

ЗАНИМАТЕЛЬНАЯ СТАТИСТИКА

МАНГА

ФАКТОРНЫЙ АНАЛИЗ

Син Такахаси
Иноуэ Ироха



OHM
Ohmsha

ОДГКА

DMK
ИЗДАТЕЛЬСТВО

ОБРАЗОВАТЕЛЬНАЯ МАНГА

ЗАНИМАТЕЛЬНАЯ СТАТИСТИКА



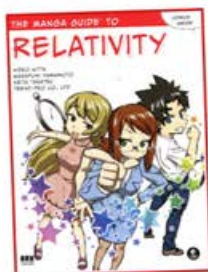
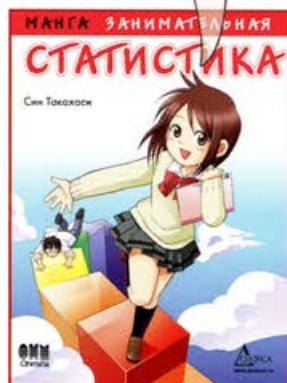
ПРОСТОЙ,
ПОСЛЕДОВАТЕЛЬНЫЙ,
НАГЛЯДНЫЙ И
НЕОБРЕМЕНИТЕЛЬНЫЙ ПУТЬ
ИЗУЧЕНИЯ —
ШКОЛЬНОЙ И
ИНСТИТУТСКОЙ
ПРОГРАММЫ!

ЯПОНЦЫ
ТАКЖЕ УМНЫЕ,
ПОТОМУ ЧТО
УЧАТСЯ
ПО МАНГЕ!



ЧИТАЙ КОМИКСЫ И
СТАНОВИСЬ ОТЛИЧНИКОМ!

ВСЕ КНИГИ ПРОВЕРЕНЫ
ПРЕПОДАВАТЕЛЯМИ
И ПРОФЕССОРАМИ
ВЕДУЩИХ ПРОФИЛЬНЫХ
ВУЗОВ
ДЛЯ ТОЧНОГО ПЕРЕВОДА
И АДАПТАЦИИ.



КНИГИ СЕРИИ "ОБРАЗОВАТЕЛЬНАЯ МАНГА"
МОЖНО ЗАКАЗАТЬ НА САЙТЕ ИЗДАТЕЛЬСТВА
WWW.DMKPRESS.COM ИЛИ WWW.DMK.RF

Нет кризису! Весь 2015 год при заказе на сайте
цена 349 руб + доставка

Занимательная статистика

ФАКТОРНЫЙ АНАЛИЗ

Манга

マンガでわかる

統計学

[因子分析編]

高橋 信 / 著
井上 いろは / 作画
トレンド・プロ / 制作



ОБРАЗОВАТЕЛЬНАЯ МАНГА

ЗАНИМАТЕЛЬНАЯ СТАТИСТИКА

ФАКТОРНЫЙ АНАЛИЗ

Син Такахаси
Иноуэ Ироха

Перевод
с японского
Клионского А.Б.



ДМК
ПРЕСС
ИЗДАТЕЛЬСТВО

ОДЭКА



Москва
Додэка, ДМК Пресс, 2015

УДК 519.233.5

ББК 22.172

T15

Такахаси, Син.

T15 Занимательная статистика. Факторный анализ. Манга. / Син Такахаси (автор), Иноуэ Ироха (худож.); пер. с яп. Клионского А. Б. — М. : ДМК Пресс, 2015. — 256 с. : ил. — (Серия «Образовательная манга»). — Доп. тит. л. яп. — ISBN 978-5-97060-270-6.

Руи влюблена в Ямамото, работающего в отделе, начальником которого является отец Руи. Но внезапно Ямамото улетает в длительную командировку. Узнав, что его работа в командировке связана с факторным анализом, Руи решает освоить эту науку, чтобы полететь к Ямамото и вместе с ним делать его работу.

А помогут ей в этом нелёгком деле её подруги, с которыми она подрабатывает в кафе Norns - аспирантка Риса и студентка Миу. Вместе с ними она узнает, для чего проводят факторный анализ и на конкретном примере освоит методы, применяемые при его проведении. Кроме того она ознакомится с методом главных компонент, который очень похож на факторный анализ, и узнает, в чём его отличие.

