

О ПРИМЕНЕНИИ МАТЕМАТИЧЕСКИХ МЕТОДОВ ПРИ АНАЛИЗЕ ИСТОРИКО-СТАТИСТИЧЕСКИХ ДАННЫХ

И. Д. КОВАЛЬЧЕНКО

Отличительная черта развития науки в последнее время — радикальное совершенствование приемов накопления, обработки и анализа фактического материала на основе широкого применения электронной техники и математических методов.

Раньше или позже этот качественный сдвиг неизбежно охватит все области науки. В настоящее время назрела необходимость применения новых методов и в исторической науке. Это прежде всего относится к такому ее ведущему разделу, каким является социально-экономическая история. То обстоятельство, что здесь, как правило, исследование основывается на обработке и анализе историко-статистических материалов, уже сейчас делает возможным широкое применение электронных машин (вычислительных и логически-информационных) и математических методов.

Первые опыты подобного рода, предпринятые в последнее время, дали весьма обнадеживающие результаты. Они показали, что широкое применение новых методов, во-первых, освободит исследователей от чрезвычайно трудоемкой, требующей огромного времени чисто технической работы. Это позволит уделить больше внимания творческому

осмысливанию конкретного материала и сократит сроки выполнения исследований. Во-вторых, появится возможность значительно расширить круг фактического материала, привлекаемого исследователями, и осуществить такую его обработку и анализ, которые невозможны при использовании традиционных приемов и методов. Это существенно повысит глубину исследований, поднимет их на качественно новую ступень. В-третьих, применение новых методов позволит исследовать на основе широкого круга источников ряд таких проблем, которые сейчас даже не ставятся. Это будет содействовать актуализации исторической проблематики и повысит значение исторических трудов.

В настоящее время, когда работа в новом направлении лишь начинается, важно, чтобы историки убедились в эффективности новых методов, осознали необходимость их внедрения, что позволит перейти к широкому и планомерному их применению.

* * *

Цель настоящего сообщения — на конкретном примере показать целесообразность и эффективность применения методов математической статистики для анализа историко-статистических материалов.

В настоящее время историки при анализе статистических данных, как правило, опираются на простейшие положения общей статистики и элементарной математики. При этом они имеют возможность выражать в количественных показателях взаимосвязь различных явлений и факторов, определять их роль в процессе, устанавливать степень сходства и различия между ними и т. п. лишь в тех случаях, когда между явлениями и процессами существует функциональная зависимость. Те или иные явления и процессы относятся к функционально зависимым в том случае, если какой-либо количественной характеристике одного явления соответствует строго определенный количественный показатель в другом явлении, зависимом от первого. Но таких взаимосвязей, когда определенное количественное изменение одного фактора ведет к строго определенному количественному изменению другого, в общественном развитии очень немного. Чаще всего изменения в тех или иных сторонах общественной жизни зависят от совокупного и одновременного воздействия ряда причин. При этом каждая из них, в зависимости от силы воздействия других факторов, влияет на зависимое от нее явление в каждом конкретном случае с различной интенсивностью и может приводить к различным количественным изменениям. Подобная зависимость между явлениями и процессами отлична от функциональной и получила название случайной, корреляционной или статистической зависимости. Понятие «случайная» в данном случае означает наличие такой сложной множественной зависимости, когда, в силу невозможности учета всех факторов, влияющих на процесс, нельзя сказать, приведут ли изменения в одном факторе к изменениям в другом, либо таких изменений не последует, во-первых, и когда нельзя точно указать, каков будет масштаб изменений, если они произойдут, во-вторых. Здесь возможны лишь приближенные исчисления, основанные на том или ином числе наблюдений (отсюда и определение подобной взаимосвязи как «статистической»).

Анализ случайных или корреляционных взаимосвязей является предметом теории вероятностей и математической статистики, важнейших разделов высшей математики. В них хорошо разработаны мощные приемы и способы анализа случайных явлений и процессов, позволяющие в точных количественных показателях раскрывать закономерности и связи, присущие этим явлениям. До сих пор историки, имея дело со статистическими данными, как правило, не прибегали при

ТАБЛИЦА I

Зависимость размеров крестьянского хозяйства от обеспеченности работниками, эксплуатации и развития промыслов

