2. Основные понятия факторного анализа

Факторный анализ был развит психологами в течение первой половины XX в., главным образом в работах Спирмена, Терстона, Кетелла и Хотелинга.

Начиная с 50-х годов он начал широко применяться в социологии, социальной психологии и во многих областях социального исследования. В последние годы — и это весьма характерно — он привлек внимание крупных специалистов математической статистики— Кенделла, Бартлета, Лоули — и представляет в настоящее время весьма разработанную математическую теорию.

Как и в случае шкалирования, мы имеем N лиц и n вопросов— переменных, т. е. таблицу (матрицу), в которой по строке расположены ответы лица на вопросы, а по столбцу — ответы лиц на определенный вопрос. Существенно, что переменные количественные, т. е. ответ выражается числом, выражающим реальное количественное отношение — возраст, доход, разряд и т. п. Смысл факторного анализа заключается в том, что принято считать данные n переменных линейными функциями меньшего числа других переменных, называемых факторами. Факторы выступают как бы более фундаментальными переменными, характеризующими явление, и исходные переменные как бы объединяются в группы, каждая из которых представляет некий фактор. Задача анализа — найти эти факторы.

Поскольку фактор представляет собой объединение определенных переменных, постольку из этого следует, что эти переменные связаны друг с другом, обладают корреляцией, причем большей между собой, чем с другими переменными, входящими в другой фактор. Методы отыскания факторов основываются на использовании именно коэффициентов корреляции между переменными. Факторный анализ дает нетривиальное решение, т. е. это решение нельзя предвидеть и усмотреть, не применяя специальную технику извлечения факторов. Вместе с тем его решение имеет большое значение для характеристики социального явления, поскольку вначале оно характеризовалось n переменными, а в результате применения анализа оказалось, что оно характеризуется меньшим числом — q — других переменных-факторов.

Первоначально Спирмен выдвинул идею одного так называемого генерального, фактора g-фактор). Это означало, что вся деятельность индивида обусловлена влиянием одного генерального фактора. Смысл факторного анализа, по Спирмену, состоит в том, что «все стороны интеллектуальной активности вообще имеют одну фундаментальную функцию... в то время как оставшиеся или специфические элементы деятельности, по-видимому, в каждом случае отличны от нее во всем другом» [2]. Если мы задаем индивидам вопросы, то, по Спирмену, необходимо установить влияние, зависимость ответов от действия генерального фактора, т. е. требуется установление корреляции переменных с фактором: большая корреляция — большая связь данной переменной (действия) с фактором и т. д.

Позднее Терстон обобщил идею Спирмена и Хольцингера в своей модели многофакторного анализа, в которой существует конечное множество факторов, которые обуславливают значение данной системы эмпирических переменных. Терстон применил аппарат матричной алгебры и создал весьма разработанный формализм факторного анализа[3]. Терстон проделал также большую работу в области методологии факторного анализа. Он подразделил факторный анализ на факторный анализ описания и факторный анализ объяснения.

Факторный анализ определенной эмпирической системы данных будет только описательным анализом данной системы, находясь в зависимости от выбора переменных и популяции. Но если при изучении явления применять факторный анализ на разных системах переменных и на разных популяциях и если при этом будут получаться однотипные факторы, то эти факторы уже будут объясняющими. Описательный факторный анализ дает факторную картину единичного явления, объясняющий позволяет найти внутренние глубинные переменные, которые обусловливают эмпирическую картину. Терстон сформулировал ряд условий, которым должна отвечать факторная система: простота, инвариантность, единственность. Собственно, это те методологические требования, которые предъявляются любой научной теории.

Факторный анализ является разделом многомерной статистики, поскольку популяция индивидов исследуется по *n* переменным (измерениям), характеризующимся *n* эмпирическими распределениями. Можно установить зависимости между ними, вычисляя коэффициенты корреляции. Переменные и их распределения распадутся на группы по величине коэффициентов корреляции. Например, первая переменная тесно связана со второй и третьей, а вторая — с третьей; с остальными переменными у этих переменных связь слабая. Они образуют как бы одно целое, одну функциональную единицу — фактор. Если переменных *n,* то коэффициенты корреляций между переменными образуют

[2] *Spearman S.* General intelligence objectwely determined and measured. - «Amer. J. Psychob», 1904, N 15, p. 201—293.

