

УДК 159.99

Психология и математическая статистика: перспективы XXI века

Александр В. Дятлов^{1*}, Ирина В. Абакумова²

¹ Южный федеральный университет, г. Ростов-на-Дону, Российская Федерация

² Донской государственный технический университет, г. Ростов-на-Дону, Российская Федерация

* E-mail: avdyatlov@sfedu.ru

Аннотация

Массовое применение компьютеров в анализе данных результатов психологических исследований и простой доступ к программному обеспечению с мощными вычислительными возможностями (SPSS, Statistica и аналогичные) – главная особенность применения методов математической статистики в психологии. Особенно интенсивно в последнее время стали применяться многомерные методы анализа психологических данных, которые из-за сложности расчетов до недавнего времени рассматривались только как теоретические. Взаимное проникновение математической статистики и психологии привело к развитию новых методов моделирования и объяснения различных типов психологических данных.

Однако этот процесс, по сути, как «палка о двух концах». Многомерные статистические методы (далее МСМ) предполагают четкое построение модели и жесткие требования к дизайну психологического эксперимента. Но популярность МСМ совершенно не делает их проще как в применении, так и последующей интерпретации результатов. Предлагаемая статья – своеобразный обзор современного состояния использования и принципов применения МСМ в психологических исследованиях. Описаны этапы использования методов математической статистики в психологических исследованиях в соответствии с глубиной и сложностью рассматриваемых моделей, а также типов данных с которыми они работают. Рассмотрены важные принципы, которые лежат в основе применения методов МСА, сформирована дефиниция, определяющая содержание понятия многомерная статистика (многомерный статистический анализ).

В данной статье представлена классификация по различным критериям методов математической статистики, наиболее часто применяемых в психологии. Кроме того, была предпринята попытка описать кратко будущее статистических методов в психологических исследованиях.

Ключевые слова

психологические исследования, моделирование, анализ данных, многомерные методы математической статистики

Psychology and mathematical statistics: prospects for the twenty-first century

Aleksandr V. Dyatlov¹, Irina V. Abakumova²

¹ Southern Federal University, Rostov-on-Don, Russian Federation

² Don State Technical University, Rostov-on-Don, Russian Federation

* Corresponding author. E-mail: avdyatlov@sfedu.ru

Abstract

The mass use of computers in the analysis of data from psychological research results and easy access to software with powerful computing capabilities (SPSS, Statistica, and similar) is the main feature of applying mathematical statistics methods in psychology. Multidimensional methods of analysis of psychological data, which were considered only theoretical until recently due to the complexity of calculations, have been used especially intensively in recent years. The mutual penetration of mathematical statistics and psychology has led to the development of new methods for modeling and explaining various types of psychological data. However, this process is essentially a "double-edged sword". Multidimensional statistical methods (hereinafter MSM) assume a clear construction of the model and strict requirements for the design of a psychological experiment. But the popularity of MSM does not make them easier to apply or interpret the results. This article is a kind of review of the current state of use and principles of MSM application in psychological research. The stages of using mathematical statistical methods in psychological research are described in accordance with the depth and complexity of the models under consideration, as well as the types of data they work with. The important principles that underlie the application of ISA methods are considered, and the definition that defines the content of the concept of multidimensional statistics (multidimensional statistical analysis) is formed.

This article presents a classification by various criteria of mathematical statistics methods that are most often used in psychology. In addition, an attempt was made to briefly describe the future of statistical methods in psychological research.

Keywords

psychological research, modeling, data analysis, multidimensional methods of mathematical statistics

Введение

Цель этой статьи – провести обзор многомерных методов математической статистики, которые применяются в психологических исследованиях. В историческом плане отношения между математической статистикой и психологией берут своё начало в конце XIX-го – начале XX-го века, несмотря на то, что еще в III веке до новой эры Аристотель использовал, по сути, многомерный подход к классификации объектов в соответствии с их сходством или различием.

Многомерный статистический анализ (далее по тексту МСА) имеет свою собственную теоретическую базу, которая оформилась в связке с психологией, и базируется, в своей основе, на трудах Карла Пирсона и Чарлза Спирмена относящихся к факторному анализу. Не вызывает сомнения, что появление многомерного статистического анализа основано на идеи, имеющей цель ответить на сложные психологические вопросы.