Годы	На двор			
	Скота в переводе на крупный (голов)	Работников-мужчин (чел.)	На работника барской запашки (дес.), оброка (руб. сер.), или тягол	Занятых промыслами мужчин (чел.)
Покровское имение Сапожковского уезда				
1813	6,6	1,9	3,8 дес.	
1817	7,1	1,9	3,7	
1823	7,4	2,0	3,7	
1826	6,0	2,0	3,3	
1831	5,6	2,0	2,9	
1835	5,4	2,0	2,9	
1856	5,2	2,0	3,3	
1856 г. в % к 1813 г.	78,2	105,3	86,8	
Коэффициент корреляции		-0,48	0,75	
Коэффициент регрессии		-0,85	1,44	
Киясовское имение Серпуховского уезда				
1817	3,5	2,0	6,6 руб.	
1822	3,8	2,3	8,2	0,4
1834	2,8	2,4	10,1	1,4
1840	2,6	2,3	10,1	1,4
1851	2,8	2,3	11,5	
1858	2,1	1,9	13,0	1,4
1858 г. в % к 1817 г.	60,0	95,0	197,0	350,0
Коэффициент корреляции		0,26	-0,12	-0,78
Коэффициент регрессии		0,80	-0,03	-1,87
Петровское имение Звенигородского уезда				
1824	3,0	1,7	1,0 тягол	1,0
1828	2,4	1,8	0,9	0,9
1834	2,6	2,0	0,9	1,2
1854	2,5	1,9	0,8	1,5
1857	2,1	1,8	0,7	1,5
1860	1,8	2,0	0,7	1,6
1860 г. в % к 1824 г.	60,0	117,6	70,0	160,0
Коэффициент корреляции		-0,52	-0,35	-0,69
Коэффициент регрессии		-0,15	-0,70	-0,48

их анализе к помощи теории вероятностей и математической статистики, так как это требует знания высшей математики. Интересы развития исторической науки требуют преодоления этого барьера. Решение этой задачи облегчается успехами в развитии вычислительной техники. Теперь историк, будучи знаком с основными принципами теории вероятностей и математической статистики, может подготовить данные и поставить соответствующую задачу, математический аппарат которой будет оформлен математиками, а необходимые подсчеты, на выполнение которых уходит основное время, произведены на вычислительных машинах.

Применение этих методов позволит более глубоко раскрыть взаимосвязи различных факторов, их роль в тех или иных процессах и степень воздействия на те или иные стороны общественного развития. Нет нужды говорить о том, насколько важно знание этих взаимо-

связей, однако именно здесь исследователи в наибольшей степени испытывают ограниченность традиционных приемов анализа и поэтому очень часто далеко не в полной мере используют имеющиеся в их распоряжении конкретные данные. Подобные затруднения испытывал и автор настоящей статьи, занимаясь изучением состояния хозяйства и положения крестьян в первой половине XIX в. Поиски наиболее эффективных методов обработки и анализа имеющихся данных и привели его в Лабораторию применения электронных вычислительных машин в гуманитарных исследованиях при Институте математики Сибирского отделения АН. Здесь, наряду с опытом обработки на электронной вычислительной машине данных подворных описаний крестьянских хозяйств, был проведен анализ ряда показателей о состоянии хозяйства и расслоении крестьян методами математической статистики. Конкретно задача сводилась к выявлению зависимости размеров и уровня крестьянского сельскохозяйственного производства от обеспеченности крестьян рабочей силой, интенсивности их эксплуатации и степени развития промыслов. Кроме того, выявлялась зависимость численности различных прослоек в среде крестьянства (т. е. их доли в общем числе дворов) от интенсивности эксплуатации крестьян и степени развития в их среде промыслов.

Для определения указанных зависимостей вычислялись коэффициенты линейной корреляции и регрессии. Коэффициент корреляции показывает степень взаимосвязи между различными факторами. Его значения могут колебаться в пределах от -1 до $+1$. Чем ближе значение коэффициента к своему предельному значению, т. е. к 1 , тем выше степень связи между явлениями. В тех случаях, когда коэффициент равен 1 (плюс или минус, все равно), мы имеем дело с функциональной зависимостью. Когда же коэффициент равен 0 , между рассматриваемыми факторами отсутствует всякая взаимосвязь.

Установление на основе коэффициента корреляции факта наличия или отсутствия зависимости между различными явлениями, в тех случаях, когда эта зависимость не очевидна, само по себе имеет большое значение при анализе исторических процессов. При наличии существенной зависимости между одними факторами и отсутствии ее между другими, коэффициент корреляции позволяет решать вопрос о роли этих факторов в изучаемом процессе. Однако, в случае существенного влияния на ход процесса двух и более факторов, различия в коэффициентах корреляции не отражают степени воздействия каждого из факторов на ход этого процесса. Ответ на этот вопрос дает коэффициент регрессии. Он показывает интенсивность или меру изменения одного фактора, в зависимости от изменения другого при прочих равных условиях.