[3] *Thurstone L. L.* Multiple factor analysis. Chicago, 1947.

квадратную симметричную матрицу порядка *n*. Здесь предполагается обычный коэффициент парной корреляции. В этом случае можно рассуждать и по-другому. В этом случае переменная может быть представлена как сумма факторов, умноженных на некоторые коэффициенты, которые определяются из матрицы корреляций. Будем искать такую переменную (фактор), когда при исключении ее влияния, частные коэффициенты корреляции между данными переменными будут равны нулю: $r_{ij,\ f1,} = 0$. Если же они не все оказались равными нулю, то ищем вторую переменную — фактор, чтобы при исключении действия этих двух факторов частные коэффициенты между данными переменными были бы равными нулю ($r_{ij,\ f1,f2} = 0$), и т. д.

Процесс обрывается, например, на *q* факторе, если при учете этих (q-факторов все частные коэффициенты между переменными будут равны нулю.

Теперь в несколько упрощенном виде выразим математически основную идею факторного анализа. Имеем N лиц и *n* переменных, т. е. эмпирические данные: $x_{ij}=1,\ \ldots,\ N;\ j=1,\ \ldots,\ n$. Можно изобразить это в виде матрицы:

$$X = \begin{Vmatrix} x_{11} \ldots x_{1N} \\ x_{21} \ldots x_{2N} \\ \ldots\ldots\ldots \\ \ldots\ldots\ldots \\ x_{n1} \ldots x_{nN} \end{Vmatrix}$$

Основная мысль факторного анализа[4] — представить эмпирические переменные в качестве линейных комбинаций меньшего числа некоторых других переменных, которые назовем факторами:

$$x_{ji} = a_{j1} f_{1i} + \ldots + a_{jq} f_{qi} = \sum_{k}^{q} a_{jk} f_{ki}. \qquad (1)$$

Оказывается, что матрица эмпирических данных имеет связь с матрицей корреляций между эмпирическими переменными. Это видно из следующих преобразований.

Прежде всего нормируем x_{ji}, т. е. положим, что

$$z_{ji} = \frac{x_{ji} - M_j}{\sigma_j},$$

[4] *Kendall M. G., Smith B. B.* Factor Analysis as Statistical technique.— «J. Roy. Statist. Soc. B», 1950, v. 12, p. 60.

где M_j — средняя арифметическая j-й переменной; σ_j — стандартное отклонение j-й переменной.
В этом случае

$$\bar{z}_j=\sum_i^N z_{ij}=0, \quad \sigma_{z_j}=\frac{1}{N}\sum_i^N z_{ji}^2=1,$$

т. е. средняя и дисперсия нормированных переменных соответственно равны нулю и единице.
Уравнение (1) можно переписать для нормированных данных:

$$z_{ji}=\sum_k^q a_{jk}f_{ki}. \qquad (1)$$

Впредь будем считать данные нормированными и использовать обозначения уравнения (1).
Уравнение (1) можно написать в матричной форме:

$$X=AF, \qquad (2)$$

где

$$A=\{a_{jk}\}=\begin{Vmatrix} a_{11}\dots a_{1q} \\ a_{21}\dots a_{2q} \\ \dots\dots \\ \dots\dots \\ a_{n1}\dots a_{nq}\end{Vmatrix};$$

$q<n,\ j=1,\ \dots,\ n;\ k=1,\ \dots,\ q;$

$$F=\{f_{ki}\}=\begin{Vmatrix} f_{11}\dots f_{1N} \\ \dots\dots \\ \dots\dots \\ f_{q1}\dots f_{qN}\end{Vmatrix};$$

$i=1,\ \dots,\ N;\quad k=1,\ \dots,\ q.$

$$r_{ke}=\frac{1}{N}\sum_a^N x_{ka}x_{ea} \qquad (3)$$

(4)

или в матричной форме

$$R=\frac{1}{N}XX'$$

где R — матрица корреляций:

$$R = \begin{vmatrix} r_{11} r_{12} \cdots r_{1n} \\ r_{21} r_{22} \cdots r_{2n} \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ r_{n1} r_{n2} \cdots r_{nn} \end{vmatrix}$$

X — матрица эмпирических данных; X'— транспонированная матрица X.