Если вас интересует, как можно узнать причины (или факторы), побуждающие людей отвечать на вопросы анкет так или иначе, то эта книга для вас.

УДК 519.233.5

ББК 22.172

Original Japanese edition
Manga de Wakaru Toukeigaku — Kaiki Bunseki-hen (Manga Guide: Statistics-Regression Analysis)
By Shin Takahashi (Author), Iroha Inoue (Illustrator) and
Trend-Pro Co., Ltd. (Producer)
Published by Ohmsha, Ltd.
3-1 Kanda Nishikicho, Chiyodaku, Tokyo, Japan
Russian language edition copyright © 2015 by DMK Press
Translation rights arranged with Ohmsha, Ltd.

Все права защищены. Никакая часть этого издания не может быть воспроизведена в любой форме или любыми средствами, электронными или механическими, включая фотографирование, ксерокопирование или иные средства копирования или сохранения информации, без письменного разрешения издательства.

ISBN 978-4-274-06662-2 (яп.) Copyright © 2006 by Shin Takahashi and Trend-Pro Co., Ltd.

ISBN 978-5-94120-265-2 (Додэка)

© Перевод, Издательский дом «Додэка-XXI», 2015

ISBN 978-5-97060-270-6 (ДМК Пресс)

© Оформление, издание, ДМК Пресс, 2015

ПРЕДИСЛОВИЕ

Данная книга посвящена в основном объяснению *факторного анализа* и *метода главных компонент*.

Для лучшего понимания материала желательно, чтобы читатели уже ознакомились с предыдущими книгами данной серии — «Занимательная статистика. Манга»*, или же их знания в этой области не уступали разъяснённым в этих книгах.

Данная книга имеет следующую структуру:

- Глава 1. Базовые знания об анкетировании;
- Глава 2. Бланки анкет и вопросы;
- Глава 3. Базовые математические знания;
- Глава 4. Метод главных компонент;
- Глава 5. Факторный анализ.

Каждая из глав, в свою очередь, состоит из собственно манги и текстового дополнения. Также в книге имеется приложение, знакомящее с несколькими общеизвестными методами анализа, включая рассмотренные в книгах «Манга о статистике» и «Манга о статистике. Регрессионный анализ».

Читателю, сведущему в математике, будет полезно внимательно изучить подробное описание процесса расчётов, которые есть в данной книге. Тому, кто не особо силён в математике, можно просто пробежать их глазами. Другими словами, читателю вполне достаточно думать: «Да, смысл этого метода мне не понятен, вычисления выглядят сложно, но, как бы там ни было, это вроде бы позволяет найти решение». Не нужно пытаться понять всё сразу. Давайте заниматься спокойно, без спешки. Однако автор просит читателя обязательно пробежать взглядом даже по непонятным объяснениям!

Скажу прямо: содержание данной книги сложнее, чем у книг «Манга о статистике» и «Манга о статистике. Регрессионный анализ». Поэтому, возможно, многие из читателей, только взглянув на какой-либо из приведённых процессов вычислений, подумают: «Это слишком сложно. Мне не хочется даже смотреть на это». Я прошу Вас приложить усилие, чтобы переступить через это чувство. Вычисления факторного анализа действительно сложны, но для читателей с познаниями в математике на уровне вступительных экзаменов научных и технических факультетов они должны быть вполне понятны.

Правда, это означает также, что для не очень разбирающихся в математике понять приведённые вычисления будут всё-таки трудно. Как бы там ни было, любой сложный предмет можно легко освоить, если двигаться постепенно, шаг за шагом.

В связи с округлением, некоторые результаты вычислений, приведённые в книге не совсем совпадают с результатами, которые читатели получают самостоятельно.

* *Син Такахаси*. Занимательная статистика. Манга. М.: Додэка, 2010. 224 с.

Син Такахаси. Занимательная статистика. Регрессионный анализ. Манга. М.: ДМК Пресс, 2014. 214 с.

В заключение я хотел бы здесь выразить благодарность коллективу Отдела разработок издательства Ohmsha, которые дали мне шанс взяться за перо. Я благодарю также сотрудников компании Trend-Pro, благодаря усилиям которых моя рукопись смогла стать мангой, господина re_akino, отвечавшего за написание сценария, а также господина Иноуэ Ироха, художника этой манги. Кроме того, я глубоко благодарен господину Сакаори Фумитакэ, преподавателю факультета социологии университета Риккё, давшему мне много ценных советов.

