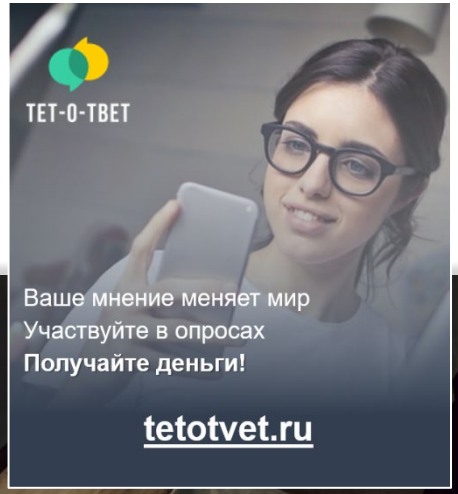
НАФИ



ТЕМА 12

МНОГОМЕРНОЕ ШКАЛИРОВАНИЕ

Для свободного использования в образовательных целях
Copyright © Академия НАФИ. Москва
Все права защищены
nafi.ru

ОГЛАВЛЕНИЕ

1. **Многомерное шкалирование: понятие и назначение**
   1.1. Метрические методы многомерного шкалирования
   1.2. Неметрические методы многомерного шкалирования

2. **Пример реализации процедуры многомерного шкалирования в SPSS**

1. МНОГОМЕРНОЕ ШКАЛИРОВАНИЕ: ПОНЯТИЕ И НАЗНАЧЕНИЕ

1. Многомерное шкалирование: понятие и назначение

Многомерное шкалирование (МШ) позволяет представить большой массив данных о различии объектов в наглядном, доступном для интерпретации графическом виде.

Методы многомерного шкалирования (МШ) применяются на практике для исследования сложных явлений и процессов, не поддающихся непосредственному описанию или моделированию.

В SPSS выводится **матрица различий** в виде одно-, двух- или трехмерного графического изображения.

- если две точки на изображении удалены друг от друга, то между соответствующими объектами имеется значительное расхождение
- если две точки на изображении расположены близко друг к другу, то между соответствующими объектами наблюдается сходство

1. Многомерное шкалирование: понятие и назначение

ИДЕЯ КЛАССИЧЕСКОГО МНОГОМЕРНОГО ШКАЛИРОВАНИЯ

Исследуется отношение совокупности людей к некоторым объектам: $a_1, a_2, \dots, a_n$.

Исходными данными для «классического» варианта МШ служат **близости между объектами.**

Пусть $S_{ij} = S(a_i, a_j)$ – это измеренная близость между объектами a_i и a_j, где

i=1, …, n; j=1,…, n. Тогда пусть исходные эмпирические данные представляют собой совокупность таких близостей, заданных в виде матрицы:

$$\|S_{ij}\|$$

Если вместо близостей будут различия, то элементы соответствующей матрицы будем обозначать через δ_{ij}, а всю матрицу через:

$$\|\delta_{ij}\|$$

- **Близости и различия** – величины обратные друг другу: чем больше близости, тем меньше различия.
- Например, если близости S_{ij} будут изменяться от 0 до 1, то в качестве различий можно взять разности $\delta_{ij} = 1 - S_{ij}$.
- **Алгоритм МШ** ищет такое числовое признаковое пространство и такое расположение рассматриваемых объектов в этом пространстве, что структура расстояний между этими объектами отвечает структуре исходной матрицы близостей.

1. Многомерное шкалирование: понятие и назначение

Традиционно в многомерном шкалировании применяются понятия, которые непривычны или вовсе не встречаются в разделах классической статистики:

- **Стимул** – некоторое свойство объекта (количественное или качественное), его признак, непосредственно не измеряемая характеристика изучаемого объекта (образ объекта).

- **Шкала** – ось теоретического пространства, которая является носителем значений обобщённого признака (фактора). Для комплексного восприятия сложной реальности требуется многомерное пространство шкал.

- **Предпочтение** – суждение об объекте с точки зрения его близости представляемому идеалу.

1. Многомерное шкалирование: понятие и назначение

- **Функция стресса** – критерий различий матрицы близостей и матрицы расстояний; изменяется от 0 до 1, когда при 0 – полное сходство, а ближе к 1 – сходства становятся меньше.

- **Функция расстояния** – двуместная функция $d(a_1, a_2)$, в качестве аргументов которой служат объекты совокупности $a_1, a_2, \dots, a_n$, и для всех i, j, k (=1, …, n) удовлетворяется следующая система аксиом:

$$\begin{cases} d\left(a_i, a_j\right) = 0, \text{если объекты } a_i \text{ и } a_j \text{ совпадают (рефлексивность расстояний)} \\ d\left(a_i, a_j\right) = d\left(a_j, a_i\right) \text{ (симметричность расстояния)} \\ d\left(a_i, a_j\right) + d\left(a_j, a_k\right) \geq d\left(a_i, a_k\right) \text{ (правило треугольника)} \end{cases}$$

1. Многомерное шкалирование: понятие и назначение

- **Квадратная матрица** – матрица, у которой число строк равно числу столбцов. Строки и столбцы матрицы представляют собой один и тот же набор объектов. Если строки и столбцы матрицы представляют собой разные объекты, тогда матрица **прямоугольная**.

