

В.Р. Орестова
А.А. Бастрон

Применение статистического пакета анализа данных SPSS Statistics в психологических исследованиях на примере факторного анализа

В статье рассматривается применение статистического пакета анализа данных SPSS Statistics для решения задач психологического исследования. Целью исследования было выявить взаимосвязь между устойчивостью личности к внешнему информационному воздействию и такими психологическими качествами, как жизнестойкость, субъектность, толерантность к неопределенности, удовлетворенность жизнью и удовлетворенность самореализацией. Для решения этой задачи используется факторный анализ. Логика применения этого метода представлена в статье.

Ключевые слова: статистический пакет анализа данных, SPSS Statistics, психологическое исследование, факторный анализ, устойчивость личности к внешнему информационному воздействию, жизнестойкость, субъектность, толерантность к неопределенности, удовлетворенность жизнью, удовлетворенность самореализацией.

Математическая обработка данных является необходимым этапом для причинного объяснения феноменов любой природы, в том числе психологических феноменов.

Существует определенная последовательность действий для получения значимой информации: выборка, предобработка, трансформация, применение метода для получения знаний, интерпретация.

Выборка – подготовка исходного набора данных и формирование некоторого объема, необходимого для дальнейшего применения методов обработки данных.

Предобработка – предварительная обработка с целью выбора метода анализа. Необходимо «сырые» данные обработать,

т. е. исключить пропущенные значения, аномальные значения, ложные значения. Иногда некоторые данные необходимо дополнить. Объем выбранных данных зависит от конкретного метода, и в том или ином случае объем выбранных данных может быть избыточен или недостаточен, и тогда либо изменяется набор данных, либо меняется метод анализа.

Трансформация – нормализация, этап приведения данных к виду, удовлетворяющему методу анализа.

Интерпретация – постобработка данных. Это обработка полученных данных к виду, пригодному для конечного пользователя, разработка алгоритмов для использования полученных результатов в реальности.

Из трех групп существующих методов интеллектуального анализа данных – математическая статистика, эволюционное моделирование и машинное обучение – ограничимся рассмотрением методов математической статистики, применяемых в психологических исследованиях.

Математическая статистика лежит в основе методов, позволяющих получить оптимальное решение в условиях случайных и непредсказуемых воздействий на объект. Статистические пакеты прикладных программ – математические системы для статистической обработки любых данных. Наиболее распространенные – Statistica, SAS и SPSS.

Выбор статистического пакета прикладных программ зависит от предметной области, в которой будут проводиться исследования. Каждый статистический пакет прикладных программ имеет свои плюсы и минусы, многие из которых обусловлены спецификой системы.

Для пакета Statistica основным направлением является обработка экспериментальных данных. Пакет SAS нацелен главным образом на анализ финансов. Программа SPSS Statistics применяется при обработке психологических, социальных, маркетинговых и других исследований.

Пакет Statistica для профессионального использования является мощным многофункциональным пакетом, содержащим разные методы анализа: непараметрические, одномерные и многомерные.

Графические возможности пакета Statistica максимальны, так как пакет содержит различные виды графиков и матриц, а также функции настройки отображения графиков, позволяет одновременно работать с несколькими графиками и предоставляет большой набор функций для работы с трехмерной графикой.

В программе возможны практически все описательные статистики, такие как вычисление медианы, моды, среднего значения и

среднеквадратичного отклонения и прочие специальные функции. Программа предоставляет возможность группировать данные по множеству заложенных параметров, а также создавать иерархические списки, поддерживает разнообразные методы определения корреляции. Поддержка многих типов распределения – неоспоримый плюс пакета Statistica.

Стандартная комплектация пакета содержит модуль, включающий методы линейной и нелинейной регрессии. Данный пакет имеет множество встроенных и подключаемых модулей, что упрощает использование программы¹.

Пакет SPSS Statistics – система Statistical Package for the Social Sciences – непревзойденный лидер в области обеспечения статистических исследований в гуманитарной сфере. Модульная схема данной системы позволяет проработать все этапы исследований: планирование, сбор данных, анализ, вывод результатов.