Ретроспективный обзор развития МСА показывает, что период наиболее интенсивного развития их теоретической базы, приходится на 20-е – 50-е годы прошлого века. С середины XX века и по настоящее время акцент делался и делается, прежде всего, на использование уже разработанных методов. Но наряду с этим создается и обосновывается и ряд новых моделей, в частности, для анализа и объяснения неметрических данных.

Теоретическое обоснование

Использование методов математической статистики в психологических исследованиях условно может быть разделено на три этапа в соответствии с глубиной и сложностью рассматриваемых моделей, а также типов данных, с которыми они работают.

Первый этап в использовании статистических методов в психологии связан, прежде всего, с установлением взаимосвязи между явлениями. Свойства шкал, по которым измеряются явления, определяют, будут ли это коэффициент Пирсона или будут основаны на таблице совместного распределения или χ^2 -квadrата. Попытки моделировать взаимосвязи и строить формальную причинно-следственную модель предполагает использование регрессионный анализ и модели лог-линейной регрессии.

На этот первый этап также приходится разработка моделей с латентной переменной. Современный подход к проверке гипотез, планированию эксперимента, анализу мощности и метод формального выбора модели может быть отнесен к этому этапу.

С ростом объема данных и изменения качества их измерения появляется необходимость новых методов, которые способны показать информацию, содержащуюся в этих данных. Эта потребность знаменует собой второй этап в использовании методов математической статистики. Примечательной особенностью является

ПСИХОЛОГИЯ

разработка структурных моделей, имеющих общее название LISREL, представляющих собой статистический программный пакет, используемый в моделировании структурных уравнений (SEM) для явных и скрытых (латентных) переменных. В них заложена следующая идея: во многих ситуациях целевые переменные в причинно-следственной модели не могут наблюдаться непосредственно. Для расчета целевых переменных могут быть измерены и другие переменные, связанные с изучаемым конструктом, при условии, что структурные модели связывают наблюдаемые переменные с изучаемым конструктом. Другим достижением в применении статистических методов на этом этапе является разработка подходов для анализа неметрических данных и, прежде всего, модель логистической регрессии и её обобщение – модель множественной логистической регрессии. Применение этих моделей выходит за рамки психологии и находит место и в таких областях знания как медицина, социология, биология, экономика. Также необходимо отметить и развитие на этом этапе методов анализа недостающих данных и анализа событий во времени (event history analysis). Особый интерес для психологии представляют иерархические модели, которые появились в середине восьмидесятых годов прошлого века. Эти модели наиболее полно учитывают структуру генеральной совокупности, из которой были извлечены данные для формирования выборки, а при анализе и моделировании этих данных объединяются идеи и классического частотного, и байесовского подходов.

Установление и моделирование причинно-следственных связей и отношений является одной из главных задач психологии. Математическая статистика избегает заниматься этим вопросом непосредственно, но у неё есть методы и средства для проверки формальных моделей, поступающих извне. Такие модели, прежде всего, являются результатами психологических экспериментов. Среди современных методов анализа экспериментальных данных особенно претерпели развитие методы, которые имеют объектом повторные измерения. Наряду с этим очевидным является и прогресс в развитии существующих и разработке новых методов планирования контролируемых экспериментов.

Третий этап в развитии применения методов математической статистики приобретает очертания в настоящее время. Особый интерес представляют современные методы сбора и генерации огромных массивов данных за короткое время. Классические методы не в состоянии справиться с анализом таких массивов данных. Например, всем известна проблема одновременной проверки многих гипотез, когда с ростом числа гипотез растёт и вероятность ошибки первого рода. Одним из первых достижений математической статистики на этом этапе является разработка методов, которые могут устранить эту проблему. Эти методы известны как «ожидаемая доля ложных отклонений» (false discovery rate). Ещё одной тенденцией является разработка новых подходов при анализе качественных и текстовых данных и их ввод в современные статистические

пакеты. Совершенствование некоторых традиционных методов также является актуальной задачей современного этапа. В качестве примера можно упомянуть назревшую необходимость усовершенствования в методах анализа неравномерности повторных измерений (Lindsey, 1999).