- **Симметричная матрица** – матрица, у которой значения над главной диагональю равны значениям под главной диагональю.

- **Ассиметричная матрица** – матрица, для которой отношение двух объектов друг к другу может быть разным.

- **Модель индивидуальных различий** – применяется в случаях, когда возможно координатное представление нескольких матриц. Например, когда несколько экспертов оценивают один и тот же набор объектов.

Основная идея метода состоит в приписывании каждому объекту значений координат, так чтобы матрица евклидовых расстояний между объектами в этих координатах оказалась близка к матрице расстояний между объектами, определенной из каких-либо соображений ранее.

1. Многомерное шкалирование: понятие и назначение

СХОДСТВО С ФАКТОРНЫМ АНАЛИЗОМ

	МШ	ФА
Сходства	Создание системы координат пространства, в котором определяется расположение точек	
	Снижение размерности и упрощение данных	
Различия	Использование **мер различия** между объектами	Использование **коэффициентов корреляции**
	Интересны – **расстояние** между точками, представляющие данные	Интересны – **углы** между точками, представляющие данные

1. Многомерное шкалирование: понятие и назначение

СХОДСТВО С КЛАСТЕРНЫМ АНАЛИЗОМ

	МШ	КА
Сходство	Анализ расстояния между объектами	
Различие	Проведение качественного анализа визуально посредством **диаграммы**	Проведение количественной процедуры **объединения объектов в группы** (кластеры)

1. Многомерное шкалирование: понятие и назначение

Обобщая, в ходе анализа **методами многомерного шкалирования** исследователь решает следующие вопросы:

1. Поиск и интерпретация латентных переменных общих факторов
2. Сжатие исходного массива данных
3. Визуализация геометрической конфигурации наблюдаемых объектов в координатном пространстве латентных признаков

В многомерном шкалировании выделяются два направления: **метрическое и неметрическое многомерное шкалирование.**

Как это используется в НАФИ

Многомерное шкалирование помогает строить карты восприятия брендов, продуктов или сервисов. На такой карте видно, какие объекты воспринимаются аудиторией как похожие, какие конкурируют между собой, а какие занимают уникальную позицию. В исследованиях здоровья бренда это помогает перейти от отдельных имиджевых показателей к целостной картине конкурентного поля.

1.1 МЕТРИЧЕСКИЕ МЕТОДЫ МНОГОМЕРНОГО ШКАЛИРОВАНИЯ

1.1. Метрические методы многомерного шкалирования

Модель метрического многомерного шкалирования имеет вид:

$$L\{S\} = D^2 + E$$

L {S} — линейное преобразование исходной матрицы расстояний

D^2 — матрица квадратов расстояний, полученная на основе созданных шкал

E — матрица отклонений модели от исходных данных.

Линейное преобразование дает матрицу преобразованных расстояний:

$$T = L\{S\}$$

Цель многомерного метрического шкалирования — поиск оптимальных шкал с помощью линейного преобразования матрицы исходных расстояний, минимизирующих ошибку *E*.

Методы метрического многомерного шкалирования применяют для обработки количественных данных.

1.1. Метрические методы многомерного шкалирования

Принципиальным для многомерного шкалирования является тип шкал данных, поскольку функция стресса чувствительна к ним.

- Элементы матрицы различий (близости) для метрического метода многомерного шкалирования (ММШ) получены по **интервальным шкалам.**

В связи с этим необходимо рассмотреть функцию стресса как в ММШ, так и в НМШ.

1.1. Метрические методы многомерного шкалирования

Если элементы матрицы различий δ_{ij} можно считать числами, полученными по интервальной шкале, то можно рассмотреть разницу δ_{ij} и d_{ij}.

Функцию стресса S построим в соответствии с принципом наименьших квадратов:

$$S = \sum_{1 \le i,j \le n} \left(\mathrm{d}_{ij} - \delta_{ij}\right)^2$$

Чем ближе каждое расстояние d_{ij} к различию δ_{ij}, тем меньше значение этой функции

Точное решение задач МШ соответствует случаю, когда *S=0*, то есть когда для всех I, j (I, j =1, … , n) $\delta_{ij} = \mathrm{d}_{ij}$.

1.2 НЕМЕТРИЧЕСКИЕ МЕТОДЫ МНОГОМЕРНОГО ШКАЛИРОВАНИЯ

1.2. Неметрические методы многомерного шкалирования

Модель неметрического многомерного шкалирования имеет вид:

$$M\{S\} = D^2 + E$$

M{S} — монотонное преобразование исходной матрицы расстояний
D^2 — матрица квадратов расстояний, полученная на основе созданных шкал
E — матрица отклонений модели от исходных данных.

Монотонное преобразование дает матрицу преобразованных расстояний:

$$T = M\{S\}$$

Цель многомерного неметрического шкалирования — нелинейное преобразование расстояний.
Методы неметрического многомерного шкалирования применяют для обработки ранговых (порядковых) данных.

1.2. Неметрические методы многомерного шкалирования

Принципиальным для многомерного шкалирования является тип шкал данных, поскольку функция стресса чувствительна к ним.