Один из модулей – SamplePower – обеспечивает этап планирования, позволяя сократить затраты на проведение данного этапа. Множество параметров позволяют находить оптимальную выборку данных для дальнейшего исследования. Для сложной выборки существует модуль ComplexSamples. Этап сбора данных обеспечивается модулем DataEntry. Модуль содержит множество вариантов для сбора данных, в том числе посредством сети Интернет.

Все существующие виды статистического анализа данных заложены в данной системе. Наличие удобного интерфейса позволяет работать с программой независимо от уровня знаний предмета статистики.

Покажем использование статистического пакета SPSS Statistics для решения задач психологического исследования.

В нашем исследовании наибольший интерес представляет способность личности сохранять устойчивость при возникшем агрессивном влиянии информации на человека. Оптимальному поведению личности в условиях агрессивного информационного воздействия помогает ее психологическая устойчивость.

При изучении психологической устойчивости личности в условиях информационного воздействия, неопределенности или отсутствия информации Д.А. Леонтьев использует понятие личностного потенциала. «Личностный потенциал – это не отдельные непосредственно поддающиеся измерению переменные, а сложно организованная система взаимосвязанных переменных»².

Результаты первых исследований Д.А. Леонтьева, направленные на выявление структуры личностного потенциала, позволили предположить, что одной из трех дополняющих друг друга групп переменных, составляющих личностный потенциал, являются пе-

ременные, «которые связаны с сохранением устойчивости и цельности на фоне неблагоприятных или враждебных обстоятельств»².

Целью исследования было выявить взаимосвязь между устойчивостью личности к внешнему информационному воздействию и такими психологическими качествами, как жизнестойкость, субъектность, толерантность к неопределенности, удовлетворенность жизнью и удовлетворенность самореализацией. Для решения этой задачи используется факторный анализ.

Опишем процесс обработки группового тестирования студентов РГГУ направления подготовки «Прикладная информатика». Для исследования взяты 13 переменных из трех методик М. Исакова «Субъектность», тест «Жизнестойкость» Мадди в адаптации Д.А. Леонтьева, «Толерантность к неопределенности» Т.В. Корниловой, тест смысловых жизненных ориентаций (СЖО) Д.А. Леонтьева.

Для проведения анализа исследования были взяты следующие переменные: ответственность, свобода, общая рефлексия, рефлексия выбора, контроль (из «Субъектности»); вовлеченность, принятие риска (из «Жизнестойкости»); толерантность к неопределенности, интолерантность, межличностная интолерантность (из «Толерантности к неопределенности»); удовлетворенность жизнью, удовлетворенность самореализацией, управляемость жизнью (из «СЖО»).

Применение факторного анализа сводится к приведению множества переменных к меньшему количеству факторов, каждый из которых объединяет сходные по смыслу первичные переменные. По утверждению В.А. Мохова, «такая подмена двух (или нескольких) сильно связанных переменных одной и есть основная техническая задача факторного анализа»³.

Основной особенностью данного метода является возможность количественного определения некоего неизмеряемого фактора за счет других измеряемых параметров. Алгоритм факторного анализа – замены двух или несколько коррелирующих переменных одним фактором, описывающим объекты, носит название метода главных компонент (principal components). Процедура факторного анализа последовательно вводит новые независимые факторы для объяснения максимально возможного процента дисперсии³.

Основные этапы факторного анализа: вычисление корреляционной матрицы для всех исходных переменных, извлечение факторов, вращение факторов, интерпретация факторов.

Корреляционная матрица – это матрица (таблица) коэффициентов корреляции. Под коэффициентом корреляции понимается показатель вероятностной связи между переменными, измеренными количественно⁴. Значение коэффициента корреляции

