основаны на понятии организационной культуры как специфического для каждой организации набора целей и соответствующих критериев эффективности деятельности, а также ключевых факторов, определяющих эту эффективность [9]. Организационная культура определяет спрос на такие свойства агента как активность и креативность – если агент обладает этими свойствами, и они востребованы, то возрастают результаты его труда и заработная плата. Таким образом, максимально реализовать свои способности агенты могут только при соответствующем профиле организационной культуры.

**Поведение.** Самостоятельные изменения характеристик агента: поступление на учебу, перемена места жительства, смена/выбор места работы, изменение усердия в труде. Так, если недовольному агенту не удастся улучшить свое положение, то снижается его усердие в труде [10, с. 61] и, как следствие, – трудовой потенциал.

Алгоритм поведения, реализованный в модели, основан на структуре поведенческого акта из теории функциональных систем физиолога П.К. Анохина [11], включающей стадии: определения уровня притязаний и идентификации ситуации взаимодействия с внешней средой как требующей принятия каких-либо мер; постановки целей действия и формирования его программы; реализации действия и оценки достигнутого результата. Схема поведения агентов в модели представлена на рисунке 1.

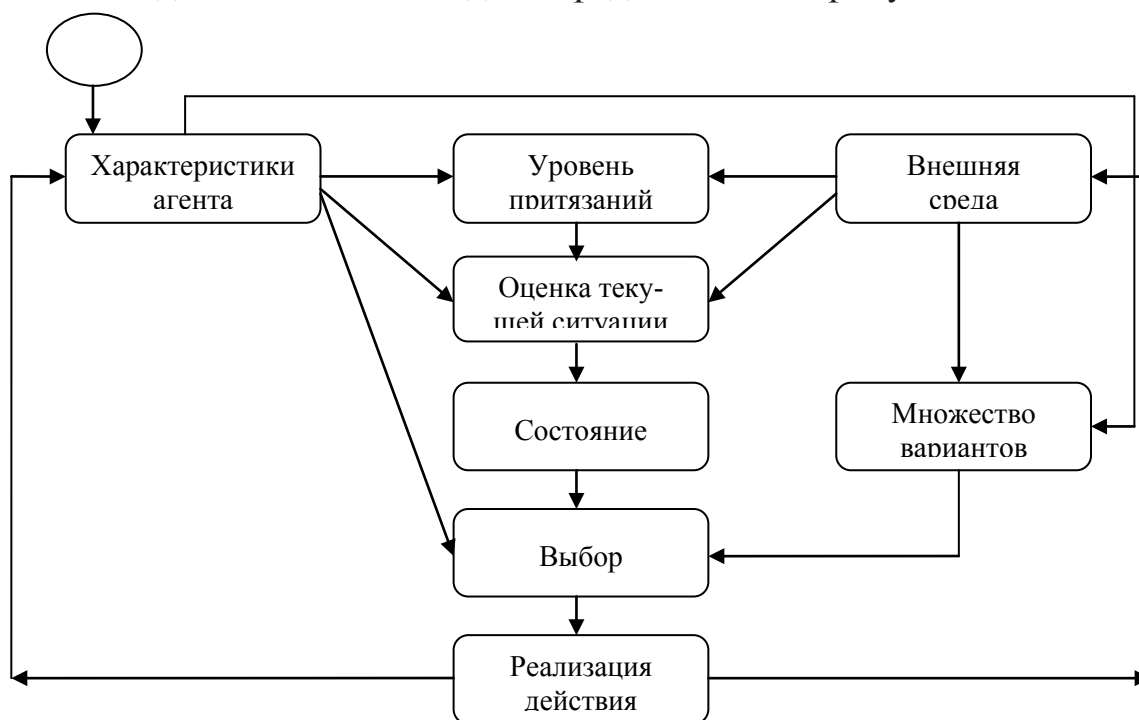


Рис. 1. Схема поведения агента

## Реализация модели

Модель реализована на примере Вологодской области в виде пользовательского программного продукта, разработанного в среде AnyLogic и названного «Агент-ориентированная Региональная Модель (АРМ) «Губернатор».

Для релевантного моделирования трудового потенциала региона, его развития и использования, а также для обеспечения комфортного для пользователя режима работы при проведении компьютерных экспериментов, модель «Губернатор» включает следующие блоки:

- блок установки начального состояния системы, соответствующего фактическим данным о социально-экономических характеристиках региона в выбранном базовом году;
- модель формирования и использования бюджета;
- модель прогнозирования демографической ситуации на территории региона и отдельных его муниципальных районов;
- модель трудового потенциала агента-человека и его поведения как работника;
- модель рынка труда и структуры занятости;
- модель производства;
- блок визуализации результатов и взаимодействия с пользователем.

Модель калибровалась на данных Росстата о состоянии Вологодской области в базовом (в качестве которого выбран 2004) и последующих годах, а также данных, полученных в ходе исследований Вологодского Института социально-экономического развития территорий РАН [12]. Модель откалибрована таким образом, чтобы статистические данные, рассчитанные на множестве агентов, совпадали с реальными данными официальной статистики о численности населения каждого муниципального района и его половозрастном составе, а также о структуре занятости по видам деятельности, полу, возрасту и уровню образования работников с соответствующей дифференциацией заработной платы. Уровень недовольства агентов-людей в модели также соответствует данным о социальных настроениях жителей Вологодской области [12, с. 78].

Автоматами в модели являются только агенты-люди, а изменение характеристик агентов остальных типов может осуществляться пользователем-экспертом для имитации управляющих воздействий и для проведения сценарных расчетов.

С использованием модели «Губернатор» были проведены численные эксперименты по имитации динамики состояния населения Вологодской области, ее муниципальных районов и предприятий на их территории при разных значениях управляемых параметров модели.

## Примеры экспериментов

**Эксперимент 1.** Управление распределением доходов консолидированного бюджета между бюджетами двух уровней – бюджетом Вологодской области и бюджетами ее муниципальных районов. Управляемые параметры – доли налоговых поступлений, собираемых на территории муниципальных образований, остающихся в их бюджетах. Общая налоговая нагрузка на предприятия различных отраслей, а также на работающих жителей области не увеличивалась, но поступающие в виде налогов средства по-разному распределялись между бюджетами двух уровней.

Целью эксперимента было проследить влияние перераспределения налоговых поступлений на такие характеристики муниципальных районов, как самообеспеченность их бюджетов и бюджетная обеспеченность на 1 жителя.

Эксперимент показал, что хотя средний объем бюджета муниципальных районов, приходящийся на одного жителя, возрастает при увеличении доли налоговых поступлений, остающихся в бюджете муниципальных районов, но одновременно возрастает и дифференциация районов по этому показателю. Таким образом, с помощью только перераспределения налоговых поступлений не удастся добиться самообеспеченности муниципальных бюджетов.

**Эксперимент 2.** Управление бюджетными расходами на здравоохранение. Управляемый параметр – норматив бюджетных расходов на здравоохранение в расчете на одного жителя области. В модели здоровье агентов-жителей может принимать значения «слабое», «нормальное» и «отличное». Причем, первоначально заданное значение для агентов старше девятнадцати лет может убывать с увеличением их возраста, а также при сокращении норматива расходов бюджета на услуги по здравоохранению. Считается, что эти факторы вносят равный вклад в вероятность ухудшения здоровья агента.

Целью эксперимента было проследить влияние снижения норматива бюджетных расходов на такие интегральные характеристики, как доля практически здоровых среди занятого населения, а также удельный трудовой потенциал агентов-жителей. Для этого были проведены сценарные расчеты на периоде от 2004 до 2008 года при различных вариантах социальной политики.

Эксперимент показал, что снижение финансирования системы здравоохранения в модели приводит к ухудшению состояния здоровья населения. А поскольку показатель здоровья входит составной частью в трудовой потенциал агентов, то и среднее значение трудового потенциала (работоспособность агентов) также снижалось при ухудшении здоровья агентов.