В уравнение (4) подставим уравнение (2):

$$R= \frac{1}{N} XX' = \frac{1}{N} AF\,(AF)\,' = A\,(\frac{1}{N} FF')\,A'$$

Выражение в скобках есть матрица корреляций между факторами. Будем считать, что факторы не коррелируют между собой или что они ортогональны. Тогда

$$\frac{1}{N} FF' = I,$$

где I — единичная матрица.

В таком случае имеем

$$R=AA'. \qquad (5)$$

Уравнение (5) запишем в алгебраической форме:

$$R_{kl}=\sum_a^q a_{ka}al_a \qquad (6)$$

$$R_{kk}=\sum_a^q a^2_{ka} \qquad (7^*)$$

Уравнение (5) является основой для реализации процедуры факторного анализа. Слева мы имеем эмпирические данные — матрицу корреляций, справа — неизвестные величины, которыми являются элементы матрицы факторных нагрузок.

Вообще говоря, уравнение (1) весьма редко имеет место. Обычно переменная не точно обусловлена факторами, а обусловлена с ошибкой:

$$Xji=\sum_k a_{jk}f_{ki}+e_{ji}$$

Где e_{ji} - величина с ошибки

Как бы мы ни подбирали факторы, они точно не воспроизведут эмпирические переменные, а всегда — с некоторым приближением. Ошибка приближения или остаток обозначается вектором *е*. Уравнение (7) является основным уравнением факторного анализа.

В нем x_{ji} — нормированы, f_{ki}— ортогональны и нормированы, e_{ji} — независимы, причем $\bar{e}=0$

Если бы мы имели дело с уравнением (1), то на главной диагонали матрицы корреляций стояли бы единицы. Но поскольку на практике приходится иметь дело с уравнением (7), то решение осложняется. Рассмотрим уравнение (6). Имеем $r_{kk}=1$. Величина $\sum a^2_{ki}$ представляет собой долю дисперсии факторов в общей дисперсии *i*-й переменной.

Обозначим

$$\sum_k a^2_{ik} = n^2_i$$

и назовем ее факторной дисперсией, которая в общем виде нам неизвестна. Она и будет стоять на диагонали матрицы корреляций.

Теперь встает проблема, как решить уравнения (5) и (7), т. е., зная эмпирическую матрицу корреляций, нужно найти неизвестную матрицу факторных нагрузок. Это означает, что,

зная $\frac{n(n+1)}{2}$ эмпирических коэффициентов корреляций, можно

найти неизвестные величины a_{jk}; $j = 1, \ldots, n$; $k=1, \ldots, q$. Естественно, что числа n и q должны быть в определенной зависимости. Аналитически можно показать [5], что они должны удовлетворять неравенству

$(n + q) < (n—q)^2$

Методы решения основных уравнений факторного анализа называются методами факторизации. В настоящее время используются в основном два метода — центроидный и максимального правдоподобия. Остановимся на первом.

В матричной форме это выглядит так:

$$X = AF+E; \qquad (8)$$

$$XX' = AA'+EE'; \qquad (9)$$

[5] *Лоули Дж., Максвелл А.* Факторный анализ как статистический метод. М., 1967, с. 136.

$$r_{ii} = \sum_k^q a_{ik}^2 + \sigma_i^2 \qquad (10)$$

$$r_{ij} = \sum_k^q a_{ik} a_{jk, i \neq j} \qquad (11)$$

Перед тем как перейти к решению уравнения (5) факторного анализа, определим значение элементов входящих в него матриц.
Имеем уравнение (1):
Множим $x_{ji} = \alpha_{j1} + f_{1i} + \ldots + \alpha_{jq} f_{qi}$
его на f1i, суммируем по всем i и делим на N:

$$\alpha_{j1} = \frac{1}{N} \sum_i x_{ji} f_{1i}$$

поскольку, как мы полагаем, факторы не коррелируют между собой и нормированы. Отсюда

$a_{i1} = r_{xif1}$

по определению коэффициента корреляции.