исходные_данные.sav [Набор данных] - Редактор данных IBM SPSS Statistics													
Файл Правка Вид Данные Преобразовать Анализ Прямой маркетинг Графика Сервис Окно Справка													
	ответствен	свобода	общ_рефлексия_в_сия	рефлексия_в_ыб	контроль	вовлеченн	прин_риска	тол_к_неопр	инtolерант	инт_к_неопр_меж	удовл_жизн...	удовл_самос_управл_жизн	
1	100,00	66,00	68,00	60,00	85,00	86,00	90,00	57,00	60,00	80,00	90,00	90,00	90,00
2	70,00	80,00	79,00	76,00	90,00	90,00	60,00	58,00	63,00	36,00	50,00	70,00	80,00
3	98,00	67,00	88,00	89,00	90,00	80,00	90,00	76,00	81,00	50,00	90,00	90,00	90,00
4	47,00	80,00	35,00	80,00	85,00	30,00	45,00	54,00	66,00	80,00	56,00	60,00	79,00
5	74,00	80,00	70,00	83,00	85,00	74,00	87,00	70,00	66,00	80,00	76,00	68,00	86,00
6	76,00	59,00	83,00	65,00	57,00	89,00	90,00	55,00	64,00	50,00	87,00	85,00	88,00
7	96,00	80,00	81,00	67,00	82,00	86,00	89,00	60,00	66,00	78,00	87,00	75,00	69,00
8	98,00	94,00	79,00	87,00	89,00	80,00	79,00	64,00	58,00	44,00	90,00	90,00	97,00
9	98,00	89,00	77,00	88,00	90,00	87,00	87,00	63,00	72,00	51,00	78,00	75,00	80,00
10	48,00	87,00	90,00	87,00	79,00	58,00	87,00	65,00	67,00	85,00	85,00	78,00	90,00
11													
12													

Рис. 1. Таблица исходных данных

измеряется в отрезке от -1 до 1 , где крайние значения показывают линейную связь, а ноль – отсутствие связи. Определяют положительную, отрицательную и нулевую корреляцию.

Первый этап при извлечении фактора – подсчет суммарного разброса значений всех переменных. Выбор взаимодействующих переменных, где наибольшая доля общей дисперсии обусловлена взаимной корреляцией, – первая задача факторного анализа. Такие переменные образуют первый фактор. После этот фактор исключается, и из оставшихся снова выбираются переменные, у которых наибольшая доля оставшейся общей дисперсии обусловлена взаимной корреляцией, – это второй фактор. Данный алгоритм проводится до тех пор, пока вся общая дисперсия не будет израсходована. Корреляции исходных переменных с выделенными факторами называются факторными нагрузками данных переменных на данный фактор (factor loadings)³.

В факторном анализе переменная имеет значение общности – доля дисперсии переменной, определенная совокупным влиянием факторов.

Выбор факторов – это логическая операция для факторного анализа, определяющая наиболее значимые факторы из общего числа. В.А. Мохов – о вращении факторов: «Для общей объяснительной силы выделенных факторов неважно, каким образом оси факторов ориентированы в пространстве факторных нагрузок, и мы можем произвольно вращать факторы в этом пространстве, стремясь максимизировать одни нагрузки и минимизировать другие. Вращение позволяет получить “хорошую”, или “простую”, структуру факторных нагрузок, при которой нагрузки одних переменных на фактор велики, а других – малы. Вращение факторов в пространстве факторных нагрузок не изменяет ни собственных значений факторов, ни общего процента объясненной дисперсии»⁴. Цель вращения – получение простой структуры, где существует большое значение нагрузки (коэффициента зависимости переменной и фактора) по одному фактору и малые значения по другим.

И в завершение проводится интерпретация факторов – процесс преобразования полученных данных в вид, понятный пользователю.

Первым этапом при проведении исследования студентов РГТУ направления подготовки «Прикладная информатика» является создание файла исходных данных, показанного на рис. 1, где строки – данные тестирования, а столбцы – переменные: ответственность, свобода, общая рефлексия, рефлексия выбора, контроль, вовлеченность, принятие риска, толерантность к неопределенности, интолерантность, межличностная интолерантность, удовлетворенность жизнью, удовлетворенность самореализацией, управляемость жизнью.

Факторный анализ

	Среднее значение	Средне-квадратичное отклонение	Анализ N
ответствен	80.5000	20.79129	10
свобода	78.2000	11.05341	10
общ_рефлексия	75.0000	15.64893	10
рефлексия_выб	78.2000	10.69579	10
контроль	83.2000	9.93087	10
вовлеченн	71.0000	21.11345	10
прин_риска	80.4000	15.46466	10
тол_к_неопр	62.2000	6.92499	10
интолерант	66.3000	6.44722	10
инт_к_неопр_меж	63.4000	18.69759	10
удовл_жизни	78.9000	14.55602	10
удовл_самореал	78.1000	10.47165	10
управл_жизн	84.9000	7.96450	10

Рис. 2. Описательные статистики

Второй этап – вычисление корреляционной матрицы для всех переменных, участвующих в анализе. В процессе работы программы SPSS создана корреляционная матрица на основе данных файла. Предварительно созданная таблица описательных статистик приведена на рис. 2, а результат действия – на рис. 3.