**Эксперимент 3.** Управление взаимоотношениями организаций и их работников. Управляемые параметры – характеристики организационной культуры агентов-предприятий, такие как ценность творчества работников



и скорость их карьерного роста. В модели принимается, что вклад каждого агента-работника в общий результат пропорционален его относительному потенциалу, который зависит как от возможностей самого агента в качестве работника, так и от востребованности его активности и креативности со стороны его организации-работодателя [4].

Целью эксперимента было проследить влияние изменений в организационной культуре агентов-предприятий на такие интегральные характеристики этих предприятий и Вологодской области в целом как доля работников, удовлетворенных жизненными обстоятельствами; а также удельный относительный потенциал работников. Для этого на данных базового года были проведены расчеты при различных вариантах изменения характеристик организационной культуры отдельных организаций.

Эксперимент показал, что изменение организационной культуры промышленных предприятий, поощрение творческой активности их работников улучшает настроение этих работников, позволяет им более полно реализовать свои возможности и повысить производительность труда.

В заключение следует отметить, что разработанная мультиагентная модель представляет собой попытку воссоздания механизмов, обеспечивающих правдоподобную имитацию социально-экономических состояний макроэкономической системы на основе индивидуального поведения включенных в нее самостоятельных агентов. Это перекликается с подходом поведенческой экономики [13, 10] и соответствует принципам объяснения социального поведения, предложенным в [14]. Так, в модели выстроены цепочки зависимостей, обеспечивающие влияние состояния и трудовой деятельности агентов-людей на экономические и социальные показатели региона, например, такие как удельный трудовой потенциал работающих, доля здоровых и удовлетворенных жизнью жителей.

Таким образом, с помощью процедур расчета параметров бюджетов двух уровней, процедур расчета состояний агентов-жителей региона и моделирования их поведения в качестве работников, а также механизмов последующего агрегирования их трудовых потенциалов, выстраиваются цепочки влияния действий пользователя на выходные параметры модели. В частности, конструкция модели призвана наглядно демонстрировать зависимость интегральных показателей региона от бюджетных расходов на социальную сферу как инвестиций в человеческий капитал.

### Библиографический список

1. Бурков В.Н., Новиков Д.А. Теория активных систем: состояние и перспективы. М.: Синтег, 1999. – 128 с.
2. Бахтизин А.Р. Агент-ориентированные модели экономики / А.Р. Бахтизин. – М.: ЗАО «Издательство «Экономика», 2008. – 279 с.



3. Handbook of Research on Agent-Based Societies: Social and Cultural Interactions / Goran Trajkovski and Samuel G. Collins, editors. – New York: Information Science Reference Hershey, 2009. – 412 p.
4. Сушко Е.Д. Мультиагентная модель региона: концепция, конструкция и реализация / Препринт # WP/2012/292. – М.: ЦЭМИ РАН, 2012. – 54 с. (Рус.).
5. Холл Кэлвин С., Линдсей Гарднер. Теории личности / Холл К., Линдсей Г.; пер. с англ. – М.: Изд-во Института психотерапии, 2008. – 672 с.
6. Bloomquist K.M. A comparison of agent-based models of income tax evasion / Social Science Computer Review, 24 (2006), 4, 411-425 pp.
7. Симонов П.В., Ершов П.М. Темперамент. Характер. Личность. М., 1983.
8. Гулин К.А. Трудовой потенциал региона [Текст] / К.А. Гулин, А.А. Шабунова, Е.А. Чекмарева; под рук. д.э.н., проф. В.А. Ильина. – Вологда: ИСЭРТ РАН, 2009. – 84 с.
9. Камерон К., Куинн Р. Диагностика и измерение организационной культуры. – СПб.: Питер, 2001.
10. Делавинья С. Психология и экономика: результаты эмпирических исследований. Часть II. Общественно ориентированные предпочтения и нестандартные убеждения // Вопросы экономики, М., 2011. — №4 – С. 56-74.
11. Судаков К.В. Общие принципы построения поведенческих актов на основе теории функциональных систем // Системные механизмы поведения / Под ред. К.В. Судакова, М. Баича. М., 1990.
12. Амелин Д.Е. Местное самоуправление в региональном развитии / Д.Е. Амелин. – Вологда: ВНКЦ ЦЭМИ РАН, 2006. – 182 с.
13. Канеман Д., Тверски А. Рациональный выбор, ценности и фреймы // Психологический журнал. – 2003. – Т. 24. – №4. – С. 31-42.
14. Эльстер Ю. Объяснение социального поведения: еще раз об основах социальных наук [Текст] / пер. с англ. И. Кушнаревой; Гос. ун-т – Высшая школа экономики. – М.: Изд. дом Гос. ун-та – Высшей школы экономики, 2011. – 472 с.

Толстова Ю. Н., Москва

## Теория измерений как основа для выявления родства науки<sup>1</sup> и искусства

---

### Аннотация

В статье показывается сходство логических рассуждений, используемых при рождении новых ветвей математики, измерении в социологии и создании произведений искусства. Суть этих рассуждений отвечает основной идее теории измерений: поиске общего в изучаемых (изображаемых) объектах и построении алгоритма, отражающего (изображающего) именно это общее.

**Ключевые слова:** измерение, теория измерений, математика, искусство, живопись, эмпирическая система, математическая система

### Суть рассматриваемой проблемы. Основная цель статьи

Несмотря на бурное развитие в последние десятилетия методов анализа данных, позволяющих эффективно решать многие социологические задачи, а также разработку все новых и новых оригинальных и полезных для социологии методов моделирования социальных явлений, использование математического аппарата в эмпирических социологических исследованиях происходит относительно редко. Встает вопрос: почему? На наш взгляд, основной причиной служит бытующее среди социологов противопоставление друг другу социологии и математики с точки зрения заложенных в них способов познания. Привлечем на помощь анализ ситуации в искусстве и попытаемся снизить остроту этого противопоставления путем обоснования следующих положений:

- (а) гносеологической основой математики является выявление «общего» в разных изучаемых объектах, четкая формулировка этого «общего» (базирующаяся на логических рассуждениях) и анализ динамики его

---

<sup>1</sup> Говоря о науке, мы, в первую очередь будем иметь в виду социологию.

(«общего») развития; это же является основой любой науки вообще и социологии, в частности; четкое выражение этот процесс получил в теории измерений;

- (б) многие художники пытались в процессе своей деятельности делать то же самое: найти и логически выразить общие свойства изображаемых ими объектов (т. е. не сводили свое творчество к отображению уникальности); при этом результатом поиска, увенчавшегося успехом, была определенная формализация процесса творчества

- (в) в творчестве социолога часто реализуются идеи того же плана, социологи действительно в определенном смысле находятся между наукой и искусством.

Надеемся, что данная статья будет способствовать пониманию того, что само рождение математики — естественный процесс, возникающий при попытке человека изучать (наука), либо изображать (искусство) любые реальные объекты.

### **Две ипостаси математики: формальная и гносеологическая. Важность для социолога гносеологической стороны**

Все знают математику как науку, в рамках которой формальные объекты изучаются по формальным правилам. Но всегда ли математика была таковой? Когда-то, в глубине веков, были рождены понятия первых формальных объектов, рождены на основе наблюдения объектов вполне реальных, и выделения в них неких общих свойств<sup>1</sup>. Эти свойства отражали лишь часть бесконечного количества качеств, в реальности присущих наблюдаемым объектам. Именно благодаря такому отвлечению от реальности полученные «урезанные» объекты стали носить формальный характер. Они стали изучаться сами по себе, формализм стал наращиваться формализмом же, родилась математика в её формальной ипостаси, математика как наука, позволяющая изучать абстрактные соотношения между абстрактными объектами, казалось бы, оторванными от реальности, но в действительности отражающие какие-то стороны вполне реальных явлений, стороны, многократно повторяющиеся в разных объектах.