Коэффициенты корреляции, определение которых требует значительных подсчетов, вычислялись на электронной машине. Вычисление коэффициентов линейной регрессии несложно, однако оно требует предварительного выведения коэффициентов корреляции.

Что дает исследователю применение указанных приемов анализа, хорошо видно на следующих примерах. В таблице 1 приведены данные, характеризующие зависимость размеров крестьянского сельскохозяйственного производства (за показатель их принято число всего скота на двор в переводе на крупный) от обеспеченности крестьянских хозяйств собственной рабочей силой, интенсивности их эксплуатации и степени развития промысловой деятельности. Данные взяты по Покровскому (Сапожковский уезд Рязанской губ.) и Киясовскому (Серпуховской уезд Московской губ.) имениям Гагариных и Петровскому имению Голицыных (Звенигородский уезд). Покровское имение было барщинным, и в нем за показатель интенсивности эксплуатации приняты размеры помещицкой запашки на работника (в дес.), Киясов-

ское — оброчным, и интенсивность эксплуатации отражена в размерах оброка (в руб. сер.), Петровское — смешанным, и за показатель эксплуатации здесь пришлось взять число тяглов на работника. Все исходные данные извлечены из подворных описей и других материалов, содержащихся в фондах Гагариных и Голицыных (ЦГАДА).

В Покровском имении с 1813 по 1856 г. количество скота на двор сократилось на 22%, т. е. размеры крестьянского хозяйства существенно уменьшились. Ведя анализ обычным способом, мы видим, что за это же время число работников на двор несколько возросло (на 5,3%). Это значит, что сокращение размеров хозяйства здесь не зависело от изменения обеспеченности дворов работниками. Размеры барской запашки на работника сократились (на 13%), т. е. интенсивность эксплуатации крестьян, если судить о ней по этому показателю, не возросла. В итоге имеющиеся данные при анализе их обычным способом не дают прямого ответа на вопрос о причинах сокращения размеров крестьянского сельскохозяйственного производства и о той роли, которую играла при этом эксплуатация крестьян. Лишь данные о том, что размеры барской запашки на работника выросли здесь еще в предыдущий период, в конце XVIII — начале XIX в. в несколько раз и что в рассматриваемое время барщина отнимала у крестьян примерно 60% рабочего времени, позволяют главную причину сокращения размеров крестьянского хозяйства усматривать в высокой интенсивности эксплуатации крестьян. Но это доказательство косвенное, и при строгом подходе к вопросу против него можно найти возражения.

Между тем определение тесноты связи между размерами хозяйства и обеспеченностью крестьян работниками и интенсивностью их эксплуатации и определение темпов изменения размеров хозяйства в зависимости от изменения указанных факторов, основанное на приемах математической статистики, подтверждает этот вывод. Для рассматриваемых явлений взаимосвязь факторов может считаться существенной при коэффициенте корреляции, равном не менее 0,70. Отсюда ясно, что между размерами крестьянского хозяйства и обеспеченностью его рабочей силой в данном случае существенной зависимости не было (коэффициент корреляции равен $-0,48$). С другой стороны, между размерами хозяйства и интенсивностью эксплуатации в Покровском имении имела место существенная зависимость (коэффициент корреляции 0,75). Это значит, что при прочих равных условиях интенсивность эксплуатации была здесь основным фактором, влиявшим на размеры хозяйства крестьян. Так как зависимость между размерами хозяйства и степенью эксплуатации по своему характеру является обратной, то корреляция должна была быть отрицательной, но в данном случае она положительная потому, что в Покровском имении эта зависимость была взаимной, т. е. не только размеры хозяйства зависели от эксплуатации, но и эксплуатация — от размеров хозяйства. Как показывают коэффициенты регрессии, изменения эксплуатации (коэффициент регрессии 1,44) воздействовали на изменения размеров хозяйства почти вдвое интенсивнее, чем изменения в обеспеченности работниками (коэффициент регрессии $-0,85$)¹.

Таким образом, анализ данных методами математической статистики позволяет получить прямой и точный ответ на вопрос, который при обычном анализе может быть решен лишь предположительно.

В Киясовском и Петровском имениях за рассматриваемый период размеры крестьянского сельскохозяйственного производства резко сократились (количество скота на двор в обоих имениях уменьшилось на 40%). Кроме данных об обеспеченности крестьян работниками и раз-

¹ Коэффициент регрессии 1,44 означает, что при изменении интенсивности эксплуатации на 1,0 размеры хозяйства изменятся на 1,44.

