Коротаев А. В., Малков А. С., Халтурина Д. А. 2007. *Законы истории: Математическое моделирование развития Мир-Системы. Демография, экономика, культура*. М.: КомКнига/URSS. С. 14–22.

Часть I

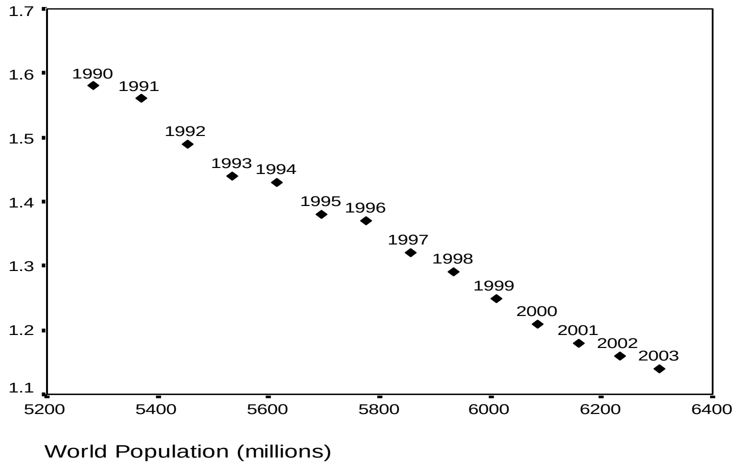
Компактные макромодели эволюции мир-системы

Глава 1

Демографическая динамика мира после 1989 г.: некоторые наблюдения

Собственно говоря, в 1990–2003 гг. мы имеем дело с исключительно сильной отрицательной корреляцией между численностью населения мира и относительными темпами его роста (см. Диаграмму 1.1):

Диаграмма 1.1. Соотношение между численностью и годовыми темпами роста населения мира, 1990–2003 гг.



Корреляционный и регрессионный анализ рассматриваемых рядов данных дает следующие результаты (см. Табл. 1.1a и 1.1b):

Таблица 1.1. Корреляция между численностью и годовыми темпами роста населения мира, 1990–2003 гг.

1.1a. Корреляционный анализ:

r	–0,996
α	0,000 000 000 000 04

ПОЯСНЕНИЯ К ТАБЛИЦЕ 1.1a: Мы отдаем себе отчет в том, что многим читателям цифры, приводимые в этой таблице, могут ничего не говорить. Однако это не так уж страшно. Рискнем утверждать, что прикладная математическая статистика "для пользователя" не так уж сложна. Для того чтобы эта книга могла быть по-настоящему полезна, в том числе и читателям, не имеющим математического образования, сделаем необходимые пояснения.

Приводимые в данной таблице числа характеризуют корреляцию между рассматриваемыми величинами. Корреляция (зависимость) между двумя переменными обычно характеризуется двумя показателями. Первый из них дает нам представление о силе связи между признаками. Чаще всего (в зависимости от типа данных) используются коэффициент корреляции Пирсона, обозначаемый обычно строчной латинской буквой r, *и коэффициент ранговой корреляции Спирмана, обозначаемый греческой буквой* ρ *(в англоязычной литературе часто используется название этой буквы в латинской графике –* Rho *или* Spearman's Rho*). Такие коэффициенты принимают значения от* – 1,0 *до* + 1,0*. Значение* + 1,0 *означает полную ("функциональную") положительную связь между признаками.*

Если между признаками существует причинно-следственная связь, это будет говорить нам о том, что увеличение значения величины x *приводит к однозначно определенному увеличению значения величины* y*. Значение* – 1,0 *означает полную ("функциональную") отрицательную связь между признаками. Если между признаками существует причинно-следственная связь, это будет свидетельствовать о том, что увеличение значения величины* x *приводит к однозначно определенному УМЕНЬШЕНИЮ значения величины* y*. Как можно видеть, коэффициент корреляции в нашем случае имеет отрицательное значение, т.е. корреляция у нас как раз отрицательная (то есть увеличение значения одной величины у нас сопровождается уменьшением значения другой). При положительной корреляции рост значения одной величины будет сопровождаться и ростом значения другой величины.*