Коррелирующие значения выделены светлым подчеркиванием, некоррелирующие – темным.

Третий этап – извлечение факторов. Был произведен подсчет суммарного разброса значений всех участвующих в анализе переменных. Далее – выбор взаимодействующих переменных, чья взаимная корреляция обуславливает наибольшую долю общей дисперсии. Эти переменные образуют первый фактор. Затем первый фактор исключается, и из оставшегося множества переменных снова выбираются те, чье взаимодействие определяет наибольшую долю оставшейся общей дисперсии. Эти переменные образуют второй фактор.

Далее идут 3-й, 4-й и 5-й факторы.

На рис. 4 и 5 показаны процессы извлечения факторов.

Корреляционная матрица ^{а,б}													
Корреляция	ответств ен	свобода	общ_реф лексия	рефлекс ия_выб	контроль	вовлечен н	прин_ри ска	топ_к_не опр	интерп ерант	инт_к_не опр_меж	удовл_ж изни	удовл_са морал	управл_ жизн
ответствен	1,000	-112	374	-154	251	795	580	279	098	-360	576	851	088
свобода	-112	1,000	-039	625	556	-250	-244	140	-179	-005	-193	-315	-053
общ_рефлексия	374	-039	1,000	117	-158	545	732	480	225	-359	551	603	255
рефлексия_выб	-154	625	117	1,000	500	-222	-115	665	486	-203	-092	-133	272
контроль	251	556	-158	500	1,000	-246	-261	405	197	-085	-259	-141	-063
вовлеченн	795	-250	545	-222	-246	1,000	899	278	092	-093	855	696	199
прин_риска	580	-244	732	-115	-261	899	1,000	460	207	075	869	665	263
топ_к_неопр	279	140	480	665	405	278	460	1,000	643	-071	377	279	335
интерперант	098	-179	225	486	197	092	207	643	1,000	-082	097	006	-144
инт_к_неопр_ме ж	-360	-005	-359	-203	-085	-093	075	-071	-082	1,000	137	-307	-199
удовл_жизни	576	-193	551	-092	-259	855	869	377	097	137	1,000	807	434
удовл_самореал	851	-315	603	-133	-141	696	665	279	006	-307	807	1,000	648
управл_жизн	088	-053	255	272	-063	199	263	335	-144	-199	434	648	1,000

Рис. 3. Корреляционная матрица

	Начальная	Извлечение
ответствен	1.000	.956
свобода	1.000	.828
общ_рефлексия	1.000	.630
рефлексия_выб	1.000	.940
контроль	1.000	.828
вовлеченн	1.000	.929
прин_риска	1.000	.938
тол_к_неопр	1.000	.909
интолерант	1.000	.935
инт_к_неопр_ме ж	1.000	.957
удовл_жизни	1.000	.947
удовл_самореал	1.000	.902
управл_жизн	1.000	.892

Рис. 4. Общности

Каждая переменная имеет единичное значение общности. Этот показатель равен доле дисперсии переменной, обусловленной совокупным влиянием факторов.

После того как процедура извлекает первый фактор, напротив его номера появляется его собственное значение; так, «ответственности» сопоставляется значение 0,956, которое пропорционально доле общей дисперсии, определяемой данным фактором, и является наибольшим. Факторы извлекаются в порядке убывания их влияния на дисперсию переменных.

Вычисляются процент дисперсии, обусловливаемый данным фактором и равный отношению собственного значения фактора к числу переменных, а также соответствующий кумулятивный (накопленный) процент.

Процент дисперсии, обусловленный факторами 1–5, равен соответственно 33,920; 16,415; 16,231; 12,115; 10,492.

С извлечением каждого нового фактора собственные значения уменьшаются, а суммарный процент равен 89%.