Аналогичную процедуру можно проделать для каждой переменной, и потому получаем равенство $a_{ja} = r_{xjfa}$.

Факторная нагрузка a_{ja} представляет собой коэффициент корреляции j-й переменной и α-го фактора.

Существо центроидного метода заключается в следующем. Рассмотрим q-мерное пространство факторов и в нем n-векторов (переменных). Можно выбрать систему векторов

$p_1 (\alpha_{11}, \ldots, \alpha_{1q})$,

$p_n (\alpha_{n1}, \ldots, \alpha_{nq})$

где a_{ij} — факторные нагрузки, $i = 1, \ldots, n; j = 1, \ldots, q$.

Ищем координаты центроида (центра тяжести) векторов. Каждая из q-координат центроида равна среднеарифметическому соответствующих координат, а именно:

$$\frac{1}{N} \sum_k^n a_{kj}, j = 1, \ldots, q$$

Полагаем, что центроид лежит на оси координат. Тогда

$$\sum_k^n a_{k2} = \ldots = \sum_k^n a_{kq} = 0$$

Его координаты суть

$$\frac{1}{N}\ \sum_{k}^{n} a_{k_1};0;...;0.$$

Найдем a_{k1}, $k=1, \ldots, n$; имеем

$r_{jk}=a_{j1} a_{k1} + a_{j2} a_{k2} + \ldots + a_{jq} a_{qk}$.

Суммируем:

$$\sum_{k}^{n} r_{jk} = a_{j1}\sum_{k}^{n} a_{k1} + \ldots + a_{jq}\sum_{k}^{n} a_{kq}$$

или

$$\sum_{k}^{n} r_{jk} = a_{j1}\sum_{k}^{n} a_{k1}$$

Еще раз суммируем:

$$\sum_{j}\ \sum_{k} r_{jk} = \sum_{j}^{n} a_{j1}\sum_{k}^{n} a_{k1} = (\sum_{k}^{n} a_{k1})^2$$

Обозначим:

$\sum_{j} r_{jk} \equiv r_k$ — сумма элементов k-го столбца матрицы корреляций,

$\sum_{j}\ \sum_{k} r_{jk} \equiv r$ — сумма всех элементов матрицы корреляций.

Имеем

$$\sum_{j} r_{jk} = a_{k1}\sqrt{r}\qquad\text{, или }\ a_{k1} = \frac{\sum_{j} r_{jk}}{\sqrt{r}}$$

Окончательно получаем $a_{k1} = r_k / \sqrt{r}$ и находим все факторные нагрузки a_{k1} первого фактора.

После этого находим остаточную матрицу корреляций:

$$r_{jk} = a_{j1}a_{k1} + a_{j2}a_{k2} + \ldots + a_{jq}a_{kq}$$

$$a_{j2}a_{k2} + \ldots + a_{jq}a_{kq} = r_{jk} - a_{j1}a_{k1}$$

Или в матричной форме:

$$W_1R=a_1a'_1,$$

где W_1 — остаточная матрица корреляций, a_1 — вектор-строка матрицы факторных весов.

С остаточной матрицей W_1 производим аналогичную процедуру, что и при извлечении первого фактора, с той только разницей, что отрицательные знаки у корреляций изменяем на положительные с тем, чтобы сделать дисперсию максимальной:

$$W_2=R-a_1a'_1-a_2a'_2.$$

Затем аналогично находим W_3 и т. д.

