

**Д. Лоули**

**Факторный анализ как  
статистический метод**

**Москва  
«Книга по Требованию»**

УДК 51  
ББК 22.1  
Д11

Д11 **Д. Лоули**  
Факторный анализ как статистический метод / Д. Лоули – М.: Книга по Тре-  
бованию, 2024. – 145 с.

**ISBN 978-5-458-27352-7**

Настоящая монография даёт чёткое представление о задачах факторного анализа, который находит всё большее применение при решении различных практических задач. Факторный анализ возник и развивался в связи с решением в первую очередь задач психологии (например, объяснение успеваемости учеников). Однако область его приложений значительно шире и охватывает в принципе все случаи применения многомерного статистического анализа.

**ISBN 978-5-458-27352-7**

© Издание на русском языке, оформление  
«YOYO Media», 2024

© Издание на русском языке, оцифровка,  
«Книга по Требованию», 2024

Эта книга является репринтом оригинала, который мы создали специально для Вас, используя запатентованные технологии производства репринтных книг и печати по требованию.

Сначала мы отсканировали каждую страницу оригинала этой редкой книги на профессиональном оборудовании. Затем с помощью специально разработанных программ мы произвели очистку изображения от пятен, клякс, перегибов и попытались отбелить и выровнять каждую страницу книги. К сожалению, некоторые страницы нельзя вернуть в изначальное состояние, и если их было трудно читать в оригинале, то даже при цифровой реставрации их невозможно улучшить.

Разумеется, автоматизированная программная обработка репринтных книг – не самое лучшее решение для восстановления текста в его первоизданном виде, однако, наша цель – вернуть читателю точную копию книги, которой может быть несколько веков.

Поэтому мы предупреждаем о возможных погрешностях восстановленного репринтного издания. В издании могут отсутствовать одна или несколько страниц текста, могут встретиться невыводимые пятна и кляксы, надписи на полях или подчеркивания в тексте, нечитаемые фрагменты текста или загибы страниц. Покупать или не покупать подобные издания – решать Вам, мы же делаем все возможное, чтобы редкие и ценные книги, еще недавно утраченные и несправедливо забытые, вновь стали доступными для всех читателей.



Серия Книжный Ренессанс

[www.samizday.ru/reprint](http://www.samizday.ru/reprint)



## ПРЕДИСЛОВИЕ ПЕРЕВОДЧИКА

В настоящее время мы часто сталкиваемся с исследованием многопараметрических «объектов», и каждый раз перед нами встает вопрос: нельзя ли отбросить часть параметров или заменить их меньшим числом каких-либо функций от них, сохранив при этом всю информацию? Однако этот вопрос приобретает смысл только в рамках какой-либо определенной модели, и его решение зависит, конечно, от структуры этой модели.

Модель факторного анализа является достаточно хорошим приближением ко многим реальным ситуациям, и для нее поставленные выше вопросы вполне осмыслимы и решаемы. Зародившийся в среде психологов еще в начале нашего века факторный анализ долгое время оставался скорее искусством, чем наукой. Только появление новых методов и переосмысление старых внутри самой математической статистики привело ко вторичному рождению факторного анализа — одного из трудных и интересных разделов современной многомерной статистики.

С каждым годом развиваются методы факторного анализа, расширяется сфера его приложений. К нему кроме психологов все чаще обращаются в своих работах медики, экономисты, химики и другие исследователи.

Авторы книги являются одними из наиболее крупных специалистов в этой области. Их содружество (один из них психолог, другой — математик) позволило создать, по-видимому, лучшую из имеющихся книг по факторному анализу.

Предназначена она, как отмечают и сами авторы, не только для специалистов по математической статистике, но и для всех тех, кто желает более строго применять факторный анализ в своих исследованиях и более объективно оценивать результаты этих исследований.

Дать математически строгое описание современного состояния факторного анализа, оттенить слабые и сильные стороны разных подходов к решаемым задачам и сделать предлагаемые алгоритмы доступными для широкого круга читателей в книге весьма небольшого объема — задача очень трудная и на первый взгляд нереальная. Однако авторам во многом удалось с ней справиться, и прежде всего за счет сжатости изложения математических фактов.

Хотя те или другие выводы, приводимые в книге, и не излагаются в скрупулезной математической последовательности, авторы опускают промежуточные выкладки лишь тогда, когда они проводятся стандартными методами. Авторы отступают от этого правила в разделах, посвященных критериям значимости, где математические результаты лишь описываются и даются ссылки на соответствующие статьи. Справедливости ради отметим, что многие из результатов по критериям значимости пока еще не обоснованы строго, а излагать «полуматематические» их обоснования едва ли полезно.