По графику собственных значений (рис. 6) можно наблюдать сохранение пяти компонентов и отсечение следующих, начиная с 6-го.

Рассмотрим факторные нагрузки до вращения.

Компонент	Начальные собственные значения			Суммы квадратов нагрузок			Суммы квадратов загрузок		
	Всего	% дисперсии	суммарный %	Всего	% дисперсии	суммарный %	Всего	% дисперсии	суммарный %
1	4.981	38.318	38.318	4.981	38.318	38.318	4.410	33.920	33.920
2	2.804	21.567	59.885	2.804	21.567	59.885	2.134	16.415	50.334
3	1.409	10.842	70.727	1.409	10.842	70.727	2.110	16.231	66.566
4	1.232	9.479	80.206	1.232	9.479	80.206	1.575	12.115	78.680
5	1.166	8.966	89.172	1.166	8.966	89.172	1.364	10.492	89.172
6	.785	6.036	95.208						
7	.356	2.737	97.945						
8	.199	1.530	99.475						
9	.068	.525	100.000						
10	.000	.000	100.000						
11	.000	.000	100.000						
12	.000	.000	100.000						
13	.000	.000	100.000						

Рис. 5. Объясненная совокупная дисперсия

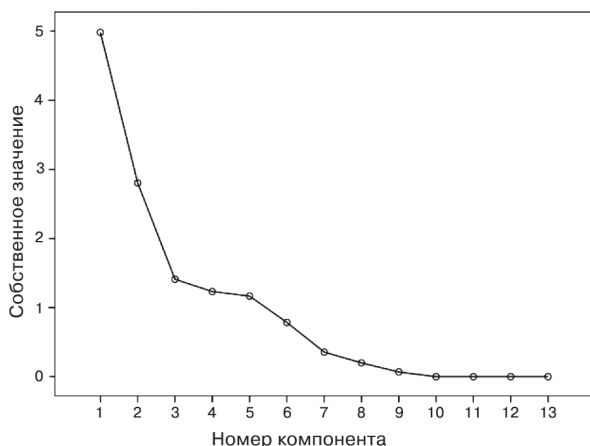


Рис. 6. График собственных значений

Наибольшие показатели имеют следующие переменные (рис. 7).

В 1-м факторе: «принятие риска» – 0,916, «удовлетворенность жизнью» – 0,902, «вовлеченность» – 0,900, «удовлетворенность самореализацией» – 0,881; во 2-м факторе: «рефлексия выбора» – 0,933; в 3-м факторе: самый высокий показатель у «интолерантности» – 0,640; в 4-м факторе: самый высокий показатель – у «управления жизнью» – 0,640; в 5-м факторе: самый высокий показатель наблюдается в «межличностной интолерантности» – 0,639.

Четвертый этап – это вращение факторов для создания упрощенной структуры. Повернутая матрица компонентов представлена на рис. 8.

В результате вращения получена более простая структура, которой соответствует большое значение нагрузки каждой переменной только по одному фактору и малое – по остальным факторам.

Нагрузка отражает связь между переменной и фактором, являясь подобием коэффициента корреляции. Значение нагрузки лежит в пределах от –1 до 1.

Выбран метод вращения «Варимакс», так как он является ортогональным и при вращении сохраняется взаимное ортогональное вращение осей (90 градусов).

Наибольший показатель имеют следующие переменные. В 1-м факторе: «вовлеченность» – 0,953 вместо 0,900 до вращения, «принятие риска» – 0,897 вместо 0,916 до вращения, «удовлетворенность

	Компонент				
	1	2	3	4	5
прин_риска	.916	-.079	.255	.034	.166
удовл_жизни	.902	-.114	.097	.168	.289
вовлеченн	.900	-.185	.076	-.217	.176
удовл_самореал	.881	-.095	-.324	.090	-.064
общ_рефлексия	.746	.175	-.023	.049	-.200
ответствен	.725	.044	-.177	-.617	.126
рефлексия_выб	-.048	.933	-.018	.256	-.029
тол_к_неопр	.498	.751	.277	.142	-.026
контроль	-.177	.731	-.148	-.395	.291
свобода	-.284	.609	-.299	-.029	.535
интолерант	.218	.536	.640	-.094	-.426
управл_жизн	.458	.154	-.498	.640	-.041
инт_к_неопр_ме ж	-.203	-.242	.589	.321	.639