Количество шагов в нахождении остаточных матриц определяется числом факторов. В первоначальной матрице корреляций взяты приближенные значения факторных дисперсий. Как только при их значении определены факторные нагрузки, можно начать итерационный процесс — улучшить факторные дисперсии с помощью найденных факторных нагрузок; снова найти факторные нагрузки и снова улучшить факторные дисперсии и т. д. Как правило, итерация происходит достаточно быстро.

Найдем корреляцию между отметками 220 школьников по шести школьным предметам — французскому языку, английскому языку, истории, арифметике, алгебре, геометрии. Получаем матрицы I, II, III, IV (табл. 6, 7, 8, 9).

Таблица 6

I. Матрица корреляций*

Гальский язык	(0,439)	0,439	0,410	0,288	0,329	0,248
Английский язык	0,439	(0,439)	0,351	0,354	0,320	0,329
История	0,410	0,351	(0,410)	0,164	0,190	0,181
Арифметика	0,288	0,354	0,164	(0,595)	0,595	0,470
Алгебра	0,329	0,320	0,190	0,595	(0,595)	0,464
Геометрия	0,248	0,329	0,181	0,470	0,464	(0,470)
Суммы	2,153	2,232	1,706	2,466	2,493	2,162
Нагрузки фактора I	0,592	0,614	0,614	0,678	0,686	0,595

* *Лоули Дж., Максвелл А.* Факторный анализ как статистический метод, с. 42—46.

Таблица 7

II. Первая матрица остаточных ковариаций

I	1	2	3	4	5	6
1	(0,086)	0,076	0,132	–0,113	–0,077	–0,104
2	0,076	(0,062)	0,063	–0,062	–0,101	–0,036
3	0,132	0,063	(0,190)	–0,154	–0,132	–0,093
4	–0,113	–0,062	–0,154	(0,135)	0,130	0,067
5	–0,077	–0,101	–0,132	0,130	(0,124)	0,056
6	–0,104	–0,036	–0,098	0,067	0,056	(0,116)

Т а б л и ц а8

III. Остаточная матрица с измененными знаками с новыми оценками факторных дисперсий

Суммы	(0,132)	0,076	0,132	0,113	0,077	0,104
	0,076	(0,101)	(0,063)	0,062	0,101	0,036
	0,132	0,063	0,154	0,154	0,132	0,098
	0,113	0,062	0,154	(0,154)	0,130	0,067
	0,077	0,101	0,132	0,130	(0,124)	0,056
	0,104	0,036	0,098	0,067	0,056	(0,104)
	0,634	0,439	0,733	0,680	0,628	0,465
Нагрузки фактора II	0,335	0,232	0,387	0,359	0,332	0,246

Таблица 9

IV. Матрица факторных нагрузок

Факторы	Переменные					
	I	2	3	4	5	6
I	0,592	0,614	0,469	0,678	0,686	0,595
II	0,335	0,232	0,387	–0,359	–0,332	–0,246

В матрице I даны корреляции шести переменных. В клетках по диагонали — приближенные значения (в одном приближении как наибольшее из чисел в столбце) факторных нагрузок. Рассчитываются суммы элементов в столбцах и их общая сумма r_1. Затем из последней суммы извлекается корень. Факторные нагрузки фактора I—a_{j1} находятся делением каждой суммы столбца на $\sqrt{r_1}$

Если обозначим полученный вектор-строку факторных нагрузок фактора I f_1, то матрица II дает элементы остаточной матрицы $R—f_1\ f_1$, где f_1—вектор-столбец II с измененными знаками у переменных 4, 5 и 6. По диагоналям расположены новые приближенные оценки факторных дисперсий, взятые как наибольшие числа в соответствующих столбцах. Из этой матрицы определяются аналогичными способами, как и для фактора I, факторные нагрузки фактора II — $\mathbf{a}_{j2}$.

IV матрица — сводная таблица факторных нагрузок.