Но этот обзор нельзя назвать исчерпывающим, ибо нашей целью было только желание наметить (довольно условно) этапы в развитии применения математической статистики.

В разработке и применении методов МСА, и в первую очередь в психологии, наиболее значительно прослеживается влияние и участие представителей трех школ (Сошникова, 1999):

- американской, с основными достижениями в области факторного анализа, многомерного шкалирования, новейшими концепциями в прикладной статистике: теорией нечетких множеств, анализом отсутствующих значений, анализом неметрических данных, графическими методами анализа и представлении многомерных данных;
- английской, с основными достижениями в области факторного анализа, дискриминантного анализа, многомерного корреляционного и регрессионного анализа, многомерными статистическими методами;
- французской, с основными достижениями в области кластерного и корреспондентного анализа.

Прежде всего, мы сформулируем, чем занимается многомерная статистика. До настоящего времени, к сожалению, не сформировано строгой дефиниции, определяющей содержание понятия «многомерная статистика». Для последующих рассуждений будем использовать достаточно гибкое и непретенциозное следующее определение: под многомерным статистическим анализом (МСА) следует понимать множество формальных методов, которые имеют в основе представление эмпирической информации в многомерном геометрическом пространстве и позволяют моделировать неявные (латентные), но объективно существующие закономерности.

Из предлагаемой формулировки становится ясно, почему геометрические представления играют столь важную роль в применении МСА в психологических исследованиях.

Рассмотрим несколько важных принципов, которые лежат в основе применения методов МСА (Сошникова, 1999).

1) Эффект действительной размерности. Этот принцип предполагает, что изучению подлежит не произвольное множество атрибутов объекта, а ограниченный круг логически связанных и дополняющих друг друга признаков, которые позволяют наиболее полно описать изучаемое явление.

2) Минимальное описание объекта. Этот принцип предполагает максимально лаконичное и строго структурированное описание объекта. Часто это означает,

ПСИХОЛОГИЯ

что данные должны быть представлены в матричном виде. Это требование заложено в большинство современных статистических пакетов.

3) Использование обучающей (дополнительной) информации при построении модели данных. Этот принцип дает более точную классификацию объектов, при сравнении изучаемого явления с уже известными явлениями. Использование такой информации значительно повышает точность статистических выводов.

4) Оптимальный подход при постановке задачи МСА. Имеется в виду рациональный выбор одного или нескольких взаимодополняющих методов, которые при применении незначительных вычислительных усилий приводят к аналитическим результатам с хорошей интерпретируемостью, которые достаточно полно и надежно представляют изучаемые явления и процессы.

Очевидно, что представленные выше принципы являются условными и в какой-то мере взаимно переплетающимися. Согласно этим принципам, мы преследуем одну цель – осмысление проблемы и способа, которым эта проблема может быть описана и решена.

Начнем с классификации методов МСА. Как любая классификация, так и эта, предлагаемая нами, является условной. Прежде всего, классические методы МСА можно разделить на методы с зависимой переменной и методы без зависимой переменной (Сошникова, 1999). Эта классификация представлена в Таблице 1.

Таблица 1. Классификация классических методов МСА.

В зависимости от наличия зависимой переменной	
методы с зависимой переменной	методы с независимой переменной
– множественная регрессия – частная корреляция – дисперсионный анализ – ковариационный анализ – дискриминантный анализ	– факторный анализ – кластерный анализ – многомерное шкалирование – анализ канонических корреляций
В зависимости от количества зависимых переменных	
одна зависимая переменная	несколько зависимых переменных
– множественная регрессия – дисперсионный анализ – ковариационный анализ – дискриминантный анализ (две группы)	– многомерная множественная регрессия – многомерный дисперсионный анализ – многомерный ковариационный анализ – множественный дискриминантный анализ

Уже упоминалось, что особый интерес представляют модели с одной или несколькими неявными (латентными) переменными (Jасq, 1998). Классификация классических методов по такому признаку представлена в таблице 2.