Для того чтобы понять "рациональный смысл" коэффициента корреляции рекомендуется возвести его в квадрат. Полученное число легче все-

го интерпретировать, если между анализируемыми показателями существует причинно-следственная связь (что наблюдается, конечно же, далеко не всегда). В этом случае, например, $r^2 = 0{,}5^2 = 0{,}25$ *будет говорить о том, что показатель* x *детерминирует вариацию показателя* y *на 25%.*

В нашем случае нет оснований говорить о причинно-следственной зависимости между признаками. В подобных случаях, коэффициент корреляции более правильно интерпретировать как количественный показатель того, насколько достоверную информацию о значении показателя y *мы будем иметь, зная значение показателя* x*. Соответственно, если* r *[а значит и* r^2*] равно* 0*, это будет говорить о том, что знание значения показателя x не дает нам никакой информации ["предикции", от англ.* predict – предсказывать*] о значении показателя* y*. А если* r *[а значит и* r^2*] равно* 1*, то зная значение показателя* x*, мы будем абсолютно достоверно знать и значение показателя* y*. Величину* x *в таких случаях будет правильно обозначать не как фактор изменения величины* y*, а как ее "предиктор".*

Обычно в математической статистике корреляция считается сильной, если она характеризуется коэффициентом со значением более 0,7*, средней – при коэффициенте со значением между* 0,5 *и* 0,7 *и слабой, если значение коэффициента меньше* 0,5.

Рассматриваемая нами корреляция, однако, охарактеризована выше еще одной величиной (α = 0,000 000 000 000 01*). Это показатель статистической значимости корреляции. В англоязычной научной литературе для его обозначения чаще используется строчная латинская буква* p *(по первой букве слова* probability*, "вероятность"). Каков смысл этой величины? Какой смысл имеет, скажем, утверждение, что статистическая значимость некой корреляции равна* 0,01 *(или, что эта корреляция значима на уровне* 0,01*)? Это значит, что вероятность того, что подобная корреляция могла появиться в результате случайности, при отсутствии реальной закономерной связи между признаками равна* 0,01*, т.е. имеется лишь один шанс из ста, что наблюдаемая корреляция является результатом случайности. Понятно, что вероятность эта довольно низка, так что обычно в таком случае гипотеза о наличии связи между признаками будет считаться нашедшей подтверждение.*

Исторически сложилось, что в качестве порогового уровня статистической значимости принимается 0,05 *(~* 5% *~* 1 *шанс из двадцати). Таким образом, если мы получили показатель значимости менее* 0,05*, то соответствующая гипотеза считается успешно прошедшей статистическую проверку, если же этот показатель более* 0,05*, то соответствующая гипотеза считается неподтвержденной. Подчеркнем, что никакого рационального основания эта конвенция не имеет. Речь идет именно об исторически сложившейся в академическом сообществе научной практике.*

Применяемый в настоящее время способ оценки статистической значимости корреляций не является единственно возможным и создает заметные трудности для восприятия у людей, начинающих осваивать прикладную матстатистику. Действительно, с трудом воспринимается то обстоятельство, что чем МЕНЬШЕ значение α*, тем ВЫШЕ статистическая значимость связи; что* $\alpha = 0{,}000001$ *является индикатором высочайшей статистической значимости связи, в то время как* $\alpha = 0{,}8$ *наоборот говорит о крайне низкой статистической значимости (собственно говоря, о том, что корреляция здесь не является статистически значимой). Однако ничего уже здесь не поделаешь. И с этой академической конвенцией нам придется считаться.*

Необходимо подчеркнуть, что связь между силой корреляции и статистической значимостью корреляции довольно сложная. Речь идет о достаточно самостоятельных величинах. Корреляция может быть сильной, и вместе с тем иметь крайне низкую статистическую значимость. И наоборот, она может быть крайне слабой и иметь вместе с тем высочайшую статистическую значимость.