меров помещичьих повинностей, здесь есть еще сведения о втянутости крестьян в промысловые занятия. Однако, анализируя все имеющиеся показатели обычными методами, исследователь не сможет правильно решить вопрос о том, какой из трех указанных факторов играл решающую роль в сокращении размеров крестьянского хозяйства. Бесспорным будет лишь тот вывод, что изменения в обеспеченности дворов работниками, в силу их незначительности, не имели существенного значения.

Далее, анализируя данные по Киясовскому имению, любой исследователь наверное пришел бы к выводу, что значительное усиление эксплуатации крестьян (рост размеров оброка на 97%) было основной причиной сокращения размеров крестьянского хозяйства. Единственно, что здесь могло натолкнуть на сомнения,—это исправная уплата крестьянами оброка в 40—50-х годах и даже погашение значительной части недоимочной суммы. Но это сомнение может быть снято фактом роста промысловых доходов крестьян. Между тем вывод о том, что основной причиной сокращения размеров крестьянского хозяйства в Киясовском имении являлось усиление эксплуатации крестьян, как показывают коэффициенты корреляции и регрессии, ошибочен. В Киясовском имении отсутствовала существенная взаимосвязь не только между размерами хозяйства и обеспеченностью работниками (коэффициент корреляции 0,26), но и между этими размерами и эксплуатацией (коэффициент корреляции всего — 0,12). Но зато имелась существенная зависимость между размерами хозяйства и степенью распространенности промысловых занятий (коэффициент корреляции — 0,78). Положение было таково, что именно этот фактор оказывал наибольшее воздействие на изменение размеров хозяйства (коэффициент регрессии — 1,87, против 0,80 по обеспеченности работниками и против — 0,03 по эксплуатации). Вполне естественно, что в условиях все большего и большего втягивания крестьян в промысловые занятия с определенного момента началось свертывание их земледельческого производства (обратная зависимость в направлении динамики промысловых и земледельческих занятий отражена в том, что коэффициенты корреляции и регрессии отрицательны).

Еще более сложна картина по Петровскому имению. Здесь, ввиду того, что изменения в интенсивности эксплуатации и размахе промысловой деятельности шли в обратных направлениях, а степень их воздействия на размеры хозяйства неизвестна, обычный анализ данных вообще ничего, кроме произвольных предположений дать не может. Коэффициент же корреляции показывает, что в Петровском имении наиболее тесной была связь между размерами хозяйства и степенью распространенности промысловых занятий (—0,69). Менее значительной была связь между размерами хозяйства и обеспеченностью дворов работниками (—0,52) и совсем несущественной — между размерами хозяйства и эксплуатацией (—0,35). Однако в динамике по силе своего воздействия на размеры хозяйства эти факторы располагались иначе. Наибольшее влияние на изменение размеров хозяйства оказывала интенсивность эксплуатации (коэффициент регрессии — 0,70), затем степень развитости промыслов (—0,48) и, наконец, обеспеченность работниками (—0,15). Таким образом, налицо совершенно определенная картина.

Возьмем другой пример. В таблице 2 указана численность (по доле дворов) различных прослоек в среде крестьянства и ее динамика. Численность существенно отличных прослоек в среде крестьянства, направление и темпы ее изменения всецело определялись соотношением факторов, углублявших и тормозивших процесс расслоения крестьян. Для правильного понимания хода и характера расслоения крестьян в тот или иной период очень важно знать, каково было соотно-

шение этих факторов, какие из них оказывали определяющее воздействие на ход процесса. Обычным путем получить ответ на этот вопрос, как правило, невозможно. Анализ же данных методами математической статистики позволяет исследователю подойти к решению