Попытки четко обрисовать выявляемое в разных объектах общее и попытаться проанализировать закономерности его появления и развития (уже безотносительно к конкретным объектам) путем изучения их изменений от объекта к объекту были предприняты уже на заре человеческой цивилизации. Поначалу выделяемые свойства объектов были относительно простыми. Примером может служить возникновение геометрии в Древнем Египте. Не сразу люди поняли, что измерять площадь участков одинаковой

---

<sup>1</sup> Выделение общих черт в рассматриваемых объектах осуществляется человеком в течение тысячелетий. Именно на основе такого выделения появились человеческие языки.

формы можно одним и тем же способом, не зависимо от того, находится ли этот участок в пустыне, в плодородной пойме Нила или в болоте. Еще несколько веков прошло прежде, чем геометрические объекты полностью формализовались, потеряли связь, скажем, с земельными участками. Это, как известно, было сделано в знаменитой работе Евклида (325-265 до н.э.) «Начала», служившей в течение более двух тысяч лет основным учебником по геометрии и образцом того, что такое математический трактат. Всеми было принято, что математика работает по формальным правилам с формально определенными объектами. «Интеллектуальные мучения» египетских распределителей земли были забыты.

Через 2-3 тысячи лет после рождения геометрии та же логика привела к рождению теории вероятностей. Примерно с 15 века ученые стали открывать законы, связывающие друг с другом частоты встречаемости разных «событий» при многократном осуществлении «эксперимента». Началось все с попыток понять, как действовать, чтобы выиграть игру в кости. Кончилось созданием теории вероятностей как строгой аксиоматической теории. Теперь о происхождении понятия вероятности, о том, что оно выросло из понятия частоты встречаемости каких-то «событий» в длинном ряду экспериментов, было забыто. Профессионалы математики – специалисты по теории вероятностей стали работать с формальными объектами по формальным правилам.

Для объектов, изучаемых современной математикой, иногда бывает весьма непросто сказать, от какой реальности отталкивалось само их возникновение. Однако многочисленные вопросы о выявлении принципов связи формальных построений с реальностью отнюдь не потеряли своего смысла. Напротив, жизнь ставит новые и новые задачи, связанные с желанием исследователя выделить в изучаемых объектах нечто общее, породить на этой основе формальные объекты и изучить их в «чистом» виде с целью получения на этой основе содержательных выводов. Отвечать на упомянутые вопросы тоже должна математика, однако здесь она проявляется в другой своей ипостаси, которую мы условно назовем гносеологической.

### **Теория измерений как основа гносеологической стороны математики (как формализация процесса отображения реальности)**

Если принять наше положение о том, что при научном или художественном познании мира человеку свойственно выделять в разных объектах нечто общее и именно это общее изучать (отображать), то рождение математики следует воспринимать как естественный шаг, олицетворяющий соответствующий этап процесса познания. Из глубины веков тянулась линия развития новых и новых ветвей математики. И лишь в середине XX века была предпринята попытка перейти на метауровень: найти общее во всех процессах выделения общего – и в участках земли, и в игре в кости. Появилась теория измерений – математика самого процесса отображения.

Мы не будем подробнее говорить о теории измерений, сославшись на работу [1], где речь идет о понятиях, ставших классическими. Отметим лишь, что ключевым моментом теории является понимание измерения как моделирования реальности. В основу моделирования положены понятия эмпирической и математической систем (ЭС и МС). Измерение – это отображение первой во вторую. Именно формируя ЭС, исследователь (художник) выражает собственный взгляд на то, что он хочет видеть в изучаемом (изображаемом объекте).

### Искусство (живопись): поиск общего в изображаемых объектах

Рассмотрим с интересующей нас точки зрения несколько отдельных эпизодов в развитии живописи.

#### Древний мир

Во многих древних культурах отображаемые в процессе художественного творчества черты рассматриваемых объектов были канонизированы. Даже если изображаемый объект вроде бы должен был принадлежать к рассматриваемому типу, но не обладал требующимися чертами, при изображении этого объекта такие черты ему приписывались. Вспомним Древний Египет с привычным для его художников изображением фараонов (и не только), когда торс человеческой фигуры изображался анфас, а голова и ноги – в профиль. Своеобразные принципы были свойственны ассирийско-вавилонскому искусству. На оставшихся до нашего времени скульптурных и живописных портретах изображались, в основном, цари. А царь – фигура сакральная, индивидуальные черты царя никогда не отмечались художниками. Зато на портретах в обязательном порядке наличествовали преувеличенная мускулатура обнаженных рук и ног, пышные длинные волосы и борода. Это отвечало господствующему в Ассирии-Вавилонии культу красоты. Нетрудно видеть, что реализация подобных принципов создания портретов очень близка к формализации. В наше время портреты, подобные рассмотренным, наверное, прекрасно мог бы создавать компьютер. Упрощая ситуацию, можно сказать, что ЭС здесь – люди, рассматриваемые как носители определенного рода красоты. При этом считается, что красивее тот, кто более знатен. Вместо элементов МС – портреты людей.

#### Развитие символики смыслов

Произведения, лежащие в рамках очень многих течений в живописи в середине второго тысячелетия начинают наполняться разными смысловыми подтекстами. Развивается символика смыслов, скрывающихся за теми или иными объектами. Так, створка раковины – символ богини, поскольку призвана скрывать в себе жемчужину. Много смысловых зна-



чений имеет яйцо, но все так или иначе сводятся к тому, что оно содержит в себе целый мир и дает новую жизнь. Красный коралл символизирует искупительную жертву Христа. Горный хрусталь — душевную чистоту верующего, ящерица — символ зла. Яблоко (груша, персик, апельсин) в руке Христа (или рядом с тронном богоматери) — символ человеческих грехов. Кувшин — символ крещения, пальмовая ветвь — стойкости верующих, колонна — то же. Собака, грызущая кость — символ алчности. Подобные предметы могут быть видны не сразу. Но тщательный просмотр картины может погрузить зрителя в мощное смысловое поле.

Во второй половине XIX века соответствующая тенденция нашла широчайшее развитие в новом течении в искусстве — символизме (сказанное выше означает, что, по существу, искусство всегда лежало в русле символизма).

Наличие такого количества смысловых условностей говорит о близости математики к живописи. Представляется возможной компьютеризация процесса вычленения на картине разных предметов с приписыванием им смыслов (из заранее заданного списка) и созданием далее общего «текста» картины, объединяя разные «единицы» скрытых в картине смыслов. Последний шаг нетривиален и напоминает выявление связей между переменными в анализе данных.

Вероятно, таким образом можно отражать на полотне и смысловой мир отдельного человека.

### Художники-математики. Идея перспективы

Мысли о возможности формализации работы художника неоднократно появлялись у известных художников прошлого. Самый яркий пример — творчество Леонардо да Винчи (1452-1519). Известно, что свои способности естествоиспытателя он пытался использовать и в живописи. Так, он внедрил в работу художника понятие перспективы и перевел соответствующую работу на язык математики. Мощная иллюстрация — знаменитая картина Леонардо «Тайная вечеря». Для нас важно отметить, что желание художника сохранить определённые качества рассматриваемых объектов (свойства ЭС, в данном случае это — то, что изображаемые объекты находятся на разном расстоянии от зрителя) влечет создание методов такого сохранения (алгоритм отображения ЭС в МС, в данном случае — известные геометрические построения, позволяющие строго определять размер изображаемого объекта в зависимости от того, на каком расстоянии от переднего плана картины этот объект находится

Можно было бы назвать еще около десятка фамилий художников, бывших математиками и пытавшихся применить в художественном творчестве свои математические знания. Это, на наш взгляд, подтверждает то, что логика, служащая гносеологической основой математики, в значительной мере свойственна и другим способам познания человеком окружающего его мира. Формирование четких представлений об ЭС дает возможность адекватно перейти к четкой МС. Аналогия между измерением и живописью представляется очевидной.