Рис. 7. Матрица компонентов

	Компонент				
	1	2	3	4	5
вовлеченн	.953	.017	-.142	-.012	-.022
прин_риска	.897	.230	-.189	.131	.166
удовл_жизни	.886	.079	-.123	.301	.222
ответствен	.844	-.049	.189	-.231	-.391
удовл_самореал	.768	-.031	-.159	.451	-.287
общ_рефлексия	.597	.345	-.108	.286	-.247
интолерант	.064	.935	-.070	-.220	-.067
тол_к_неопр	.348	.792	.325	.235	.020
рефлексия_выб	-.210	.638	.601	.352	-.060
свобода	-.168	-.051	.879	.111	.106
контроль	-.060	.192	.848	-.203	-.163
управл_жизн	.223	.001	.023	.911	-.114
инт_к_неопр_ме ж	-.037	-.069	-.034	-.136	.965

Рис. 8. Повернутая матрица компонентов

жизнью» – 0,886 вместо 0,902 до вращения, «ответственность» – 0,844 (0,725 до вращения). «Ответственность» после вращения вытесняет «удовлетворенность самореализацией», имея более высокий показатель. «Удовлетворенность самореализацией» составляет 0,768 (0,881 до вращения).

Во 2-м факторе: «интолерантность» – имеет самый высокий показатель – 0,935, заменив «рефлексию выбора» (0,933 до вращения), «толерантность к неопределенности» – 0,792 (до вращения имел низкий показатель). В 3-м факторе: «свобода» – 0,879, контроль – 0,848, «интолерантность» – невысокий показатель до вращения 0,640, но наибольший из всех в 3-м факторе. В 4-м факторе: «управлению жизнью» – 0,911, после вращения показатель увеличился с 0,640. В 5-м факторе: «межличностная интолерантность» – 0,965, после вращения показатель увеличился с 0,639.

Таким образом, после вращения матрицы наблюдается увеличение показателей переменных в факторах (рис. 9).

Распределение переменных по факторам имеет следующий вид: в 1-й фактор «Жизнестойкость» вошли следующие переменные: «вовлеченность», «принятие риска», «удовлетворенность жизнью», «ответственность», «удовлетворенность самореализацией»; во 2-й фактор «Толерантность к неопределенности»: «интолерантность», «толерантность к неопределенности»; в 3-й фактор «Субъектность»: «свобода», «контроль»; в 4-й фактор «Смыслжизненные ориентации» вошла одна переменная: «управление жизнью»; в 5-й фактор – также одна переменная: «межличностная интолерантность».

Устойчивость личности к внешнему информационному воздействию обуславливает наличие перечисленных выше переменных, распределенных по факторам «Жизнестойкость», «Толерантность к неопределенности», «Субъектность», «Смыслжизненные ориентации», «Межличностная интолерантность».

Компонент	1	2	3	4	5
1	.919	.212	-.153	.266	-.128
2	-.101	.659	.711	.155	-.160
3	.030	.590	-.309	-.474	.575
4	-.244	.125	-.230	.823	.441
5	.291	-.397	.568	-.048	.658

Метод выделения факторов: метод главных компонент.

Рис. 9. Матрица преобразований компонентов

Таким образом, цель исследования – выявление взаимосвязи между устойчивостью личности к внешнему информационному воздействию и такими психологическими качествами, как жизнестойкость, субъектность, толерантность к неопределенности, удовлетворенность жизнью и удовлетворенность самореализацией, была достигнута. Для решения этой задачи использовался факторный анализ. На примере статистического пакета анализа данных SPSS Statistics была показана логика применения этого метода для решения задач психологического исследования.

Примечания

- ¹ *Боровиков В.П.* Прогнозирование в системе STATISTICA в среде WINDOWS. М.: Финансы и статистика, 2006.
- ² *Личностный потенциал: структура и диагностика / Под ред. Д.А. Леонтьева.* М.: Смысл, 2011.
- ³ *Мохов В.А.* Математические методы в психологии: Учеб. пособие для студентов и преподавателей. М., 2012.
- ⁴ *Мохов В.А.* Указ. соч.