Таблица 2. Классификация классических методов МСА по наличию латентной переменной

Методы без латентной переменной
– множественная регрессия
– частная корреляция
– дисперсионный анализ
– ковариационный анализ
– дискриминантный анализ
– анализ связанности (path analysis)
Методы с одной латентной переменной
– одномерное шкалирование
– однофакторный анализ
Методы с несколькими латентными переменными
– факторный анализ
– многомерное шкалирование
– LISREL
Методы с группами латентных переменных
– анализ канонических корреляций
– обобщенный анализ канонических корреляций

В представленных классификациях не отражен один важный момент при выборе подходящей модели данных – способ, согласно которому эти данные измеряются. Предлагаемая классификация отражает свойства соответствующей измерительной шкалы.

Рассмотрим следующие формальные модели данных при предположении, что у нас есть одна или несколько зависимых переменных: $Y = f(X)$, где Y это зависимая переменная, X – независимая переменная, а f – связывающая функция.

Варианты различных комбинаций этих трех компонентов и определяют тип модели. Например, связывающая функция f может быть линейной или нелинейной, и это приводит к соответствующей модели. С другой стороны, если вектор Y имеет только один элемент у нас будет модель с одной зависимой переменной. Размерности X определяет и размерность модели. Различные варианты отражены в таблице 3.

Таблица 3. Классификация наиболее часто используемых моделей с зависимой переменной

Связывающая функция	Переменные				
	Y (зависимая)		X (независимая)		Модель
	Размерность	Шкала	Размерность	Шкала	Модель
Линейная	1	метрическая	1	метрическая	Простая линейная регрессия
	1	метрическая	1	неметрическая	Однофакторный одномерный ДА
	1	метрическая	>1	метрическая	Множественная линейная регрессия
	1	метрическая	> 1	неметрическая	Многофакторный ДА
	1	метрическая	> 1	смешенная	Ковариационный анализ
	>1	метрическая	1	неметрическая	Однофакторный многомерный ДА
	>1	метрическая	>1	метрическая	Каноничная корреляция
	1	неметрическая	> 1	неметрическая	Лог-линейный анализ
	1	неметрическая	>1	метрическая	Дискриминантный анализ
Нелинейная	1	метрическая	>1	метрическая	Нелинейная регрессия
	1	неметрическая	>1	смешенная	Логистическая регрессия

Обсуждение результатов

Обратим внимание, что все показанные здесь классификации являются: условными и неполными. Но, несмотря на это, они полезны, потому что дают возможность исследователю оценить свойства данных, которыми он располагает и выбрать такую модель, которая позволяет описать эти данные лучшее по предварительно выбранному критерию.

Очевидно, описание указанных здесь методов МСА требует значительного внимания. Нам же целесообразно ограничиться перечислением небольшой части источников, посвященных этим методам (Гнеденко, 2012; Дятлов, 2015, Романко, 2017; Stevens, 2002).

В XIX веке при применении методов математической статистики в психологических исследованиях в основном применялся байесовский подход. В XX веке акцент стал смещаться в сторону частотного подхода. Скорее всего XXI век будет отмечен взаимным проникновением этих двух подходов. Это хорошо видно по уже нашедших применение в психологии иерархическим и МСМС (Markov Chain Monte Carl) моделям. Также будет продолжено интенсивное применение методов анализа недостающих и неполных данных в психологических исследованиях (Schafer, 1997). Особое место занимают новые методы для анализа повторных измерений. Этот тип измерения уже становятся стандартом при сборе данных в психологии.

Современная вычислительная техника обеспечивает возможность для быстроты принятия и простоты использования графических методов для представления многомерных данных в психологии, известных как perceptual mapping (PM). На основе этих методов разработаны технологии, которые приводят к нормализации размерности. Это может быть, как факторный анализ, так и кластерный анализ, многомерное шкалирование, подход дискретного выбора, корреспондентский анализ, дискриминантный анализ и другие (Дятлов, 2018; Кляцкин, 2015; Козлов 2018; Tabachnik, 2012; Симушкин, 2009; Наследов, 2016; Куликов, 2018; Neter, 2004; Johnson, 2007). Основной целью RM является выявление скрытых структур, которые могут содержаться в сложных многомерных данных. Основной подход – это так называемый биplot. Через создание соответствующей матрицы связи между объектами и их характеристиками графически могут быть представлены взаимосвязи между этими объектами, а также сгенерированы соответствующие гипотезы. Ниже показан этот подход для типичного примера о предпочтениях к автомобилям. Разные автомашины описываются через несколько важных атрибутов и после этого группе людей, как правило, выборке из изучаемой генеральной совокупности, задают соответствующие вопросы о предпочтениях каждого из этих атрибутов. Часто, как это было сделано и в данном примере, фигурируют важные характеристики респондентов. В этом примере все переменные являются категориальными (неметрическими), а используемый метод – множественный корреспондентский анализ. Биplot, выполненный при помощи статистического пакета SAS, показан на рисунке1.