Такой подход позволил авторам охватить большинство применяемых в настоящее время методов, оценить их и дать широкое представление о состоянии факторного анализа как науки.

С другой стороны, это позволило четко изложить алгоритмы решений различных задач разными методами на примерах (в большинстве своем взятых из конкретных работ психологов), дать подробные объяснения по выбору модели, интерпретации выводов, оценки точности решения и степени достоверности выбранной модели на тех же примерах. Эти разделы могут использоваться для приложения факторного анализа и неспециалисты по математической статистике.

В конце книги авторы дают приложение, содержащее необходимые сведения из линейной алгебры. Переводчиком написано еще одно приложение, в котором дается обзор основных статистических методов, связанных по существу с задачами, рассматриваемыми авторами; как и авторское приложение, оно рассчитано на неспециалистов.

Думается, выход этой книги на русском языке восполнит серьезный пробел в отечественной литературе по статистике и привлечет к факторному анализу не только «потребителей», но и математиков, что будет способствовать развитию этой науки в нашей стране.

*Ю. Н. БЛАГОВЕЩЕНСКИЙ*

## ПРЕДИСЛОВИЕ АВТОРОВ

Мы весьма обязаны издательству Баттерворс, предложившему нам написать эту книгу. Решившись принять предложение, мы имели в виду две цели. Первая — преподнести статистическую теорию факторного анализа в ее развитии до современного уровня в сжатой форме, которая, однако, позволила бы математику-статистику легко понять, о чем идет речь. Вторая цель — проиллюстрировать теорию на некоторых тщательно подобранных примерах так, чтобы пользующиеся факторным анализом, которых становится все больше и больше, могли применять его более строго и точно, чем раньше.

В факторном анализе более чем в других ветвях многомерного анализа приходится сталкиваться с большим объемом вычислений. Однако появление электронных вычислительных машин в значительной степени способствовало преодолению этого затруднения.

Мы благодарны сотрудникам Вычислительного центра Лондонского университета В. Никсону и Дж. Галлагер, которые помогли нам составить программу для двух наиболее трудоемких итерационных процессов.

Научных работников часто осуждают за плохой язык. Любые литературные достоинства, какими, может быть, обладает эта книга, в большой степени являются заслугой наших жен.

*Д. Н. ЛОУЛИ  
А. Э. МАКСВЕЛЛ*

## Глава I

### СОДЕРЖАНИЕ ФАКТОРНОГО АНАЛИЗА

#### 1.1. Введение

Факторный анализ является той ветвью многомерного анализа, которая исследует внутреннюю структуру матриц ковариаций и корреляций. Это направление первоначально разрабатывалось главным образом психологами, наиболее видные из которых — Спирман, Томсон, Тэрстон и Барт, и в первую очередь касалось гипотез о характере умственных способностей, поданных исследованием матрицы корреляций между специально подобранными показателями. Ранние работы в этой трудной области вызвали затянувшиеся дискуссии о психологической стороне вопроса, из-за чего проявленный математиками интерес к входящим сюда статистическим проблемам на время упал, и этот раздел стал запущенной главой в статистической теории. Однако «факторный анализ, — как заметил Бартлетт (Bartlett, 1953, стр. 23), — должен рассматриваться как естественное и неизбежное развитие анализа коррелированного набора контрольных тестов и других величин, с которыми психологи могут иметь дело»; постепенно статистическая теория такого анализа (см. ссылки на литературу в конце этой главы) была создана. В настоящее время факторный анализ представляет большую часть широко (хотя и не всегда по назначению) используемой многомерной техники, как указывает Хотеллинг (Hotelling, 1957). Применение факторного анализа облегчено также появлением электронных вычислительных машин и распространяется, кроме психологии, на другие дисциплины: экологию, ботанику, биологию, а также и на социальные науки, так что попытка изложить статистическую теорию этого раздела весьма желательна.

В этой вводной главе будет дан общий обзор проблем, с которыми сталкивается факторный анализ, и моделей, которые при этом анализе используются; в последующих главах дается разработка конкретных тем. Вследствие ограниченности места иллюстративные примеры, используемые в этой книге, даны в незначительном количестве и не вполне типичны для исследования факторов, где матрицы корреляций, содержащие сорок и более переменных, не удивительны, однако они достаточны, чтобы проиллюстрировать теорию предмета. Эти примеры главным образом психологические, поскольку и большинство опубликованных работ — работы психологические, но общие принципы факторного анализа могут быть легко обобщены и применены в других областях.