На примере с перспективой можно показать также, что методы формализации отражаемого фрагмента реальности, представляющиеся адекватными одному художнику, могут не быть принятыми другим. Смотря на одну и ту же реальность, разные люди видят разные ЭС. И для некоторых ЭС теория перспективы Леонардо может оказаться совсем непригодной. Во многих культурах более значимые личности (понимание значимости могло быть разным) изображались более крупными, чем менее значимые. Это было в Древнем Иране, где простой воин имеет гораздо меньший рост, чем стоящий рядом с ним военачальник, в Японии XVIII века, где знаменитые гейши на рисунке гораздо крупнее своих учениц и помощниц, в русской миниатюра первой половины второго тысячелетия, в русской иконописи.

### **Рождение «измов»: выделение общего в наблюдаемых объектах и попытка четко его выразить**

Следующим шагом в интересующем нас направлении явилось то, что в середине второго тысячелетия художники сами начали говорить о том, что именно они стремятся отражать в своих картинах: родилось огромное количество разных «измов»: маньеризм, спиритуализм, классицизм, неоклассицизм, романтизм, реализм, сюрреализм, импрессионизм, экспрессионизм, дивизионизм, футуризм, символизм, пуантилизм, кубизм и т. д.

Рассмотрим один из моментов творчества крупнейшего представителя *импрессионизма* Клода Оскара Моне (1840-1926). В 1877 году он потряс парижскую публику, представив семь работ на тему железнодорожной станции, показав тем самым, что самые прозаические вещи могут заключать в себе удивительную поэзию. Общество считало, что данная тема далека от искусства. А тут — такой шок! Вероятно, если бы какой-либо художник взялся за изображение паровоза в той интерпретации, которая господствовала в обществе до Моне, то, наверное, в качестве ЭС у него бы выступал набор технических, противостоящих человеку качеств: мощь, жесткость, опасность и т. д. «Портретом» паровоза послужила бы некая мощная фигура, надвигающаяся на человека. У Моне же совсем другая ЭС: паровоз характеризуется быстротой, туманностью, расплывчатостью и т. д. И в итоге на картинах Моне паровоз — это пушистое, весело бегущее вперед создание, отнюдь не вызывающее страха.

Идея о том, что произведение искусства зависит не только от изображаемого объекта, но и от изображающего его субъекта, заложена уже в самом понятии ЭС: чтобы её сформировать, мы, исходим и из того, что знаем об объекте, и из своих личностных качеств и субъективных взглядов на мир. Конечно, от того же зависит и выбираемый аналог МС, и способ отображения в него построенной ЭС. Проведём аналогию с анализом данных.

Известно, что для решения практически любой социологической задачи существует множество разных методов. И каждый, вообще говоря, дает свой результат. Более того, разные исследователи будут получать

разные результаты даже при использовании одного и того же алгоритма, если, к примеру, по-разному разобьют диапазон измерения какого-либо признака на интервалы

Перейдем к неоимпрессионизму и рассмотрим лежащий в его рамках *пуантилизм* (дивизионизм), основателем которого считают Жоржа Сёра (1859-1891). Отличительной особенностью этого направления является то, что изображение создаётся отдельными точечными мазками «чистого» цвета, близко положенными один к другому. Краски не смешиваются на палитре: нужный оттенок создаётся благодаря выверенному соседству локальных цветов. Достичь этой точности художнику помогло изучение новейших открытий в оптике. Вблизи картина, написанная таким «научным» методом, являет собой хаотичное месиво, зато на расстоянии мазки, выполненные «чистыми» красками, сливаются в соответствии с законами восприятия цвета и выявляют задумку художника

Напрашивается аналогия с тем, как с помощью небольшого количества отдельных значений признаков составляется огромное количество свойств респондентов, детерминирующих самые разные виды их поведения (соответствующие наборы значений разных признаков называются взаимодействиями). То общее, что ищется здесь в разных объектах, – это те исходные наборы «чистых» цветов, присущих каждому изображаемому объекту.

В большей степени, чем у любого другого художника его времени, у Сёра чувство подчинено рассудку. Ж.Сёра отображает то, что можно назвать ЭС, в картину, похожую на то, что делали импрессионисты, но способ отображения – совершенно другой и, как следствие, картина производит другое впечатление на зрителя. Проводя аналогию с социологией, можно сказать, что можно по-разному измерять, скажем, связь между двумя переменными, получая при этом разные содержательные выводы.

Еще один пример, теперь из области русского искусства – «лучизм», главными представителями которого являются **Михаила Федоровича Ларионова** (1881-1964) и **Натальи Сергеевны Гончаровой** (1881-1962).

«Лучисты» полагали, что образ предмета в глазах человека – это результат пересечения лучей, исходящих от разных сторон этого предмета. А такая картина вполне укладывается в теорию лучизма. И, на наш взгляд, например, замечательная картина Гончаровой «Желто-зелёный лес» (1913) прекрасно отражает лес как пересечение огромного количества зеленых оттенков цвета, блеска листьев, сияния отдельных кусков ярко синего неба, солнечных пятен на траве, темных «пещер» под густыми ёлками и т. д.

В каждом рассматриваемом объекте выделяется одно свойство: поток исходящих от него лучей (это можно рассматривать и как многомерную величину, отдельными компонентами которой являются лучи разного цвета, разной интенсивности, разной направленности). Отображение каждого предмета в поток лучей – это «измерение». Способы анализа этих потоков (их сравнение, создание комбинаций) – «математические» методы.

Нетрудно видеть, что реализация идей лучизма тоже в чем-то схожа с реализацией алгоритмов поиска взаимодействий (два потока лучей могут сформировать новый образ, не похожий на эти потоки). Кроме того здесь

могут использоваться ассоциативные методы для выявления связи разных сочетаний лучей с теми эмоциями, которые они производят на человека и т. д.

### Может ли искусство быть полностью формализовано?

Наш ответ: вряд ли. Какие бы сложные свойства ни научились художники выделять в интересующих их объектах, какие бы методы их изображения ни придумали, всегда найдутся еще более сложные и тонкие свойства, не поддающиеся четкому выражению с помощью кисти. Как формализовать улыбку Моны Лизы? Несмотря на свое стремление научиться математически чётко выражать в живописи интересующие его явления, Леонардо все же не сводил живопись к тому, что можно четко выразить. В знак того, что в человеке всегда остается что-то не отраженное в его портрете, художник-ученый ввел метод «сфумато» - смягчение линий рисунка, придание линии расплывчатости. Формы смягчались обволакивающей их дымкой, благодаря чему сквозь ясность мира как бы проступала его таинственность.

Вероятно, искусство символизирует ту бесконечность мира, наличие которой мешает допускать, что мы когда-нибудь сможем найти все возможные элементы общности между любыми изучаемыми (отражаемыми) объектами. Хотя как знать ... Существует же в Музыкальной академии имени Гнесиных кафедра компьютерной музыки!

Нам бы хотелось, чтобы читатель не воспринимал эту статью как игру в формалистику. Конечно, мы вычленим в творчестве художников нечто, о чем сами они не говорят и, вероятно, мало думают. Но такое вычленения является необходимым, если мы хотим изучить логику познания человеком окружающего его мира, что, с нашей точки зрения является необходимым условием обеспечения адекватности такого отображения (с учетом возможности разного понимания самого понятия адекватности). Надеемся, что статья будет способствовать пониманию того, что математика нам задает некую генеральную линию в способах познания окружающего нас мира, результатом чего явится более широкое её использование в социологии.

### Библиографический список

1. Толстова Ю.Н. Измерение в социологии. М.: ИДУ, 2009.

Шведовский В. А., Москва

## Социальные прогнозы на топологических скрытых цепях Маркова

---

### Аннотация

Анонсируется методика регулярных построений социальных прогнозов, использующая применение топологических цепей Маркова к обобщённой модели К.Левина жизненного пространства типичных личностей, репрезентирующих социальные группы.