ПСИХОЛОГИЯ

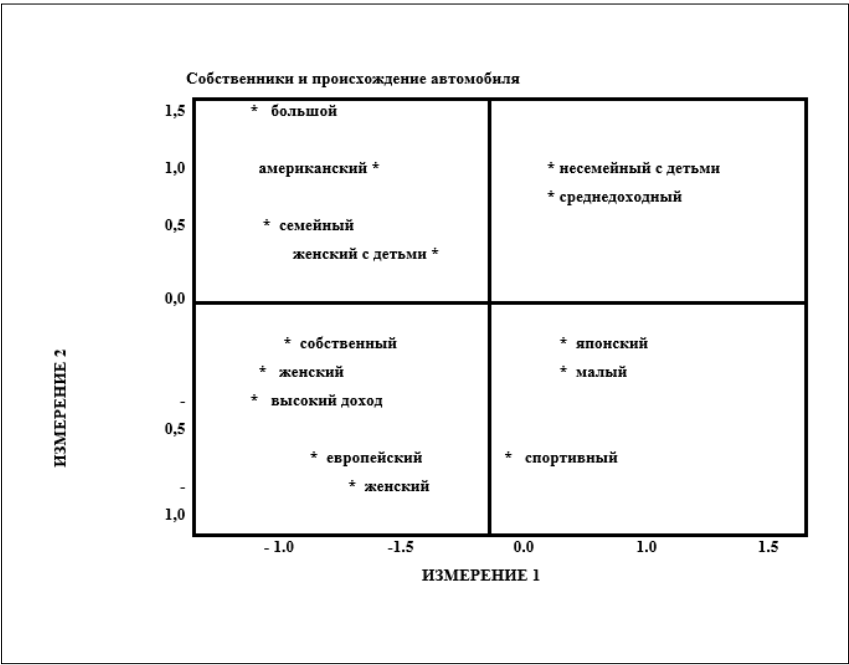


Рисунок 1. Биплот, выполненный при помощи статистического пакета SAS

Методы многомерной статистики являются неотъемлемой частью психологических исследований. Использование этих методов не является очевидным и простым, как часто используемые стандартные одномерные методы. Эффективное и актуальное применение ММА предполагает наличие предварительно известных условий:

- 1) четко сформулированная цель исследования;
- 2) четко определенная модель отношений между атрибутами (переменными);
- 3) ясность в отношении шкал, по которым измеряются переменные.

Кроме того, содержательная интерпретация результатов предполагает, по крайней мере, первичное знание математических методов.

По всей видимости, с появлением компьютеров и доступных статистических пакетов роль многомерных методов в психологических исследованиях будет расти. Наряду с этим будут возрастать и требования к подготовке психологов, которые будут намерены использовать эти методы.

Авторы заявляют об отсутствии конфликта интересов.

Литература

Гнеденко Б. В. Математические методы в теории надежности: Основные характеристики надежности и их статистический анализ. М.: КД Либроком, 2019. 584 с.

Дятлов А. В., Гугуева Д. А. Анализ данных в социологии. Ростов-на-Дону: Издательство Южного федерального университета, 2018. 205 с.

Дятлов А. В., Сажин П. В., Анисстратенко Т. Г. Поведенческие науки: применение статистических методов. Ростов-на-Дону: Фонд науки и образования, 2015. 242 с.

Кляцкин В. И. Статистический анализ когерентных явлений в стохастических динамических системах. М.: Красанд, 2015. 776 с.

Козлов А. Ю. Статистический анализ данных в MS Excel: Учебное пособие. М.: Инфра-М, 2018. 80 с.