ТАБЛИЦА 2

Зависимость численности групп крестьян от эксплуатации и развития промыслов

Прослойки крестьян	Беднейшая			Средняя			Зажиточная		
	Дворов в % к общему числу	На работника повинностей	На двор занятых промыслами (мужч.)	Дворов в % к общему числу	На работника повинностей	На двор занятых промыслами (мужч.)	Дворов в % к общему числу	На работника повинностей	На двор занятых промыслами (мужч.)
Покровское имение Сапожковского уезда									
1813	13,3	1,7 дес.		62,8	3,8 дес.		23,9	4,3 дес.	
1817	11,7	3,3		71,9	3,7		16,4	3,7	
1823	12,6	2,8		55,5	3,7		31,9	3,7	
1826	14,4	2,6		69,9	3,3		15,7	3,7	
1831	15,2	2,5		68,9	2,9		15,9	3,2	
1835	14,7	2,2		70,9	2,9		14,4	3,2	
1856	22,3	2,5		65,5	3,3		12,2	3,3	
Коэффициент корреляции		0,18			-0,51			0,58	
Киясовское имение Серпуховского уезда									
1817	37,4	6,8 руб.		43,9	6,6 руб.		18,7	6,0 руб.	
1822	21,3	7,8	0,4	46,9	9,2	0,4	31,8	7,9	0,5
1834	52,6	9,8	1,2	31,4	9,8	1,5	16,0	10,8	1,9
1840	47,2	10,4	1,1	39,5	10,7	1,8	13,3	9,7	1,7
1851	51,2	12,4		38,8	11,3		10,0	12,5	
1858	53,8	12,2	1,1	39,0	12,7	1,8	7,2	13,4	1,8
Коэффициент корреляции		0,67	0,65		0,39	-0,80		0,90	-0,60
Коэффициент регрессии		3,59	8,45		1,00	-6,1		3,60	-6,18
Петровское имение Звенигородского уезда									
1824	36,3	0,9 тягол	0,5	42,2	1,0 тягол	1,1	21,5	1,1 тягол	1,8
1828	46,3	0,9	0,4	44,7	0,9	1,1	9,0	0,9	2,4
1834	41,7	1,0	0,6	51,2	0,9	1,4	7,1	1,0	3,4
1854	46,0	0,7	1,2	47,4	0,9	1,7	6,6	0,7	2,8
1857	49,7	0,7	1,2	48,3	0,8	1,7	2,0	0,8	3,0
1860	51,3	0,7	1,3	44,0	0,7	1,9	4,7	0,9	3,3
Коэффициент корреляции		-0,12	0,31		0,75	-0,84		0,21	-0,71
Коэффициент регрессии		-0,36	3,19		1,58	-7,64		0,46	-1,70

вопроса. Указанные в таблице 2 коэффициенты корреляции и регрессии отражают зависимость численности групп и ее динамики от интенсивности эксплуатации, т. е. фактора, тормозившего процесс расслоения, и степени развития промыслов, т. е. фактора, в определенной мере отражавшего глубину втянутости крестьянского хозяйства в товарно-денежные отношения.

По Покровскому имению, к сожалению, удалось определить лишь наличие зависимости между численностью групп и интенсивностью эксплуатации. Положение здесь было таково, что ни в одной из групп

существенной зависимости между этими факторами не было (наибольший коэффициент корреляции равен 0,58). Это значит, что изменение численности групп в имении определялось прежде всего втянутостью крестьянского хозяйства в товарно-денежные отношения. Несущественные изменения в соотношении групп в рассматриваемое время говорят о еще слабой связи крестьянского хозяйства с рынком.

В Киясовском имении взаимосвязь между численностью групп и эксплуатацией была существенной в группе зажиточного крестьянства, близкой к существенной — у беднейшего и несущественной — у среднего. Взаимосвязь же между численностью групп и степенью развития промыслов во всех группах была близкой к существенной или существенной.

Что касается соотношения этих факторов, то степень развития промыслов оказывала большее влияние на изменение численности групп, чем интенсивность эксплуатации (коэффициенты регрессии по промыслам значительно выше коэффициентов регрессии по эксплуатации). Лишь у зажиточного и в меньшей мере у беднейшего крестьянства степень воздействия этих факторов была более или менее близкой и влияние эксплуатации существенно сказывалось на изменении численности этих групп. Но в целом изменения в численности групп вызывались прежде всего степенью втянутости крестьян в товарно-денежные отношения. Более существенный характер изменений в численности групп в Киясовском имении сравнительно с Покровским свидетельствует о значительно более глубокой связи здесь крестьян с рынком. Это всецело подтверждается другими данными.

Не рассматривая подробно данных по Петровскому имению, отметим, что здесь во всех группах степень развития промыслов оказывала на изменение численности групп в несколько раз большее влияние, чем интенсивность эксплуатации (см. коэффициенты регрессии). Даже резкое сокращение численности зажиточной прослойки здесь почти не зависело от эксплуатации, а было обусловлено широким развитием промыслов в форме работы по найму (в этом особенность Петровского имения, а в несколько меньшей мере и Киясовского).

Таким образом более совершенные методы анализа позволяют находить решение даже такого вопроса, на который, казалось бы, вообще невозможно получить определенный ответ.

Думается, что рассмотренные примеры достаточно убедительно иллюстрируют преимущества анализа данных, основанного на достижениях высшей математики, сравнительно с приемами, базирующимися на элементарной математике. Заметим, что в приведенных примерах использован хотя и очень важный, но всего один прием из арсенала математической статистики. Имеется немало других методов, разработанных в теории вероятностей и математической статистике, которые могут быть с успехом применены при анализе историко-статистических данных. Поиски, освоение и внедрение новых методов обработки и анализа данных — вот та новая задача, которая возникла сейчас перед историками.