В случае с Табл. 1.1а мы имеем дело с корреляцией высочайшего уровня статистической значимости ($\alpha = 0{,}000\,000\,000\,000\,04$*). Т.е. имеется лишь четыре шанса из СТА ТРИЛЛИОНОВ, что наблюдаемая корреляция является результатом случайности, а закономерная связь между двумя рассматриваемыми переменными отсутствует. А значит, можно совершенно уверенно говорить о существовании закономерной связи между двумя данными признаками. Отметим, что обычно при значении показателя статистически значимости ниже* $0{,}0001$ *(а иногда даже* $0{,}001$*) точное число не указывается, т.е. нередко ограничиваются указаниями типа* $\alpha < 0{,}001$ *или* $\alpha < 0{,}0001$*, так как считается, что в таких случаях речь идет о заведомо статистически достоверной связи и бóльшая точность здесь уже не нужна.*

Наконец, поясним, что корреляция между значениями, предсказанными моделью, и актуально наблюдаемыми данными, обычно измеряется при помощи коэффициента корреляции R*, который принимает значения от* 0 *(полное несоответствие) до* 1 *(полное соответствие), и который еще неоднократно встретится нам на страницах этой книги.*

1.1b. Регрессионный анализ

Модель	Нестандиртизированный коэффициент B	Ст. ошибка	Стандиртизированный коэффициент β	t	Статистическая значимость (α)
(Константа)	3,90	0,064		61,29	0,0000000000000003
Население мира (в миллиардах)	–0,44	0,011	–0,996	–40,3	0,00000000000004
Зависимая переменная: **Относительная годовая скорость роста населения мира** (%)					

ПРИМЕЧАНИЕ: R = 0,996, R^2 = 0,993

ПОЯСНЕНИЯ К ТАБЛИЦЕ 1.1b: При внимательном изучении Табл. 2b в ней нетрудно заметить два числа, которые нам уже попадались в Табл. 2a. Действительно, значение стандартизированного β-коэффициента здесь совпадает со значением коэффициента корреляции Пирсона в Табл. 2a; полностью совпадают для обоих коэффициентов и показатели статистической значимости. Таким образом, регрессионный анализ позволяет нам установить все основные показатели корреляции между рассматриваемыми переменными.

Однако регрессионный анализ дает нам и другую важную информацию. Строго говоря, в таблице приводятся данные линейного регрессионного анализа, который наряду с прочим используется для проверки гипотезы о наличии между соответствующими переменными простой линейной зависимости, имеющей вид Y = a + bX. *Однако линейный регрессионный анализ позволяет не только установить сам факт наличия между признаками прямолинейной зависимости, но и выяснить основные характеристики этой зависимости. В качестве этих характеристик выступают константа* a *и коэффициент* b.

Первое число в строке "Константа" и дает нам значение константы a (3,9*). В качестве независимой переменной (т.е. переменной* X*) в нашем регрессионном анализе выступает численность населения мира. Соответственно, первое число в строке "Население мира (в миллиардах)" и будет представлять значение коэффициента* b (–0,44*). То, что этот коэффициент имеет данную величину, означает, что на рассматриваемом нами временном отрезке увеличение населения мира на миллиард человек сопровождалось уменьшением относительных темпов роста населения мира на* 0,44%.

Таким образом, мы получаем уравнение связи между численностью населения мира и относительными темпами его роста для периода 1990–2003 гг. В качестве "зависимой переменной" (Y) *у нас выступает относительная годовая скорость населения мира (в %%), в качестве "незави-*

*симой" (*X*) – численность населения мира в миллиардах. Обозначим эти величины соответственно как* V *и* N. *Теперь возьмем базовую формулу линейной зависимости* Y = a + bX*, подставим туда* V *(относительные темпы роста населения мира в %%) вместо* Y*, и вместо* X *поставим* N *(т.е. численность населения мира в миллиардах). Мы уже установили, что константа* a *в этом уравнении будет равна* 3,9 *(%), а коэффициент* b *будет иметь значение* –0,44*. В итоге мы получаем уравнение*

$$V = 3{,}9 - 0{,}44N.$$

Отметим, что данное уравнение уже представляет собой математическую модель мировой демографической макродинамики, при помощи которой мы даже уже можем прогнозировать рост населения мира в будущем. Для этого надо подставить в эту формулу численность населения мира на этот год, и мы сразу получим прогноз, насколько население мира вырастет в следующем году. Таким образом, мы предположительно узнаем, и какой будет через год численность населения мира. Подставив в формулу полученное значение численности населения мира в следующем году, мы сможем предположительно узнать, насколько оно вырастет через два года, и т.д. Таким образом, мы сможем сделать и прогноз того, каким население мира будет через 50, или, скажем, через 100 лет. Другой вопрос, конечно, насколько этот прогноз будет точным.