**Ключевые слова:** образ жизни, сфера образа жизни, жизненное пространство личности, типичная личность, образ сферы образа жизни, цепь Маркова, двойная (скрытая) цепь Маркова, индикаторная функция, социальный прогноз

### Введение

В социальных науках давно утвердился социальный прогноз, опирающийся на аппарат марковских цепей: это и прогнозирование социальной мобильности, рождаемости и численности различных социальных групп и т. д. В данной работе предлагается методика регулярных построений социальных прогнозов для социальных процессов, детерминированных устойчивым образом жизни больших социальных общностей, репрезентируемых основным рядом типичных личностей этого социума.

В работе [8] на основании данных социологического исследования РГСУ-2010 при использовании аппарата цепей Маркова был построен прогноз численностей основных групп иммигрантов в Москву до 2015 года. Каждой из этих групп сопоставлялся соответствующий ей образ жизни, что позволяло репрезентировать эти характерные социальные группы соответствующим им спектром типичных личностей, их «социальных портретов». Таким образом, выдвигается методический приём замены изучения поведения социальных групп или общностей, детерминированного их траекториями по сферам образа жизни, на изучение траекторий между образами сфер жизнедеятельности - ячейками жизненном пространстве типичных для этих групп или общностей личностей. В качестве модели жизненного пространства личности используется обобщение модели К.Левина [9]. Сама концепция личности выстраивается в контексте идей структурной

стыковки, самореферирования и автономности Н.Лумана [4], и ориентирована, прежде всего, на моделирование механизма прогноза численности носителей веры: в верховное существо, в будущее, в себя и т. п.

### Основные идеи выстраиваемой методологии

- Прогноз динамики носителей веры для группы подобен прогнозу степени веры для её типичной личности; траектории личности в жизни аналогичны траекториям в её жизненном пространстве (ЖПЛ).

- Собственное вероятностное поведение личности, т. е. исключая подражательное поведение и давление фактов конкретной социальной ситуации, определяется её установками, как правило, амбивалентными.

- Установки личности формируются её устойчивым образом жизни, который задаётся набором сфер образа жизни и структурой связей между ними, что представимо оргграфом  $G(X, U)$  связей  $u_{ij}$  между сферами  $x_i$  образа жизни – с.о.ж., т. е.  $u_{ij} = (x_i, x_j)$ .

- Образ жизни представим однородным б-орграфом, организующим трафики переходов без пересечений на 2- мерной компактной, ориентированной поверхности рода  $p \geq 2$ . В силу т. Пуанкаре  $\int K ds = 2-2p$  такая поверхность обладает отрицательной гауссовой кривизной, т. е.  $K_c < 0$ , что составляет основу для возникновения гиперболической д.с. (г.д.с).

- С позиций символической динамики [1] эта компонента фазового пространства д.с. есть его конфигурационное подпространство  $Q$ , порождающее гиперболическую динамическую систему – г.д.с., для фазового пространства которых академик Синай Я.Г. доказал существование марковских разбиений и обосновал построение топологических цепей Маркова.

- Конфигурационное подпространство  $Q$  отождествляется с топологическим обобщением [9] в модели К.Левина жизненного пространства личности (ЖПЛ):  $L$  – жизненное пространство,  $E$  – психологическая среда,  $I$  – информация,  $p$  – сама личность с её ячеистой структурой; ячейками являются образы сфер жизнедеятельности ( $T$  – труд,  $B$  – быт,  $K$  – культура,  $O$  – образование,  $C$  – непосредственное общение,  $P$  – общественно-политическая деятельность);

- каждая сфера представима бюджетом времени занятости в ней  $\Delta t_{ii}$ , и эти времена сопоставляются с недельным бюджетом, равным 168 часам; кроме того, в меру  $i$ -ой сферы о.ж.  $\mu(q_i)$  включена оценка её сложности  $H_i$ , по аналогии с оценкой сложности рабочего места [10], т. е.  $\mu_i = \mu(\Delta t_{ii}, H_i)$  [11].

Движение фокуса психической активности по ячейкам структуры личности есть процесс её самоидентификации и самореферирования - считывания информации –  $I$ , необходимой для принятия решения. Считывается, прежде всего, баланс отрицательных  $S^-$  и позитивных  $S^+$  фактов, как ослабляющих, так и усиливающих сложившиеся у личности установки, ценностные ориентации, отстаиваемые позиции. Эти факты,



предстоящие как конкретные объекты внешнего мира, в ЖПЛ выступают как их референты (денотаты)  $R(S^-)$ ,  $R(S^+)$ , как правило, в эмоционально окрашенной форме, и личность «работает» с ними как с внутренними стимулами + или - реакции. В начальном приближении веса этих референтов считаются одинаковыми и неизменными.

У разных типических личностей свой спектр траекторий движения по образам с.о.ж. Их число особенностей определяется по схеме совмещения итогового орграфа ТМЦ с 2-поверхностью (компактной и ориентированной, рода  $p \geq 2$ ), являющейся носителем точек конфигурационного пространства  $Q$ , т. е. подпространства фазового пространства исследуемых г.д.с. Этим особенностям отвечают вершины б-орграфа, ибо именно в них не определён вектор скорости г.д.с., что и позволяет вводить вероятности перехода в ТМЦ.

## Основные понятия модели

В основании модели ТМЦ лежит представление о двойной (скрытой) марковской цепи [5]. Ниже приведена диаграмма (см. рис. 1), из которой видно соотношение между наблюдаемыми признаками  $y(t)$  и латентными переменными  $x(t)$  для скрытых состояний. При этом значение скрытой переменной  $x(t)$  (в момент времени  $t$ ) зависит только от значения скрытой переменной  $x(t-1)$  (в момент  $t-1$ ). Это и есть свойство Маркова простой случайной последовательности состояний, описываемых переменными  $x(t)$ . Важно отметить, что в одно и то же время  $t$  значение наблюдаемой переменной  $y(t)$  зависит только от значения скрытой переменной  $x(t)$ .

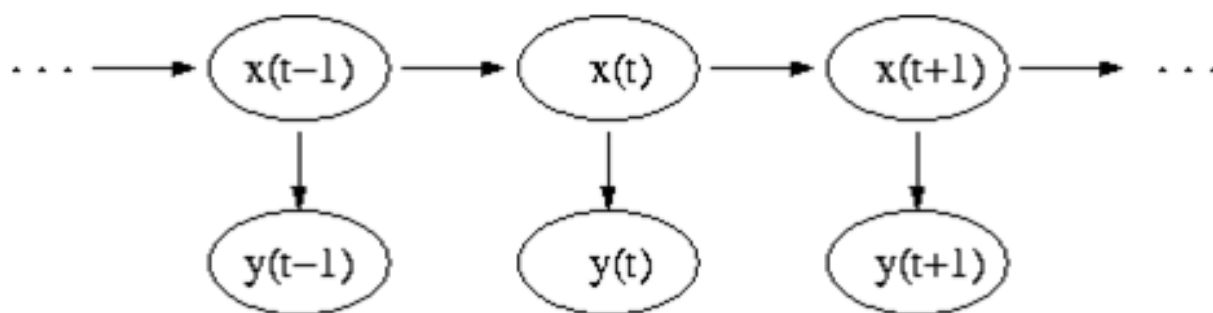


Рис. 1

Вероятность увидеть последовательность  $Y = y(0), y(1), \dots, y(L-1)$  длины  $L$  равна

$$P(Y) = \sum_x P(Y | X)P(X), \quad (1)$$

здесь сумма считается по всем возможным последовательностям скрытых узлов состояний и этих возможных последовательностей очень велико. При этом  $X$  в качестве однородной марковской цепи рассчитывается, опираясь на матрицу переходных вероятностей  $A = \{a_{ij}\}$ , где  $a_{ij} = p(q_{t+1} = x_j | q_t = x_i)$  – вероятности перехода из  $i$  в  $j$ , не зависящие от времени, а  $\sum_j a_{ij} = 1$ . Расчёт  $a_{ij}$  осуществляется в соответствии



с теорией ТМЦ, т. е. разбиения на ячейки жизненного пространства личности, реализованного в виде 2-мерной ориентированной компактной поверхности [1, с.209].