- Элементы матрицы различий (близости) для неметрического метода многомерного шкалирования (НМШ) получены по **порядковым шкалам.**

- Неметрический метод важен, потому что данные, полученные от респондентов, легче и чаще всего даются в порядковой, а не интервальной шкале

1.2. Неметрические методы многомерного шкалирования

При НМШ элементы матрицы различий δ_{ij} и S_{ij} получены по порядковой шкале. **Например**: респонденты оценили близость следующим образом: a_1 и a_6 близки друг к другу, а a_{15} и a_2 стоят дальше от первых упомянутых.

В данном случае, не получится применить функцию стресса как у ММШ

$$S = \sum_{1 \le i,j \le n} \left(d_{ij} - \delta_{ij}\right)^2$$

В подобной ситуации координаты рассматриваемых объектов ищутся так, чтобы упорядочение расстояний d_{ij} максимально соответствовало упорядочению, обратному упорядочению близостей S_{ij}.

1.2. Неметрические методы многомерного шкалирования

Требования к критерию

- должен измерять степень упорядоченности d_{ij} и последовательности, обратной S_{ij}
- должен быть минимален при совпадении названных упорядочений
- не должен изменять своего значения при замене последовательности чисел последовательностью с тем же порядком элементов

Применяется коэффициента Краскала

$$S = \sqrt{\frac{\sum_{1 \le i,j \le n} \left(d_{ij} - \hat{d}_{ij}\right)^2}{\sum_{1 \le i,j \le n} \left(d_{ij} - \bar{d}\right)^2}}, \text{ где}$$

$\bar{d}$ - среднее значение всех расстояний d_{ij}

$\hat{d}_{ij}$ - член последовательности чисел, называемой «монотонной регрессией членов последовательности $\{d_{ij}\}$»

Числа $\hat{d}_{ij}$ совпадают по упорядочению с числами S_{ij} и при этом минимально в смысле суммы квадратов отклонений отличаются от чисел d_{ij}.

1.2. Неметрические методы многомерного шкалирования

Также можно применить другой критерий:

Коэффициент отчуждения

$$k = \sqrt{1 - \mu^2}, \text{ где}$$

μ - это коэффициент монотонности, который равен:

$$\mu = \frac{\sum_{1 \le i,j \le n} d_{ij} \cdot \hat{d}_{ij}}{\sqrt{\sum_{1 \le i,j \le n} d_{ij}^2 \cdot \sum_{1 \le i,j \le n} \hat{d}_{ij}^2}}$$

Максимизация μ равносильна минимизации S. Чем лучше неметрическая модель подогнана к исходным данным, тем меньше коэффициент отчуждения *k*.

АЛГОРИТМ НЕМЕТРИЧЕСКОГО МНОГОМЕРНОГО ШКАЛИРОВАНИЯ

1. **Создание матрицы различий, содержащей ранговые данные - характеристики непохожести анализируемых объектов**

   Существуют различные приемы получения исходных ранговых данных, наиболее распространены в анализе из них следующие:

   - **метод последовательной рандомизации** - его сущность в последовательно проводимом делении совокупности наблюдаемых объектов на группы. При первом делении появляются две группы — пары похожих объектов и пары непохожих объектов. Затем в каждой группе соответственно находят пары с наиболее и менее похожими объектами и т. д.

   - **метод исходной (якорной) точки** - из общего числа n-объектов на первом шаге отбирают один и его положение в совокупности принимается за исходное относительно других объектов. Степень сходства всех прочих объектов с первым (якорным) оценивается экспертами с присвоением ранга. На следующем шаге якорным становится другой, следующий из совокупности объект. И так для всех объектов.

   - **метод рейтинговой оценки** – экспертам предлагается шкала с некоторым числом делений (обычно 7—9), позволяющих оценивать каждую пару объектов по степени их сходства.

1.2. Неметрические методы многомерного шкалирования

2. **Поиск стартовой конфигурации**
   - решается формальная задача построения координатного пространства и размещения в нем точек-объектов таким образом, чтобы расстояния между ними, определяемые по введенной метрике, соответствовали исходным различиям
   - используется аппарат линейной или нелинейной оптимизации.

3. **Стандартизация оценок координат и расстояний**
   - стандартизация проводится с целью сохранения пропорций ортонормированного стимульного пространства и избегания вырожденных решений, когда пространство стимулов сжимается до размеров точки, и анализ не дает сколько-нибудь значимых результатов.
   - допускаются различные варианты стандартизации.

4. **Неметрический этап**
   - в теоретическом пространстве шкал монотонность исходных данных может быть нарушена. Корректировка теоретических величин расстояний производится при неизменных оценках координат стимулов и таким образом, чтобы восстановить общую тенденцию к возрастанию в исходных данных о различиях.

5. **Метрический этап**
   - на данном этапе имеющимся исходным и уточненным величинам расстояний находят уточненные оценки координат.

1.2. Неметрические методы многомерного шкалирования

6. **Качество подгонки модели – оценка соответствий монотонных ранговых эмпирических и теоретических данных**

- проверке на монотонность подлежат теоретические данные и рассматривается степень их улучшения на прошлой итерации. Если улучшение существенно, итерация возобновляется после стандартизации полученных на шаге 5 оценок координат и расстояний. Если же улучшение мало, итерации заканчиваются, и приступают к интерпретации итогов анализа.