## 1.2. Модели факторного анализа

При анализе структуры ковариационных (или корреляционных) матриц чаще всего используются два подхода, формально до некоторой степени похожих друг на друга, однако различных по целям. Одним из них является метод главных компонент по Пирсону (Pearson, 1901) и Хотеллингу (Hotelling, 1933), другим — факторный анализ, возникший как метод в работах Спирмана (Spearman, 1904, 1926). Следует различать эти два подхода, и, хотя в этой книге изучается прежде всего факторный анализ, в ней будет описан и метод главных компонент. Последний по существу совпадает с методом расчленения ковариационной или корреляционной матрицы на совокупность ортогональных векторов (компонент) или направлений по числу рассмотренных переменных. Эти векторы соответствуют собственным векторам и собственным значениям корреляционных матриц (см. приложение I). По этому методу собственные значения выделяются в порядке убывания их величины, что становится существенным, если для описания данных должно быть использовано лишь незначительное число компонент. Векторы попарно ортогональны, и компоненты, полученные по ним, некоррелированы. Хотя

несколько компонент могут выделить большую часть суммарной дисперсии переменных, однако для точного воспроизведения корреляций между переменными требуются все компоненты. В тех случаях, когда применяется метод главных компонент, не нужно делать никаких гипотез о переменных, они не обязаны даже быть случайными величинами, хотя на практике их наблюдаемые значения рассматриваются как выборки из некоторой популяции.

В противоположность методу главных компонент факторный анализ заранее объясняет матрицу ковариаций наличием минимального или по крайней мере небольшого числа гипотетических переменных или факторов.

Переведем это на язык корреляций. Первый возникающий вопрос: существуют ли значимые корреляции, т. е. отличается ли корреляционная матрица значимо от единичной матрицы? Если исследователь убежден, что это так, тогда он спрашивает: существует ли случайная величина  $f_1$ , такая, что попарные корреляции между переменными равны нулю, если влияние  $f_1$  уже учтено? Если после этого корреляционная матрица еще не объяснена, он исследует: существуют ли такие две случайные величины  $f_1$  и  $f_2$ , что корреляции между парами переменных равны нулю, если влияние  $f_1$  и  $f_2$  учтено, и так далее... (Howe, 1955, стр. 8 и след.). Таким образом, можно сказать, что, в то время как метод главных компонент ориентирован на *дисперсии*, факторный анализ ориентирован на *ковариации* (или на корреляционную связь).

Цели двух этих методов анализа можно также сформулировать в терминах наблюдаемых переменных и искомых компонент или факторов. Пусть  $x_1, x_2, \dots, x_p$  суть  $p$  наблюдаемых переменных. В компонентном анализе к ним применяется ортогональное преобразование для получения некоррелированных переменных  $y_1, y_2, \dots, y_p$ . Они выбираются так, что  $y_1$  имеет максимум дисперсии,  $y_2$  — максимум дисперсии при требовании некоррелированности с  $y_1$  и т. д., причем относительно переменных  $x_i$  не надо делать никаких предположений.

В факторном анализе основным предположением является равенство

$$x_i = \sum_{r=1}^k l_{ir} f_r + e_i \quad (i = 1, 2, \dots, p), \quad (1.1)$$

где  $f_r$  есть  $r$ -й простой фактор<sup>1)</sup>,  $k$  точно задано и  $e_i$  — остатки, которые представляют источники отклонений, действующие только на  $x_i$ . Эти  $p$  случайных величин  $e_i$  предполагаются независимыми как между собой, так и с  $k$  величинами  $f_r$ . Вначале мы будем предполагать, что  $f_r$  некоррелированы, однако позднее мы рассмотрим и случай коррелированных факторов. Не умаляя общности, мы можем считать дисперсию каждого из  $f_r$  равной единице. Дисперсии  $e_i$  обозначим через  $v_i$ . Все средние предполагаются равными нулю. Коэффициент  $l_{ir}$  обычно называется или нагрузкой  $r$ -го фактора в  $i$ -й переменной, или нагрузкой  $i$ -й переменной на  $r$ -й фактор. Значения  $l_{ir}$ , а также  $v_i$  считаются обычно неизвестными параметрами, подлежащими оценке. Здесь можно упомянуть, что были сделаны попытки считать индивидуальные значения  $f$  параметрами, но это привело к трудностям, важнейшая из которых состоит в том, что число параметров растет до бесконечности при неограниченном увеличении объема выборки (Whittle, 1953). Этот подход в нашей книге не рассматривается.