Вершины орграфа, представляющие эти ячейки жизненного пространства, составляют алфавит символов в порождаемой символической динамике -  $\mathfrak{L}\{T, B, K, O, C, \Pi\}$ . Символическая динамика [1, с.196] представима пространством  $\Sigma$  бесконечных последовательностей  $\{\omega_n\}_{-\infty}^{\infty}$  из символов алфавита  $\mathfrak{L}$ , например, ...Б-Т-С-Т-Б-О- Б-Т-С-Т-Б-К-...и оператора сдвига на этих последовательностях  $\sigma$  так, что  $\sigma\omega_n = \omega_{n+1}$ , где получаемые последовательности отвечают траекториям г.д.с., - см. Рис.2. Основная догма символической динамики представлена из [1, с. 197] формулой (2):

$$\Psi(x) = \omega^- = \{\omega_n\} \leftrightarrow f^n x \in E\omega_n \leftrightarrow x \in \bigcap_n f^{-n} E\omega_n, \quad (2)$$

определяющей соответствие между последовательностями отображений  $f^n$  точки  $x$  – носителя координат конфигурационного пространства, принадлежащей ячейки разбиения этого пространства  $E\omega_n$ , и соответствующим ей символом из алфавита  $\mathfrak{L}$ , помеченным номером отображения  $n$ .

На рис. 2 показано упрощённое разбиение всего  $X$  на 4 сферы образа жизни, которое допускает ещё полный граф, натягиваемый без пересечений рёбер, на замкнутую ориентированную компактную поверхность рода 0, т. е. сферу. Этот граф представляет собой остов тетраэдра с вершинами Т, Б, К, С.

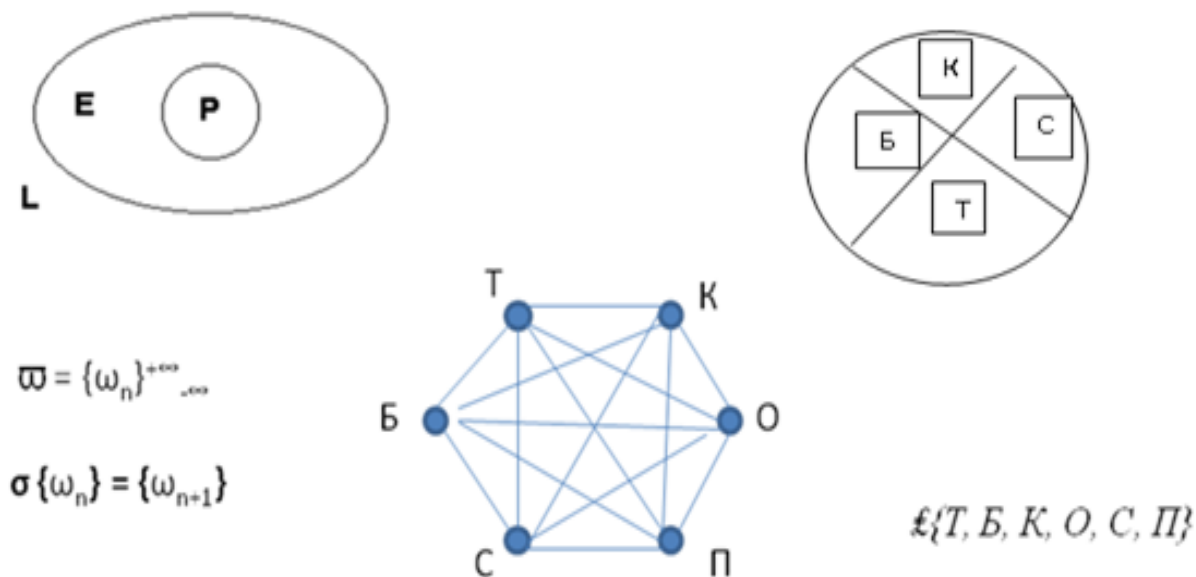


Рис. 2. Представление ЖПЛ и его разбиения на ячейки активности – образы сфер образа жизни

На рис. 2. показан 6-вершинный полный граф, который по т. Куратовского-Понтрягина уже не является планарным, и требуется найти для него замкнутую ориентированную 2-поверхность, с соответствующей ей родом. Для упрощённых 2-х и 3-х графов даны отвечающие им матрицы на рис. 3

$$A_2 = \left\{ \begin{array}{ccc} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{array} \right\}$$

Рис. 3. Матрицы переходных вероятностей цепей Маркова для 2-х, 3-х сферных (труд-быт-общение) моделей образа жизни -  $\lambda$

## Постановка задачи

Требуется представить движение фокуса психической активности личности в её жизненном пространстве – ЖПЛ как однородную эргодическую марковскую цепь на б-орграфе образов сфер жизнедеятельности – «скрытые, латентные состояния», отражающих в ЖПЛ реально существующие сферы образа жизни данной типичной личности. Задав стартовый вектор  $\omega$  альтернативных её свойств, например, религиозность/нерелигиозность, а также построив на основе значений показателей индикаторной функции в каждой из сфер образа жизни (наблюдаемые переменные) агрегированные оценки проявлений указанных свойств личности в виде матрицы переходных вероятностей «2x2», получить прогноз динамики этих свойств.

## Построение топологической марковской цепи.

Для последовательности  $A_1(s), A_2(s), \dots, A_k(s)$  – несовместимых событий ( $s$ - номер испытания) рассматривается простая цепь Маркова без поглощений  $0 < p_{ij} < 1$  ( $\sum_j p_{ij} = 1$ ), где вероятность  $P(A_i(s+1)) = f(A_j(s))$ , при этом матрица  $\mathbf{P}$  переходных вероятностей  $P_{ij}$ : (3), а вектор вероятностей  $k$ -го состояния  $P(k) = P(0) * \mathbf{P}^k$ , где  $P(0)$  – вектор начального состояния.

Условие (по Колмогорову) существования в представленной выше постановке задачи, стационарного марковского случайного процесса, отвечающего однородной, т. е. не зависящей от времени, марковской переходной функции (непрерывный случай) или матрицы (дискретный случай, как у нас -  $p_{ij}(k) = p_{ij}$ ) выглядит так: необходимо и достаточно, чтобы нашелся такой класс  $Q$  сообщающихся состояний  $Q \subset X$ , что математическое ожидание времени попадания из  $q_1 \in Q$  в  $q_2 \in Q$  было конечно для любых  $q_i \in Q$ . Технологические детали по идентификации цепей Маркова см. [8, с.27].

## Вывод расчётных формул для двойной цепи Маркова на базе ТМЦ

Ключевым в методологии построения двойной ТМЦ являются три момента:

1. Определение индикаторной функции –  $b_j(d_t)$  – вероятность в  $j$  –м состоянии иметь наблюдаемое свойство  $d_t$  в момент времени  $t$ , связывающей сферы жизни с предметом исследования, например, с уровнем религиозности.

2. Генерирование разных моделей образования вероятностей в матрице переходных вероятностей, в которой учитывается индикаторная функция.

3. Повторная «перенормировка» переходной матрицы вероятностей с учётом индикаторной функции.

В качестве первого примера приводится «Смена ценностных ориентаций женщин России фертильного возраста», выявленной в ходе социологического опроса [2]. Среди множества вопросов анкеты имеется такой критериальный вопрос (см. таблицу 1).