Куликов Е. И. Прикладной статистический анализ. М.: ГЛТ, 2018. 464 с.

Наследов А. Д. IBM SPSS Statistics 20 и AMOS: профессиональный статистический анализ данных. СПб.: Питер, 2016. 416 с.

Романко В. К. Статистический анализ данных в психологии: Учебное пособие. М.: БИНОМ. ЛЗ, 2017. 312 с.

Сошникова А. А., Томашевич В. Н., Уعبة Г., Шефер М. Многомерный статистический анализ в экономике. М.: Юнити, 1999. 598 с.

Симушкин С. В. Многомерный статистический анализ. Казань: Казанский государственный университет, 2009. 114 с.

Терещенко О. В. Многомерный статистический анализ данных в социальных науках. Минск: БГУ, 2012. 239 с.

Johnson R. A., Wichern D.W. Applied multivariate analysis (6th ed.), New Jersey: Pearson Education, 2007. 769 p.

Neter J., Wasserman W., Kutner M.H. Applied linear statistical models (2nd ed.), Homewood, Illinois: McGraw-Hill Irwin, 2004. 1424 p.

Schafer J. L. Analysis of incomplete multivariate data // Monographs on statistics and applied probability. Taylor & Francis, 1997. 430 p.

Stevens J. Applied multivariate statistics for the social sciences (4rd ed.). New Jersey: LEA, 2002. 699 p.

Tabachnik B. G., Fidel L. S. Using multivariate statistics (6th ed.). Boston: Allyn and Bacon, 2012. 983 p.

Jacq J. Multivariate analysis techniques in social science research from problem to analysis. London: Sage Publications, 1998. 432 p.

References

Dyatlov, A. V., Gugueva, D. A. (2018). *Data Analysis in sociology*. Rostov-on-don: Southern Federal University Press. (in Russ.).

ПСИХОЛОГИЯ

Dyatlov, A.V., Sazhin, P. V., Anistratenko, T. G. (2015). *Behavioral Sciences: application of statistical methods*. Rostov-on-don: Foundation for science and education. (in Russ.).

Gnedenko, B. V. (2019). *Mathematical methods in the theory of reliability: the Main characteristics of reliability and their statistical analysis*. Moscow: Librokom KD. (in Russ.).

Jacq, J. (1998). *Multivariate analysis techniques in social science research from problem to analysis*. London: Sage Publications.

Johnson, R.A., Wichern, D.W. (2007). *Applied multivariate analysis (6th ed.)*. New Jersey: Pearson Education.

Klyatskin, V. I. (2015). *Statistical analysis of coherent phenomena in stochastic dynamic systems*. M.: Krasang. (in Russ.).

Kozlov, A. Yu. (2018). *Statistical data analysis in MS Excel: textbook*. Moscow: Infra-M. (in Russ.).

Kulikov, E. I. (2018). *Applied statistical analysis*. M.: GLT. (in Russ.).

Legacies, A.D. (2016). *IMB SPSS Statistics 20 and AMOS: professional statistical data analysis*. SPb.: Piter. (in Russ.).

Neter, J., Wasserman, W., Kutner, M. H. (2004). *Applied linear statistical models (2nd ed.)*. Homewood, Illinois: McGraw-Hill Irwin.

Romanko, V. K. (2017). *Statistical data analysis in psychology: textbook*. M.: BINOM. LZ. (in Russ.).

Schafer, J. L. (1997). *Analysis of incomplete multivariate data*. Taylor & Francis.

Simushkin, S. V. (2009). *Multivariate statistical analysis*. Kazan: Kazan state University. (in Russ.).

Soshnikova, L. A., Tomashevich, V. N., Uchebe G., Schaefer M. (1999). *Multivariate statistical analysis in Economics*. M.: Unity.

Stevens J. (2002). *Applied multivariate statistics for the social sciences (4rd ed.)*. New Jersey: LEA.

Tabachnik, B. G., Fidel L.S. (2012). *Using multivariate statistics (6th ed.)*. Boston: Allyn and Bacon.

Telesena, O. V. (2012). *Multivariate statistical analysis in the social sciences*. Minsk: BSU. (in Russ.).