Очевидно, что уравнения (1.1) нельзя проверить непосредственно, поскольку  $p$  переменных  $x_i$  выражены в них через  $(p+k)$  других ненаблюдаемых переменных. Но эти уравнения заключают в себе гипотезу о ковариациях и дисперсиях  $x_i$ , которую можно проверить (см. гл. 2).

Следует указать еще на одно различие между методом главных компонент и факторным анализом, замеченное Бартлеттом (Bartlett, 1953, стр. 32 и след.). Первый из этих методов по определению является линейным и аддитивным и даже не ставит вопроса о

<sup>1)</sup> Термин «простой фактор» („common factor“) часто будет заменяться термином «фактор», что используют и авторы этой книги. — *Прим. перев.*

гипотезе, тогда как второй несет в себе так называемую *гипотезу линейности*, которая, если она не верна, может быть использована как первое приближение. Однако она будет отвергнута как линейная модель, постулируемая уравнением (1.1), если фактические данные приведут к этому.

Так как корреляция по существу относится к линейным связям, то модель в таком виде не пригодна для манипуляций, и Бартлетт вкратце показывает, как можно подправить структуру уравнений для улучшения соответствия модели: добавить (как следующее приближение) вторые степени и произведения факторов. В подправленных уравнениях особый интерес представляет произведение факторов; так, равенства

$$x_i = l_{11}f_1 + l_{12}f_2 + e_i, \quad (1.2)$$

которые содержат ровно два фактора, станут такими:

$$x_i = l_{11}f_1 + l_{12}f_2 + l_{13}f_3 + e_i, \quad (1.3)$$

где  $f_3 = f_1f_2$ . Несмотря на то, что детали этого более сложного варианта еще только разрабатываются, упоминание о нем полезно, чтобы напомнить предположения линейности, содержащиеся в уравнении (1.1), и подчеркнуть различие между факторным анализом и эмпирической природой компонентного анализа.

### 1.3. Факторные концепции и вращение

В предыдущем параграфе была предпринята попытка объяснить различие между методом главных компонент и факторным анализом. В главе 2 будет упомянуто еще одно важное отличие, касающееся определенных свойств инвариантности при изменении масштаба в переменных, которыми обладают факторные нагрузки и не обладают веса в компонентном анализе.

В любом из этих случаев либо компоненты (в компонентном анализе), либо факторы (в факторном анализе) получаются как взвешенные суммы переменных,

и решение того, какой метод анализа применить в каждом практическом исследовании, должно зависеть от целей и вопросов, поставленных перед исследованием. Большинство психологов сходится, однако, на том, что факторная модель лучше согласуется с их нуждами. Это подтверждается и теоретически, и практически. С практической стороны факторная модель дает возможность адекватно рассчитывать ковариационную структуру между относительно большим числом наблюдаемых переменных посредством меньшего числа простых факторов. С теоретической стороны — это вера в то, что характеристики человеческого поведения, какова бы ни была их природа, можно описать в терминах сравнительно небольшого числа психологических факторов или признаков или во всяком случае показать, что они действуют так, как если бы они были устроены таким образом.

В ситуациях с одним фактором ( $k=1$ ) достаточно рассчитать корреляцию между переменными, чтобы нагрузки однозначно определились. Однако, когда  $k>1$ , ни факторы, ни нагрузки не определяются однозначно, поскольку в уравнении (1.1) факторы  $f_r$  могут быть заменены любым ортогональным преобразованием их с соответствующим преобразованием нагрузок. Это свойство использовано для преобразования или вращения факторов, полученных в каком-либо практическом исследовании. Вращение подбирается так, чтобы переменные, которые в большей или меньшей степени измеряют некоторые легко опознаваемые стороны поведения (например, «Двигательная реакция» или «Общая активность»), имели бы столь высокие нагрузки на один фактор, насколько это возможно (этому фактору и присваивается соответствующее название), и нулевые или почти нулевые нагрузки на другие факторы.

Отсутствие единственности сочетается с тем, что разные экспериментаторы иногда имеют дело с факторами, повернутыми к различным положениям, но называют эти факторы одинаково; это приводит к большому количеству недоразумений при интерпретации факторов.