- для измерения качества модели используется **S-stress показатель Юнга:**

$$S\text{-}stress = \left(\frac{\|E\|}{\|T\|} \right)^{1/2}$$

- показатель изменяется от 0 до 1. Равенство нулю означает точную подгонку модели, единице - полную ее бессмысленность.

- кроме того, оценить качество модели можно с помощью показателя **Stress-index Краскала**, который получается с использованием матрицы не квадратов расстояний, а расстояний.

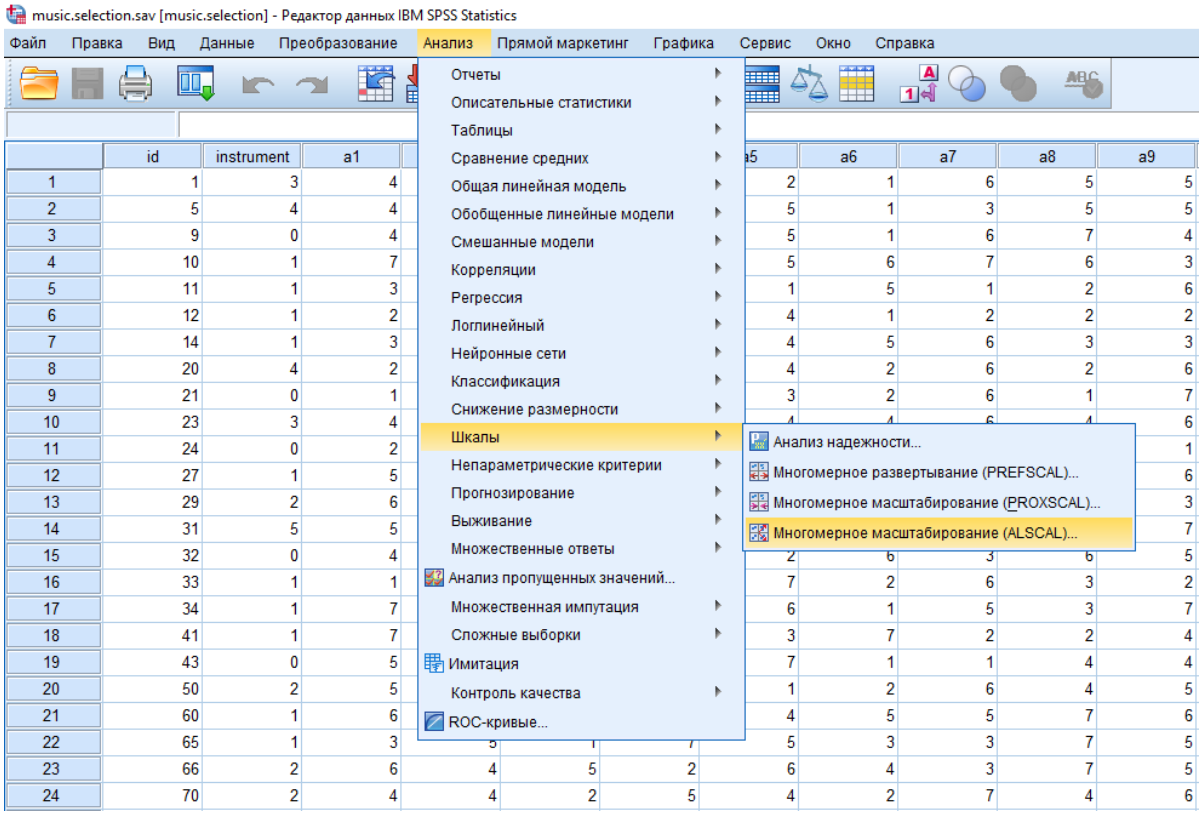
- еще один показатель качества модели – **RSQ** – представляет квадрат коэффициента корреляции между матрицами T и D.

2. ПРИМЕР РЕАЛИЗАЦИИ ПРОЦЕДУРЫ МНОГОМЕРНОГО ШКАЛИРОВАНИЯ В SPSS

2. Пример реализации процедуры многомерного шкалирования в SPSS

Процедура многомерного шкалирования включает:

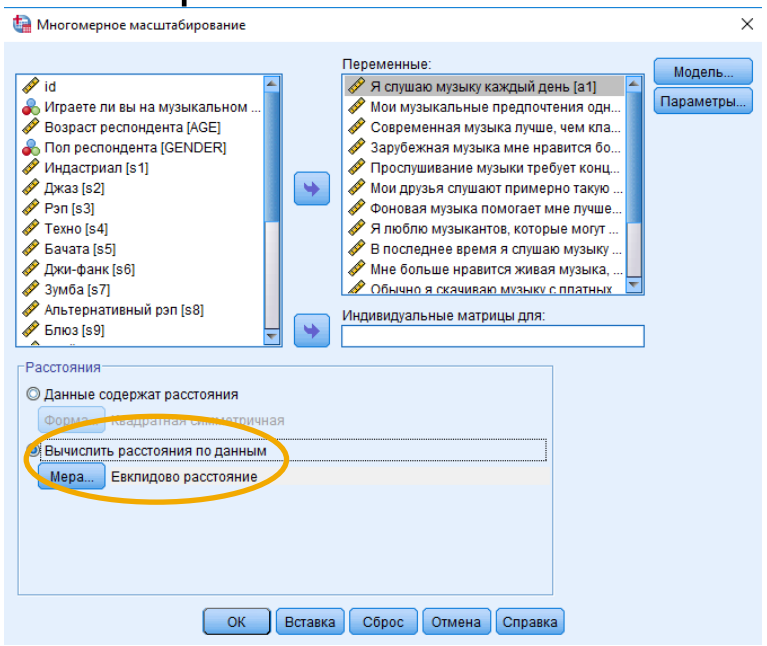
1. Меню **«Анализ»** → **«Шкалы»** → **«Многомерное масштабирование» (ALSCAL)**