Таблица 1

Ответы (по годам) матерей девочек - подростков на вопрос «с чем связываете будущее своей дочери?», %

	1990	1995	2001	2002	2005
С удачным замужеством ( $\psi$ )	64	46	27	26	27
С хорошей работой ( $\phi$ )	19	36	64	62	62

В качестве второго примера приводится «Общемировая закономерность спада степени распространённости религиозности по мере роста уровня образованности населения», например, горожане Санкт-Петербурга верующими себя считают среди лиц, имеющих высшее и н/в образование – 52%, среди лиц со средним образованием – 67%, с неполным средним – 88%. Для целей формирования индикаторной функции к данным [3] применён закон Ципфа (5) (см. таблицу 2).

Таблица 2

Распределение занятости мужчин по уровню полученного образования в областях РФ

	Ярославская область	Воронежская область	Башкортостан
	1	2	3
Рабочие (Р)	38	56	54
ИТР	26	38	49
Интеллигенция (И)	3	3	6

$$f_r(i) = C(i) / r^{\alpha(i)} \quad C(1) = 0.38; \alpha(1) = 0.6 ; r - \text{ранг группы (5)}$$

Опираясь на этот закон, определена степень религиозности занятого населения в Северо-Западном федеральном округе, и для него посчитан среднесрочный прогноз этой степени.

Уточнение: исследователи [3] отмечают существенную разницу между кругом респондентов, относящих себя к той или иной конфессии, например, православной, и кругом тех, кто отнёс себя к верующим в Бога. На наш взгляд, самоидентификацию с православием осуществляют не только «воцерковленные» верующие, но и широкий круг респондентов - приверженцев православной культуры. Например, имеет распространение такая самоидентификация как «православный атеист».

На наш взгляд, такое положение вещей связано с тем, что церковь является всего лишь одним из депозитариев информационно-культурных кодов, социокультурной генетики данного социума [12], наряду с другими воплощениями его социальных институтов, например, семьёй.

Влияет на динамику приверженцев религии и половая принадлежность. Так в те же 90-е годы, когда стабилизировалась численность верующих, «чаще называют себя православными женщины: их число составило 66%, мужчин – 43%» [7].

В итоге в современной социологии религии уживаются следующие оценки степени религиозности [3]: 1) по самооценке; 2) по статистике ответивших о частоте соблюдения религиозных ритуалов; 3) по конфессиональной самоидентификации. В данной работе, с учётом этих сложностей, решено измерять степень религиозности по двухбалльной шкале (верующий, неверующий). К неверующим отнесены ответившие «нет ответа, не знаю и т. п.», а также «сомневающиеся». Основной формулой, позволяющей вычислять вероятность нахождения наблюдаемых свойств, т. е. данных  $D$  при условии модели  $\lambda = (A, B, \pi)$ , является формула, в которой учитываются скрытые состояния, т. е. переменные  $Q$  [5]:

$$p(D | \lambda) = \sum_Q p(D | Q, \lambda) p(Q | \lambda), \quad (6)$$

где  $p(Q | \lambda)$  – вероятность последовательности  $Q = q_1 q_2 \dots q_T$  при условии выбранной модели  $\lambda = (A, B, \pi)$  – см. выше Рис.3, т. е. выбрав в качестве первого элемента последовательности  $q_1$ , перейдём в состояние  $q_2$  и т. д. до состояния  $q_T$ . Каждая такая последовательность считается как отдельное событие и необходимо просуммировать по всем таким событиям для изучаемой типичной личности с учётом вероятности иметь искомое свойство в каждом из состояний последовательности:

$$p(D | \lambda) = \sum_Q p(D | Q, \lambda) p(Q | \lambda) = \sum q_1 q_2 \dots q_T b_{q_1}(d_1) \dots b_{q_T}(d_T) \pi_{q_1} a_{q_1} q_2 \dots a_{q_{T-1}} q_T \quad (7).$$

При расчёте  $p(D | \lambda)$  возникает несколько проблем, в частности:

1. выбора модели - см. выше рис.3,
2. выбора начального состояния, т. е. сферы  $q_1$  из набора сфер (Т, Б, К, С, О, П),
3. определения последовательностей состояний  $q_1 q_2 \dots q_T$ .

В соответствии с постановкой задачи независимо от выбранной для описания движения фокуса активности внутри жизненного пространства типичной личности марковской модели  $\lambda_i$ , где  $i = 2, 3, 4, 5$  и 6 итоговая марковская цепь описывается матрицей «2 x 2», т. е. выглядит как модель  $\lambda_2$ . Таким образом, итоговая матрица переходных вероятностей, в общем виде, выглядит:

$$P = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix} \quad (9)$$

Удобство этой матрицы для марковских цепей заключается в том, что в случае, когда  $|p_{00} + p_{11}| < 1$ , то вычисление часто необходимое для получения прогноза через  $n$  – временных интервалов сводится к , что легко считается для  $n \rightarrow \infty$  как  $P^n \rightarrow x$  (10)

Используя полученное выражение и формулу, определяющую текущее значение  $P^n$ , можно получить оценку второй систематической ошибки метода, а сравнив её с ошибкой измерений  $\delta_{ij}$ , можно получить  $n$ :

$$n = \log(\max \delta_{ij}) / \log(|p_{00} + p_{11}| - 1) / (2 - p_{00} - p_{11}) \quad (11)$$

В случае, когда условие  $|p_{00} + p_{11}| < 1$  не выполняется, и известно лишь, что  $0 < p_{01}, p_{10} < 1$ , то можно вывести другое выражение для  $P^n$ :

$$P^n = \frac{1}{p_{01} - p_{10}} \begin{pmatrix} p_{10} + p_{01} \times (p_{00} - p_{10})^n & p_{01} \times (1 - (p_{00} - p_{10})^n) \\ p_{10} \times (1 - (p_{11} - p_{01})^n) & p_{01} - p_{10} \times (p_{11} - p_{01})^n \end{pmatrix} \quad (12)$$

Из формул (10), (12) вытекает, что в пределе для  $n \rightarrow \infty$  стабилизированная положительно определённая матрица переходных вероятностей приобретает в качестве своих одинаковых строк компоненты вектора стационарного распределения, не зависящего от вектора начального распределения вероятностей. Эта матрица, полученная из формулы (10) тождественна матрице, получающейся из (12) в силу того, что  $0 < p_{01}, p_{10} < 1$ , для этого достаточно вспомнить, что  $1 - p_{11} = p_{10}$ , а  $1 - p_{00} = p_{01}$ .

## Классификация типичных цепочек «сфер образа жизни» в жизненном пространстве личности

Для использования индикаторных функций, формирующихся, в том числе, и на основе таблиц 2 и 3, приведём классификацию типов личностей. Ниже в таблице 4 представлены приведённые ранее цепочки с.о.ж., соотнесённые с известными из ранее проведённых социологических исследований типами трудового поведения.

Таблица 4

Варианты цепочек сфер образа жизни  
в жизненном пространстве личности (фрагмент)

Социальные портреты типичных личностей	Трудоголики (1)	Высокая оплата – высокий мотиватор труда (2)	Не высокая, но постоянная зарплата, - мотиватор монотонного труда (3)	Не интегрированные в сферу труда (4)
Распространённость - sd	10	46	34	10
Траектория цикла 1	ТБТСТБТКБ – ж БТСБТКБТ - с	ТБТСТБТБ (в основном мужчины)	БТБСТСБ (в основном женщины)	БСТСБ
Траектория цикла 2(учащийся)				ОБОСОБОКБ
Траектория цикла 3(учащийся, с укор. рабочим днём)			ОБОСТБОБ	

## Пример построения расчётного алгоритма

Предполагается, что приведённый набор последовательностей вместе с «близкими»<sup>1</sup> к каждой из 4-х групп, т. е. образующей в совокупности долю  $sd(i)$ , образует всё пространство событий.