Можно строго показать, что факторы определяются с точностью до ортогонального преобразования или — в переводе на геометрический язык — с точностью до вращения. Можно так подобрать оси координат, чтобы переменные имели возможно большие нагрузки на один фактор и возможно меньшие (лучше нулевые) нагрузки на другие факторы. В этом случае факторы, по Терстону, образуют так называемую простую структуру.

Центроидный метод нашел широкое применение в силу своей простоты и доступности. Но в статистическом отношении он не совсем корректен, поскольку не дает возможности сделать выборочную оценку результатов. Наиболее разработанная процедура оценки факторных нагрузок предложена Лоули посредством метода максимального правдоподобия.

Другая проблема факторного анализа — проблема количественных и качественных данных. Техника извлечения факторов основывается на количественных данных. Используя другие коэффициенты корреляции, можно применять и качественные данные. Но в этом случае еще более неопределенней становится задача статистической оценки полученных факторных нагрузок.

Первоначально факторный анализ использовался в психологии. Известны работы Спирмена, Терстона, Томсона, Барта, Хорста, Гилфорда по применению факторного анализа в исследовании интеллекта, темперамента, памяти, способностей, сенситивных характеристик и прочих психологических элементов.

Начиная с 30-х годов факторный анализ используется в социальной психологии, социологии и других социальных науках.

У. Белл [6] применил факторный анализ к данным переписи по семи переменным.

1. Число рабочих на тысячу занятых лиц.
2. Число лиц 25 лет и старше с законченным или незаконченным средним образованием на тысячу лиц 25 лет и старше.
3. Средний доход.
4. Число детей на тысячу женщин в возрасте до 50 лет.
5. Число работающих на производстве женщин на тысячу женщин в возрасте от 17 лет и старше.
6. Процент семей, живущих в отдельных квартирах или домах.
7. Число иммигрантов на тысячу лиц.

Таблица 10

Матрица корреляций

	1	2	3		5	6	7
1		0,730	0,710	0,810	0,560	0,373	0,319
2	0,0730		0,696	0,650	0,277	0,047	0,649
3	0,710	0,696		0,538	0,311	0,049	0,356
4	0,810	0,650	0,538		0,690	0,560	0,383
5	0,560	0,277	0,311	0,690		0,680	-0,063
6	0,373	0,047	0,049	0,560	0,680		-0,030
7	0,319	0,469	0,356	0,383	-0,063	-0,030	

Таблица 11

Матрица факторных весов

Переменные	Факторы			
	I	II	III	IV
1	0,886	0,075	-0,233	0,845
2	0,777	0,511	0,088	0,873
3	0,693	0,390	-0,361	0,763
4	0,913	-0,189	0,089	0,877
5	0,646	-0,560	-0,185	0,765
6	0,485	-0,635	0,065	0,643
7	0,465	0,447	0,447	0,613

[6] *Bell W.* Economic, Family and Ethnic Status: an Empirical Test— «Amer. Soc, Rev.», 1955, v. 20, p. 45.

Таблица 12
Простая структура

Переменные	Факторы I	II	III	Переменные	Факторы I	II	III
1	0,482	0,193	–0,094	5	0,148	0,617	–0,193
2	0,319	–0,044	0,282	6	–0,147	0,727	0,015
3	0,653	–0,192	–0,189	7	–0,109	0,047	0,576
4	0,109	0,562	0,176				

Из матрицы простых структур следует, что выделено три фактора, которые суть:

I—экономический статус, объединяющий 1, 2, 3 переменные;

II—семейный статус, объединяющий 4, 5, 6 переменные;

III—национальный статус, обусловленный 7-й переменной.

Прайс рассмотрел 93 города США по 15 рубрикам для 1930 г.[7]:

1)население;
2)процент занятости в необслуживающей сфере;
3) соотношение полов;
4)процент прироста населения с 1925 по 1930 г.
5)средний месячный доход;
6)процент незанятого населения;
7)возраст города;
8)процент населения в возрасте от 15 до 50 лет;
9 процент работающих лиц со стажем в 10 лет и выше;
10)процент семейных рабочих;
11)средний объем семьи;
12)оптовая торговля на душу населения;
13)розничная торговля на душу населения;
141 относительный рост заработка;
15) процент налогоплательщиков.