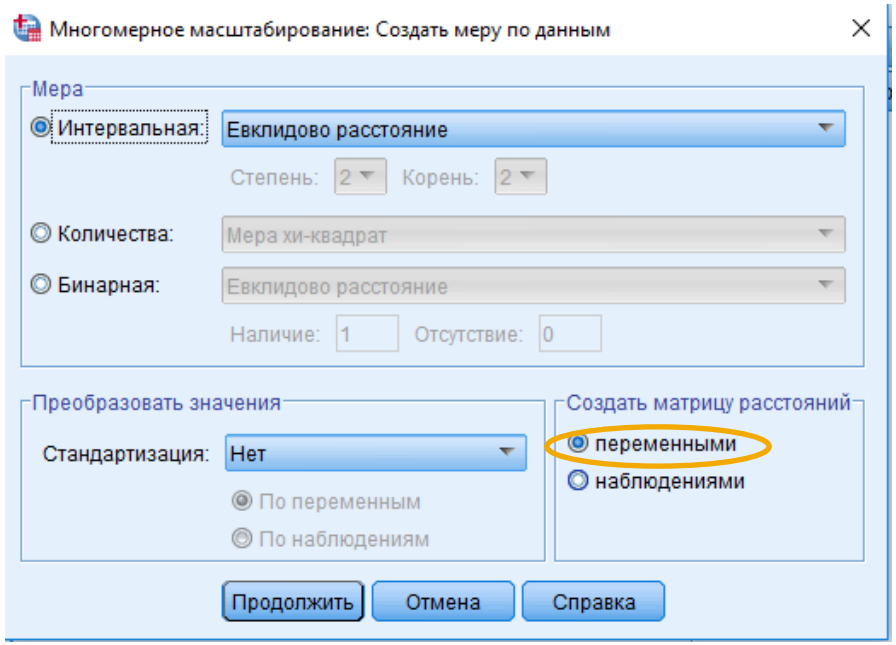


2. Пример реализации процедуры многомерного шкалирования в SPSS

Процедура многомерного шкалирования включает:

2. Переменные a1 – a15 переместите в поле **«Переменные»**
3. Так как данные не содержат матрицу расстояний, то ее необходимо вычислить. Выбираем поле **«Вычислить расстояния по данным»** и создаем матрицу расстояний по переменным