И еще одно замечание. Как мы видим, в рассмотренном нами случае корреляционный анализ не дает нам никакой полезной информации, которую мы бы не получили в процессе регрессионного анализа, а регрессионный анализ дает нам наряду с информацией, получаемой в процессе простого корреляционного анализа, еще и важную дополнительную информацию. Поэтому ниже в случаях, аналогичных рассмотренному выше, мы будем приводить данные лишь регрессионного анализа. Собственно говоря, данные простого корреляционного анализа были выше приведены нами лишь для того, что объяснить читателям, не имеющим математического образования, что такое корреляция, какие характеристики она имеет, и при помощи каких математических символов эти характеристики обозначаются, так как соответствующие понятия и символы будут неоднократно встречаться читателям на страницах этой книги.

Проведенный анализ заставляет предполагать, что 99,3% всей мировой макродемографической вариации в 1990–2003 гг. описывается при помощи следующего предельно простого уравнения:

$$V = 3{,}9 - 0{,}44N\,, \tag{1.1}$$

где N – население мира (в миллиардах чел.), а V – относительная годовая скорость роста населения мира[1] (в %%).

[1] Обычно эта величина обозначается как r (от английского *rate*), однако выше этот символ уже был использован для обозначения коэффициента корреляции Пирсона, поэтому во из-

Естественно, это позволяет оценивать будущую динамику народонаселения мира (при условия сохранения наблюдаемого в последнее время паттерна соотношений между N и V) при помощи следующего разностного уравнения:

$$N_{i+1} = N_i\left(1 + \frac{3{,}9 - 0{,}44N_i}{100}\right), \qquad (1.2)$$

где N_i – население мира (в миллиардах чел.) в году i, а N_{i+1} – численность, которой население мира достигнет через год.

ПОЯСНЕНИЯ К МОДЕЛИ (1.2)*: Поясним, как из уравнения* (1.1)*, т.е.*

$$V = 3{,}9 - 0{,}44N,$$

было получено уравнение (1.2)*, т.е.*

$$N_{i+1} = N_i\left(1 + \frac{3{,}9 - 0{,}44N_i}{100}\right).$$

Допустим, нам известно, что численность населения мира в этом году равняется N_i*, а относительная скорость роста населения мира в следующем году составит* V%. *Как на основании этих данных вычислить, какой численности население мира достигнет в следующем году?*

Ясно, что надо взять численность населения мира в этом году, т.е. N_i*, умножить на тот процент, на который оно должно вырасти в следующем году, т.е. на* V*, и поделить полученное число на* 100. *Т.е. ожидаемый прирост населения в следующем году составит* $N_iV/100$. *Теперь ожидаемый прирост населения за следующий год сложим с численностью населения мира в этом году. Таким образом, численность населения мира* (N_{i+1}) *в следующем году окажется равной* $N_i + N_iV/100$. *В результате мы получаем следующее уравнение:* $N_{i+1} = N_i + N_iV/100$.

Теперь вынесем N_i *за скобки и получим:* $N_{i+1} = N_i(1 + V/100)$. *Для того чтобы, зная численность населения мира в этом году, мы могли бы предположительно подсчитать, каким население мира станет в следующем году, нам осталось установить, какой будет относительная годовая скорость роста населения мира в следующем году. Мы можем сделать это, используя уравнение* (1.1)*, т.е.* $V = 3{,}9 - 0{,}44N$.

Как мы помним, при помощи этого уравнения, зная, какова численность населения мира в этом году, мы как раз можем вычислить, на сколько процентов она вырастет в следующем году. Допустим, в этом

бежание возможной путаницы мы решили использовать здесь хорошо известный всем читателям символ, широко применяемый для обозначения скорости, *V*.