Таким образом, дана эмпирическая матрица. По строкам расположены значения данных 15 переменных для каждого из 93 городов, по столбцам — значения каждой переменной для всех 93 городов. Матрица имеет 93 строки и 15 столбцов. Были получены коэффициенты корреляций для данных 15 переменных и благодаря матрице корреляций (15x15) найдены четыре фактора. Первый фактор наиболее сильно коррелирует с переменными 7, 1, 15, т. е. возрастом города, объемом населения, числом налогоплательщиков, торговлей на душу населения. Он может рассматриваться как экономический фактор (табл. 13).

[7] Price D. O. Factor Analysis in the Study of Methropoliten Centres.— «Social Forces», 1942, v. 20, N 4, p. 449.

Далее выводятся индексы городов по каждому фактору по

формуле $f_1 = \sum_{i-1}^{15} a_{ij}z_i$,

где f_j — индекс для фактора j, a_{ij} — факторный вес i–й переменной по j–му фактору, z_i — стандартный балл *i–й* переменной. Для каждого фактора города ранжируются по величине его индекса.

Таблица 13

Таблица факторных весов для 15–ти переменных

	Факторы			
Переменные	I	II	III	IV
1	0,6401	0,0767	0,2927	–0,1172
2	–0,0439	–0,7251	–0,0181	–0,0888
3	–0,0792	–0,3338	0,7905	0,1609
4	–0,3761	0,3891	0,3023	0,0259
5	0,5124	0,2212	0,6022	0,0090
6	0,1178	–0,3537	0,0620	–0,0343
7	0,7734	0,0698	0,0357	–0,0707
8	0,0057	0,6399	0,4464	0,0652
9	0,2233	0,6283	–0,0588	0,0306
10	0,2546	0,5121	–0,7299	–0,1637
11	0,0732	–0,6473	–0,2331	–0,1915
12	0,3123	0,2410	–0,0801	0,7991
13	0,3345	0,2564	0,1236	0,8367
14	0,2063	–0,2817	0,7849	0,0986
15	0,5008	0,1332	0,3609	0,0908

Мозер исследовал 157 городов по 57 характеристикам и получил такие четыре фактора: классовость, изменение населения за 1931 – 1951 гг., изменение населения за 1951–1958 гг. и пере–населенность города[8].

Из советских социологов Т. И, Заславская применила фактор–ный анализ в исследовании причин миграции сельского на–селения*. По результатам анализа она пришла к выводу,

[8] *Moser* C., *Scott W.* British Town. Edinburgh – London, 1961.
[9] Количественные методы в социальных исследованиях. – Информ. бюлл. ИКСИ АН СССР, вып. 9. М.,1968, с. 34.

Таблица 14
Распределение признаков по факторам (в зависимости от максимальных весов)

№ фактора	№ признака	**Признак, имеющий максимальный вес по данному фактору**	**Коэффициент связи между показателем и фактором**
I	29	Число врачей на 1000 сельских жителей	0,90
	49	Оборот розничной торговли на сельского жителя	—0,89
	2	Средняя оплата рабочего дня в совхозах	0,88
	42	Число кинопосещений на одного жителя в год	0,84
	31	Число медработников на 1000 сельских жителей	0,82
	1	Изменение численности рабочей силы совхозов	0,74
	13	Потребление электроэнергии в быту	0,67
	12	Обеспеченность жильем за счет совхозов	0,66
	25	Доля молодежи среди сельского населения	0,66
	47	Число учителей на 1000 сельских жителей	0,58
	41	Процент детей в детских учреждениях	0,34
II	57	Естественный прирост населения, %	0,68
	26	Доля лиц со средним и высшим образованием	0,60
	7	Плотность сельского населения	0,60
	11	Плотность железных и шоссейных дорог	0,57
	10	Доля лиц коренной национальности	0,54
III	15	Процент домов без электричества	0,67
	3	Число рабочих дней в году на работника	0,66
	53	Доля женщин среди работников совхозов	0,61
	14	Процент сельского населения в районе	0,52
IV	6	Средний размер населенного пункта	0,67
	4	Средний доход от личного подсобного хозяйства	0,43

что в миграции играют роль два главных фактора. Первый связан с материальным и культурным благосостоянием сельского населения района. Второй—с уровнем жилищно-бытового строительства.