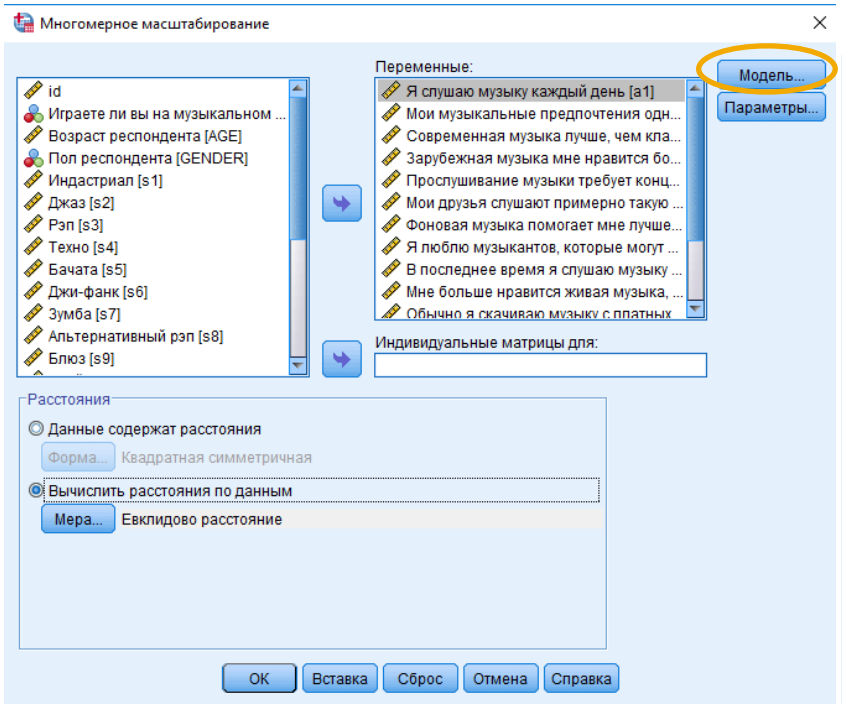


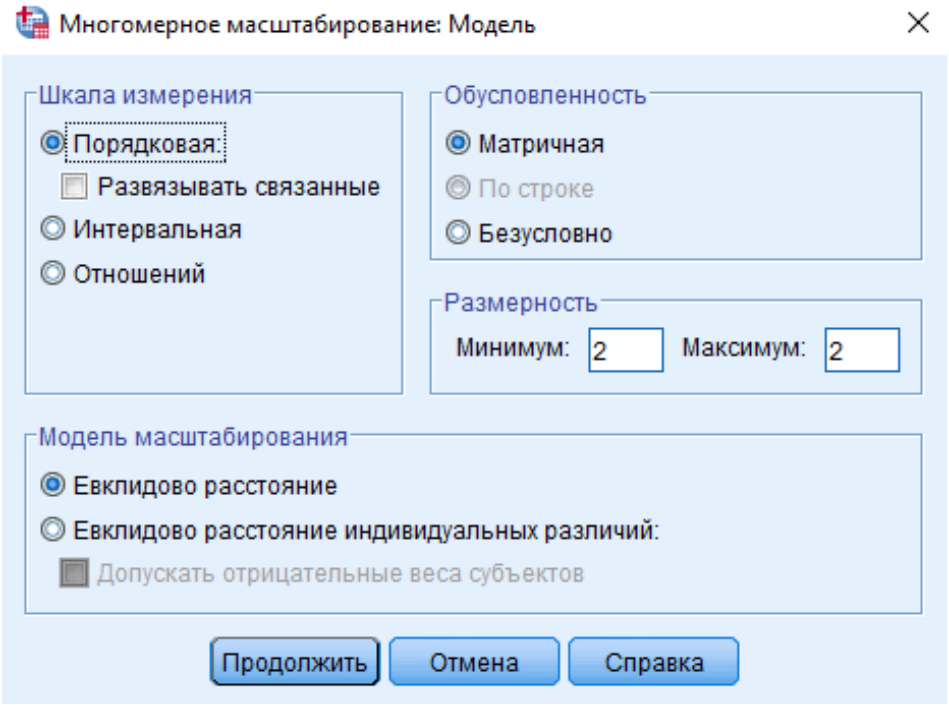


2. Пример реализации процедуры многомерного шкалирования в SPSS

Процедура многомерного шкалирования включает:

4. Открываем вкладку **«Модель»** и выбираем необходимую шкалу измерения, модель масштабирования, обусловленность.