году численность населения мира равна N_i. *Соответственно, в следующем году она вырастет на* ($3{,}9 - 0{,}44N_i$)%. *Теперь осталось подставить это выражение на место* V *в уравнение* $N_{i+1} = N_i(1 + V/100)$ *и мы получим разностное уравнение* (1.2)*:*

$$N_{i+1} = N_i(1 + \frac{3{,}9 - 0{,}44N_i}{100}) .$$

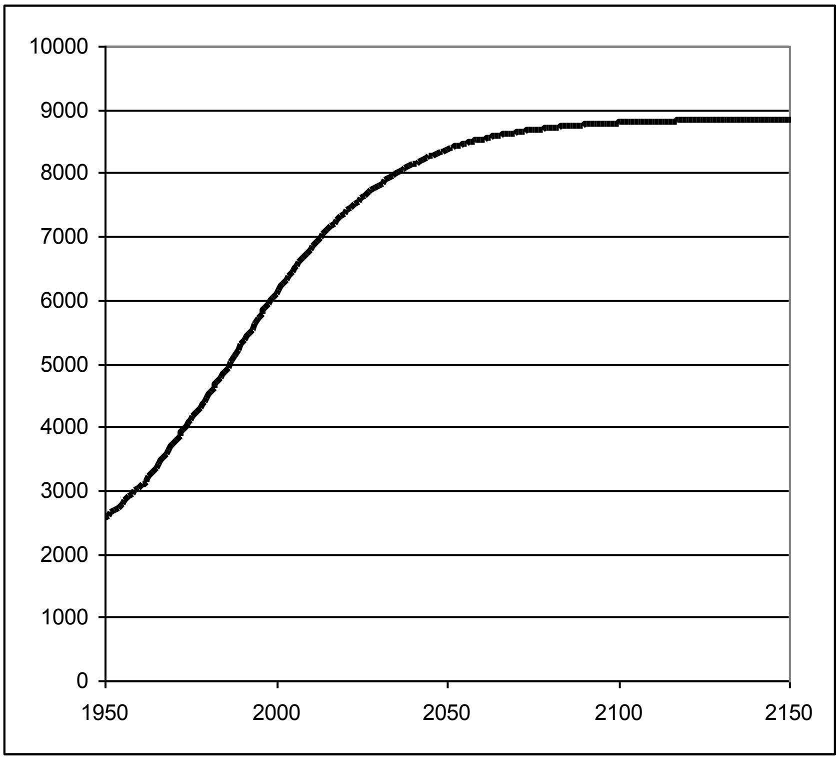
Конечно же, это уравнение позволяет, зная значение N_i , *установить (с достаточно небольшой погрешностью) значение* N_{i+1} *только для 1990–2003 гг. Неизбежно возникает вопрос, насколько оправданно использование этого уравнения для прогнозирования численности населения мира в последующие годы. Ответу на данный вопрос и посвящена заметная часть этой книги.*

Результаты соответствующей компьютерной симуляции с началом в 2003 г. и стартовым значением N = 6305,15 миллионов чел. выглядят следующим образом (см. Табл. 1.2 и Диаграмму 1.2):

Таблица 1.2. Прогноз численности населения мира (в миллионах чел.), оценки, сгенерированные компьютерной симуляцией с использованием модели (1.1)

Год	*2010*	*2020*	*2030*	*2040*	*2050*	*2060*	*2070*
Население	6785,6	7360,3	7801,6	8126,0	8356,8	8517,2	8626,8
Год	*2080*	*2090*	*2100*	*2110*	*2120*	*2130*	*2150*
Население	8700,9	8750,6	8783,8	8805,8	8820,5	8830,2	8840,8

Диаграмма 1.2. Население мира (в миллионах) в 1950–2003 гг., с экстраполяцией динамического тренда 1990–2003 гг. до 2150 г.



Насколько вероятно, что реальный рост населения мира будет идти по данному паттерну? Как мы увидим ниже, имеются вполне серьезные теоретические и эмпирические основания утверждать, что подобное развитие не может рассматриваться как совершенно невероятное.