На основании приведённых выше формул (1), (2), (5-9), а также Таблиц 2-4 выводится алгоритм расчёта вероятностей  $p_{00}$  и  $p_{11}$ , где  $p_{00}$  – вероятность иметь признаки религиозности,  $p_{11}$  - вероятность не иметь признаки религиозности:

$$p_{00} = sd(1) * p_1(D | \lambda) + sd(2) * p_2(D | \lambda) + sd(3) * p_3(D | \lambda) + sd(4) * p_4(D | \lambda) \\ = sd(1) * b_T(fr(1)) \dots b_B(IndB) \pi(T) a_{TT} a_{TB} a_{BT} \dots a_{KB} a_{BB} + \dots$$

$$sd(4) * b_O(IndO) * b_B(IndB) \dots b_B(IndB) \pi(B) a_{OO} a_{OB} a_{BB} a_{BO} a_{OO} a_{OC} a_{CC} a_{BB} \\ (13).$$

<sup>1</sup> «Близость» определяется количеством символов, отличающих эталонную последовательность (Табл.4) от иной учтённой.



Аналогично  $p_{00}$  вычисляется  $p_{11}$ :

$$P_{11} = sd(1) * p_1(D^1 | \lambda) + sd(2) * p_2(D^1 | \lambda) + sd(3) * p_3(D^1 | \lambda) + sd(4) * p_4(D^1 | \lambda) \quad (14)$$

Последовательности  $D^1/D$  задают отсутствие / наличие религиозности.

Стабилизировавшуюся матрицу переходных вероятностей получаем (15) согласно формуле (10):

$$P = \frac{1}{2 - (\sum_1^4 sd(i)p(D|\lambda)_i) - (\sum_1^4 sd(i)p(D^1|\lambda)_i) \times \times} \quad (15)$$

где  $P$  - стабилизированная матрица переходов однородной простой эргодической цепи Маркова [4].

## Вычислительные эксперименты с двойными цепями Маркова и обсуждение результатов

Сначала рассматривался вариант образа жизни с двумя сферами жизнедеятельности (труда и быта) (см. рис. 4).

$$R1^2 = \begin{pmatrix} 0.496 & 0.504 \\ 0.487 & 0.513 \end{pmatrix}$$

$$R1^5 = \begin{pmatrix} 0.492 & 0.508 \\ 0.492 & 0.508 \end{pmatrix} \quad R0 := \begin{pmatrix} 56.6 \\ 43.4 \end{pmatrix} \quad R1 \cdot R0 = \begin{pmatrix} 49.889 \\ 49.89 \end{pmatrix}$$

$$R1^5 \cdot R0 = \begin{pmatrix} 49.89 \\ 49.89 \end{pmatrix}$$

Рис.4. Первый фрагмент рабочего поля с итоговым результатом прогноза религиозности в России (С-ЗФО)

Результат, представленный на Рис.4., полученный с использованием методики двойной ТМЦ, нашёл своё подтверждение в выводе Европейского социального исследования о стабильном сосуществовании в ближайшей перспективе 49% от всего населения России последователей религий с 48% нерелигиозного населения [6]. Аналогичный результат получается и для других стартовых значений переходных вероятностей, в частности, если 17 - процент «воцерковленных», т. е. *активных верующих*, и 22 - процент атеистов, т. е. *активных неверующих* (по данным опроса Аналитического центра Юрия Левады - Росбалт, 08/09/2011), то в перспективе - вектор стационарного распределения на полюсах степеней религиозности: «воцерковленных» - 19%; атеистов – 19%.

## Расчёт тестового примера

В основе расчёта положена упрощённая модель российского социума занятого (работающего и рабочего) населения в конкретном регионе, описанная в Таблице 4, но с учётом в нём долей женщин – 0,61 и мужчин.

При использовании массовых цепочек в Табл.4 (БТБТБ, БТБСТСБ, ...) при вычислении вероятностей согласно формулам (13), (14) обнаруживается, что итоговые величины много меньше 1, т. е. чем сложнее цепочки, тем меньше их вклад в итоговых выражениях для рассчитываемых оценок.

Аналогично в качестве тестовых были получены оценки усреднённой степени религиозности для Приволжского и Северо-Западного округов РФ. Показатель  $P_{ver} = 0.402$  характеризует асимптотику доли воцерковленного населения при неизменной структуре социума. Показатель  $P_{unver} = 0.598$  характеризует асимптотику суммарной доли неверующих и колеблющихся в вере. К примеру – в такой области, как Нижегородская, на 1995-96 гг доля населения, вовлечённого в религиозную жизнь, т. е. имеющего дома иконы и т. п. культовые предметы, составило 39.2% [20], что близко к полученной оценке.

## Выводы и заключение

1. В работе поставлена, теоретически обоснована, математически формализована и решена задача расчёта регулярных социальных прогнозов на скрытых топологических однородных марковских цепях.

2. Эффективность работы представленного метода на примере расчёта прогноза степени религиозности российского социума подтверждена в аналогичном прогнозе Европейской комиссии социологов.

3. Точками роста развиваемого метода являются 1) его обобщение с матрицы перехода «2 x 2» на матрицы перехода «n x n», где  $n = 3 - 6$ .

4. Обобщение метода для однородной матрицы вероятностей перехода на неоднородную матрицу.

## Библиографический список

1. Алексеев В.М., Якобсон М.В. Символическая динамика и гиперболические динамические системы – Добавление в книге Р.Боуэна «Методы символической динамики», серия «Математика» - Новое в зарубежной науке 13, Мир, М., 1979.
2. Вовк Е. Гендерная асимметрия и женские роли в современной России / Е. Вовк // Социальная реальность. - 2006. - N 3. - С. 61-73.

3. Кублицкая Е.А. Особенности религиозности в современной России // Социс. 2009, №4 . - С. 96-108.
4. Луман Н. Теория общества // Теория общества (фундаментальные проблемы) / Под ред. А.Ф.Филиппова.- М.: «КАНОН-пресс-Ц», «Кучково поле», 1999. С.196-232.
5. Николенко С.И. Скрытые марковские модели, Академический университет, С.-Пб.,2011 - URL: [logic.pdmi.ras.ru/~sergey/teaching/mlaptu11/06-hmm.pdf](http://logic.pdmi.ras.ru/~sergey/teaching/mlaptu11/06-hmm.pdf).
6. Кофанова Е. и М.Мчедлова М. Религиозность россиян и европейцев // Перспективы, 2010-05-20.htm.
7. Цеханская К.В. Россия: тенденции религиозности в XX веке (некоторые статистические данные) // Исторический вестник, 2000, № 5.
8. Шведовский В.А. О гипотезе ограничения рождаемости в основных этнических группах иммигрантов Москвы, Научно-практический журнал «Человеческий капитал», № 9 (21), - М.: Союз, 2010, С. 24-29.
9. Шведовский В.А., Шведовский О.В. Тополого-сетевая модель жизненного пространства личности как обобщение гипотезы К.Левина // Математическое моделирование социальных процессов, МГУ Социологический ф-т, вып.11., М.,УНИВЕРСИТЕТ – Книжный дом, 2010.
10. Шведовский В.А. Информационный критерий в обобщении уравнения Н.Д.Кондратьева для современной социально-экономической ситуации в России // Математическое моделирование социальных процессов, МГУ Социологический ф-т, вып.9., М.,УНИВЕРСИТЕТ – Книжный дом, 2007. С. 73.
11. Шведовский В.А. К построению модели комплекса факторов воспитания личности в трудовом коллективе // Роль трудовых коллективов в формировании активной жизненной позиции, Институт социологических исследований АН СССР, - М.: 1979, С.52-55.
12. Шведовский В.А. Проблемы московской самобытности // Социология, Журнал Российской социологической ассоциации № 3. М.: 2011, С. 57-75.
13. Широкалова Г.С. Сравнительные характеристики верующих и неверующих нижегородцев // Социол. исслед. 2001.№ 7. С.80-88.