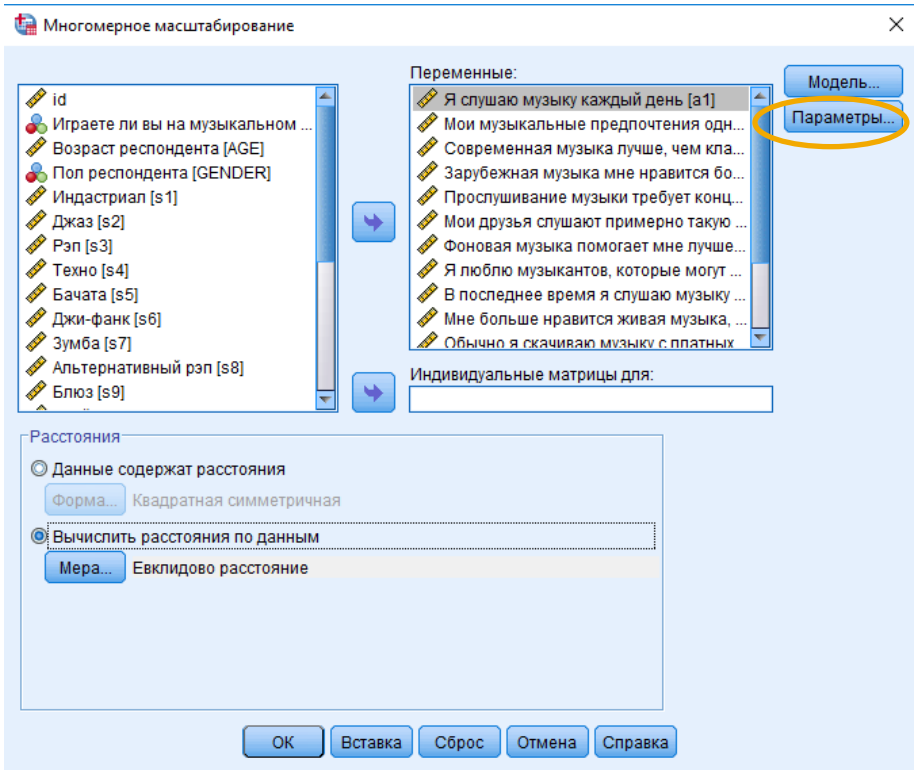


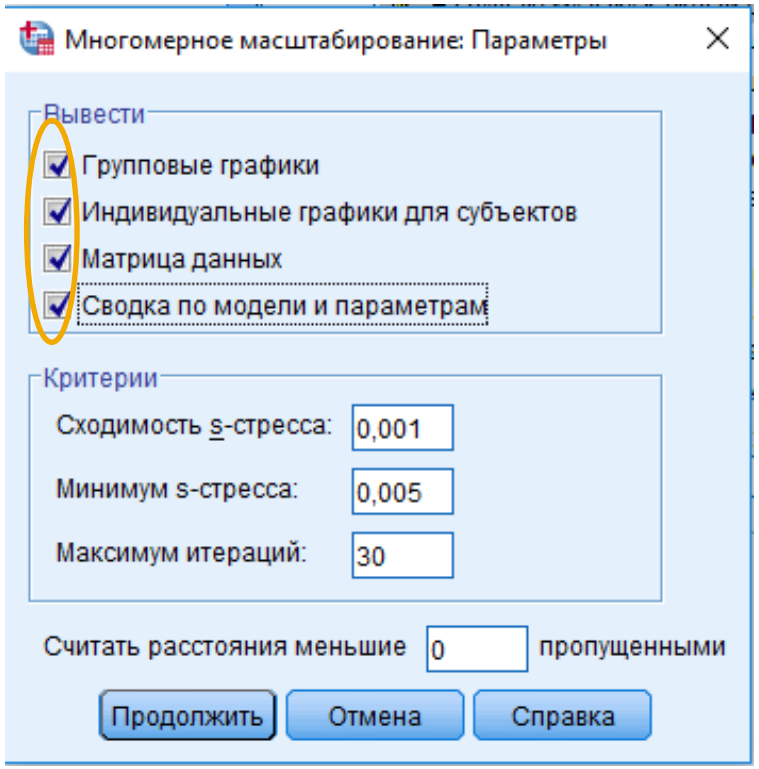


2. Пример реализации процедуры многомерного шкалирования в SPSS

Процедура многомерного шкалирования включает:

5. Открываем вкладку **«Параметры»** и выводим все элементы для анализа.





Представление результатов

1. **«Raw (unscaled) Data for Subject 1»** - это базовые Евклидовы расстояния между объектами

Raw (unscaled) Data for Subject 1

	1	2	3	4	5	6	7	8	9	10
1	,000									
2	24,839	,000								
3	25,417	22,561	,000							
4	25,199	25,768	21,886	,000						
5	24,799	25,259	25,981	24,980	,000					
6	24,860	24,062	26,683	27,000	26,363	,000				
7	26,476	25,338	29,275	28,948	25,884	28,018	,000			
8	25,239	26,192	26,851	25,573	25,456	26,552	27,423	,000		
9	25,495	25,904	26,268	24,021	27,074	25,456	25,397	24,759	,000	
10	23,854	23,875	27,148	24,042	23,707	26,439	27,604	24,860	23,558	,000
11	25,942	24,860	26,058	25,768	24,000	27,767	24,413	26,907	24,434	24,042
12	25,865	24,249	25,159	23,917	25,179	29,240	26,646	26,833	25,357	23,791
13	27,055	22,782	24,900	27,477	28,018	27,166	28,125	30,116	25,456	26,173
14	25,140	23,431	24,207	27,695	26,096	24,331	27,404	27,000	27,019	27,074
15	24,249	23,173	27,604	25,515	24,269	27,495	27,839	27,404	26,683	23,601

	11	12	13	14	15
11	,000				
12	22,716	,000			
13	23,812	26,249	,000		
14	23,937	27,258	23,622	,000	
15	24,434	26,552	24,779	26,833	,000

Представление результатов

2. **«Iteration history for the 2 dimensional solution (in squared distances)»** - это расчёт характеристик при итерациях

Значения S-stress характеризуют отклонение результата от идеального значения, точно соответствующего матрице различий, при различных итерациях. Если число итераций больше 30, то существуют проблемы с исходными данными.

Iteration history for the 2 dimensional solution (in squared distances)

Young's S-stress formula 1 is used.

Iteration	S-stress	Improvement
1	,43717	
2	,38042	,05674
3	,37043	,01000
4	,36332	,00711
5	,35846	,00485
6	,35605	,00241
7	,35482	,00124
8	,35391	,00091

Iterations stopped because
S-stress improvement is less than ,001000

Представление результатов

3. **«Stress and squared correlation (RSQ) in distances»** - это стресс и коэффициент R^2

Чем выше коэффициент R^2, тем лучше модель.

```
                 For  matrix
Stress  =   ,25971      RSQ =  ,57120
```

Представление результатов

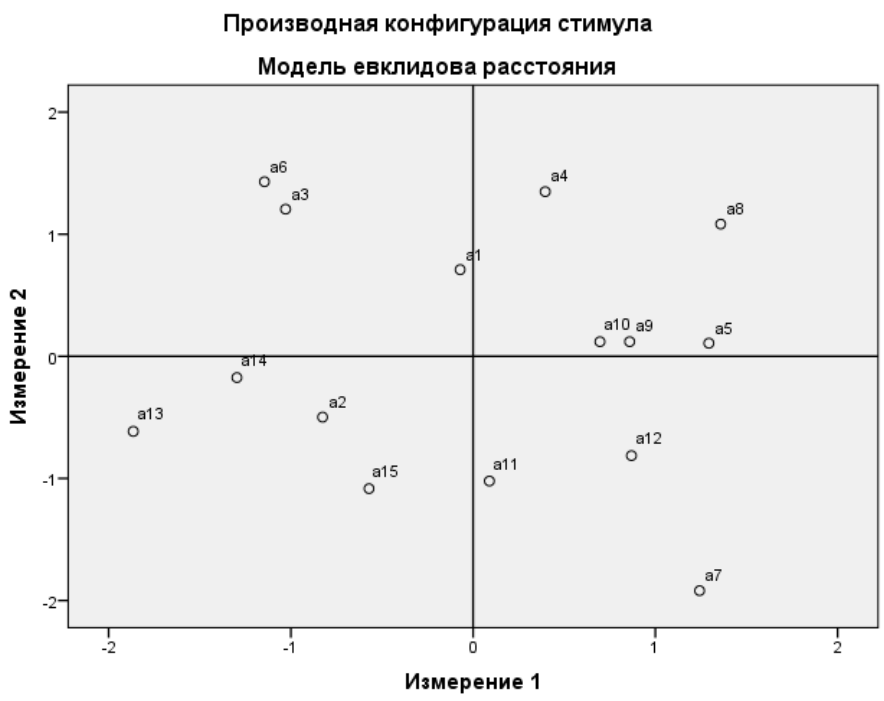
4. **«Stimulus Coordinates »** - это координаты стимулов по каждой шкале X и Y.