Другим примером применения факторного анализа может служить анализ структуры признакового пространства, описывающего условия труда и жизни сельского населения различных районов[10]. Для испытания было отобрано 22 показателя. Весь анализ можно разделить на четыре стадии. Первая стадия — получение так называемой матрицы интеркорреляций.

Вторая стадия — это последовательное преобразование исходной матрицы и выполнение расчетов, направленных на «извлечение» независимых факторов, характеризующих внутреннюю структуру изучаемого признакового пространства.

Третья стадия представляет собою специальную операцию — поворот осей, которая результируется в составлении окончательной таблицы данных связи между признаками и факторами; Рассматривая, как улучшились качественные характеристики матрицы в результате поворота осей, авторы Т. И. Заславская и Е. В. Виноградова делают следующее заключение: «Несмотря на то, что использованные методы поворота осей носили приближенный характер и не обеспечивали оптимального результата, эффективность этой операции очевидна. Количество нежелательных средних весов уменьшилось почти вдвое, заметно повысилось число показателей, имеющих четко выраженные максимумы по отдельным факторам при малых значениях весов по другим. Показатели более равномерно распределились по факторам, что облегчило возможность предметного толкования последних» [11].

Последняя стадия факторного анализа заключается в трактовке результатов. Анализируя данные о распределении признаков по факторам в зависимости от максимальных весов, сведенные в специальную таблицу, авторы дают специфическое толкование каждому из четырех выделенных факторов. Тем самым каждый из выделяемых факторов получает содержательную характеристику через систему отношений к заданным внешним признакам. Первый фактор, объединяющий признаки 29, 49, 2, 42, 31, 1, 13, 12, 25, 47, 41, характеризуется авторами как уровень материально-бытовых и социально-культурных условий жизни сельского населения; второй, объединяющий признаки 57, 26, 7, 11, 10,—как структура сельского населения районов; третий, объединяющий признаки 15, 3, 53, 14, — как уровень экономического и технического развития района; и, наконец, четвертый, объединяющий признаки 6 и 4,— как характер сельского расселения (табл. 14).

[10] Социальные проблемы трудовых ресурсов села. Под ред. Т. И. Заславской. Новосибирск, 1968, с. 107—124.
[11] Социальные проблемы трудовых ресурсов села, с. 116.

Во всех рассмотренных случаях использовались корреляции между переменными. Математически совершенно равноправна операция использования корреляций между лицами, т. е. между строками в эмпирической матрице. Это так называемая Q-техника, в отличие от наиболее употребительной .R-техники. Q-техника приводит к нахождению факторов среди лиц (объектов), т. е. лица объединяются в группы-факторы. Эта техника весьма перспективна в социологии, хотя она и сопряжена с более трудоемкими операциями в сравнении с R-техникой [12].

Применение факторного анализа связано с математическими трудностями и с вопросом содержательной интерпретации факторов. Преодолеть эти трудности можно только широким экспериментированием по трем направлениям, применяя ралличные методы факторизации к разным выборкам, разным лицам и разным проблемам, что в целом и делается в большей части современных социологических исследований. По словам известного математика и психолога П. Хорста, «многие другие возможности -применения факторного анализа, без сомнения, будут обнаружены в будущем, потому что роль факторного анализа значительна в систематическом научном исследовании во всех областях; его использование будет расширяться, его техника улучшаться, методы анализа — становиться более общими и доступными благодаря вычислительным машинам с большими скоростями работы» [13].

[12] *Cattet R. B.* Personality and Social Psychology. San Diego, 1964.
[13] *Horst P.* Factor analysis of data matrices. N. Y., 1965, p. 142.