Stimulus Coordinates

Stimulus Number	Stimulus Name	Dimension 1	Dimension 2
1	a1	-,0712	,7099
2	a2	-,8258	-,4980
3	a3	-1,0289	1,2054
4	a4	,3956	1,3489
5	a5	1,2937	,1073
6	a6	-1,1449	1,4303
7	a7	1,2424	-1,9198
8	a8	1,3585	1,0831
9	a9	,8587	,1202
10	a10	,6960	,1200
11	a11	,0894	-1,0214
12	a12	,8687	-,8134
13	a13	-1,8649	-,6145
14	a14	-1,2960	-,1743
15	a15	-,5712	-1,0835

Представление результатов

5. **«Производная конфигурация стимула»** - это диаграмма, на которой изображена итоговая конфигурация сравниваемых респондентами способов прослушивания музыки.



Таким образом, можно выделить четыре смысловые группы утверждений о прослушивании музыки:

1. Слушают современную музыку (a3, a6)
2. Интересуются жизнью исполнителей, скачивают музыку с платных ресурсов и предпочитают фоновую музыку для работы/учебы (a12, a11, a7)
3. Слушают музыку одного стиля, стараются соответствовать в одежде и поведении стилю предпочитаемой музыки, любят ходить в клубы, на концерты и хорошо танцуют (a2, a13, a14, a15)
4. Слушают музыку каждый день, больше предпочитают зарубежные композиции, концентрируются, любят исполнение без фонограммы и живую музыку (a1, a4, a5, a8, a9, a10)

Литература по Теме 12

1. **Толстова Ю.Н. Основы многомерного шкалирования: учебное пособие. – М.: КДУ, 2006**

2. **Наследов А. IBM SPSS Statistics 20 и AMOS: профессиональный статистический анализ данных. – СПб., 2013**
   - Глава 23. Многомерное шкалирование

3. **Крыштановский А. О. Анализ социологических данных с помощью пакета SPSS. – М.: Изд. дом ГУ ВШЭ, 2006**
   - Глава 5. Исследование структуры данных

4. **Большаков А. А., Каримов Р. Н. Методы обработки многомерных данных и временных рядов. – М.: Горячая линия – Телеком, 2007**
   - Глава 9. Анализ соответствий и многомерное шкалирование

Чек-лист аналитика

Перед интерпретацией карты многомерного шкалирования проверьте:

- Понятно, какие объекты сравниваются: бренды, продукты, сервисы, работодатели, утверждения или другие стимулы.
- Исходные данные подходят для МШ: есть матрица сходств / различий или корректно рассчитанные расстояния по исходным переменным.
- Выбран подходящий тип модели: метрическое МШ - для количественных различий, неметрическое МШ - для ранговых / порядковых данных.
- Задано обоснованное число измерений: чаще начинают с двухмерного решения, но проверяют, не ухудшает ли оно качество модели.
- Проверено качество подгонки: Stress достаточно низкий, RSQ достаточно высокий, динамика итераций не указывает на проблемы сходимости.
- Карта интерпретируема содержательно: близкие объекты действительно похожи по смыслу, а удаленные - различаются.
- Оси названы: они не заданы заранее и требуют содержательной интерпретации на основе расположения объектов.
- Визуально близкие группы не выданы за формальные кластеры: для строгого выделения групп нужен отдельный кластерный анализ.
- Вывод не сводится к описанию картинки: объяснено, что карта означает для позиционирования, конкурентного поля или восприятия аудитории.
- Результат переведен на язык решения: какие объекты конкурируют между собой, кто занимает уникальную позицию, где есть риск смешения образов и какие направления позиционирования можно усилить.

НАФИ

О том, как на практике применяется многомерное шкалирование, вы можете увидеть, изучив, например, страницу решения «Исследование показателей «здоровья» бренда (ВНТ)» на нашем сайте.

Хотите узнать, как проводятся онлайн-опросы, и попробовать себя в роли респондента? Регистрируйтесь на нашей платформе https://tetotvet.ru или скачивайте наше приложение «Тет-о-твет» в любом из магазинов приложений.

ДОСТУПНО В Google Play

ЗАГРУЗИТЕ В App Store

ОТКРОЙТЕ В AppGallery

Для свободного использования в образовательных целях
Copyright © Академия НАФИ. Москва
Все права защищены
nafi.ru