Син Такахаси
октябрь, 2006

СОДЕРЖАНИЕ

Пролог.

НАШ С ТОБОЙ ФАКТОРНЫЙ АНАЛИЗ	1
---	----------

Глава 1.

БАЗОВЫЕ ЗНАНИЯ ОБ АНКЕТИРОВАНИИ	15
--	-----------

1. Методы взятия выборок	17
2. Методы опроса	30
3. Критерии размера выборки	34
4. Методы случайного выбора и методы преднамеренного выбора	36
5. Количественное исследование и качественное исследование	38
6. Подходы к анализу данных	40

Глава 2.

БЛАНКИ АНКЕТ И ВОПРОСЫ	41
-------------------------------------	-----------

1. Структура бланка анкеты	43
2. Типы вопросов	48
3. Каких вопросов нужно избегать	52
4. Каких вопросов нужно избегать (продолжение)	60
5. Наличие нейтрального варианта ответа	62

Глава 3.

БАЗОВЫЕ ЗНАНИЯ ПО МАТЕМАТИКЕ	65
---	-----------

1. Матрицы корреляции	67
2. Единичные матрицы	68
3. Вращение	70
4. Собственные значения и собственные векторы	73

5. Симметричные матрицы	76
6. Дополнительные сведения о матрицах	80
7. Сумма квадратов отклонений, дисперсия случайной величины, квадратичное отклонение	86

Глава 4.

МЕТОД ГЛАВНЫХ КОМПОНЕНТ 89

1. Что такое метод главных компонент	93
2. Важные моменты метода главных компонент	98
3. Конкретный пример использования метода главных компонент ...	104
4. Выбор переменных и первая главная компонента.....	118
5. Первая главная компонента и общий показатель	124
6. Критерий суммарного вклада	125
7. Главные компоненты, начиная со второй	126
8. Дисперсия и собственные значения	127

Глава 5.

ФАКТОРНЫЙ АНАЛИЗ 129

1. Что такое факторный анализ	132
2. Важные моменты факторного анализа	139
3. Конкретный пример факторного анализа	150
4. Выборка в примере данной главы	192
5. Дополнения к важным моментам	193
6. Переменные с малыми значениями факторных нагрузок.....	197
7. Метод максимального правдоподобия	198
8. Почему вращение подразумевает метод варимакс	199
9. Матрица факторных нагрузок и матрица факторной структуры ...	200
10. Метод промакс	202
11. Верхний предел задаваемого числа общих факторов	208
12. Стоит ли считать метод главных факторов и метод варимакс устаревшими?	209
13. Терминология факторного анализа	210

Приложение.

РАЗЛИЧНЫЕ МЕТОДЫ АНАЛИЗА	219
1. Многомерный анализ	220
1.1. Общие сведения о многомерном анализе	220
1.2. Множественная регрессия	221
1.3. Логистическая регрессия	222
1.4. Кластерный анализ	224
1.5. Анализ соответствий и квантификация Хаяши методом III	226
1.6. Моделирование структурными уравнениями	229
2. Другое	231
2.1. Проверка статистической гипотезы	231
2.2. Метод Каплана-Мейера	233
Список литературы	235
Предметный указатель	236

ДЕЙСТВУЮЩИЕ ЛИЦА



ТАКАЦУ РУИ

Студентка 1-го курса. Оптимистка, необузданно эмоциональна, но в то же время серьёзна. Подрабатывает в кафе NORN'S.



ТАКАЦУ ХИРОСИ

Отец Руи. Начальник отдела маркетинговой компании. Из-за чрезмерной привязанности к дочери иногда ведёт себя по-детски.



ЯМАМОТО МАМОРУ

Подчинённый Такацу. В прошлом занимался с Руи статистикой. С Руи они больше чем друзья, но меньше чем возлюбленные.



ИГАРАСИ МИУ

Студентка 3-го курса. Подрабатывает вместе с Руи. В прошлом боялась статистики, но с некоторых пор погрузилась в неё с головой.



МИЯНО РИСА

Аспирантка. Подрабатывает вместе с Руи. Холодный характер, но добра к младшим. Талантлива и увлечена исследованиями.

Пролог

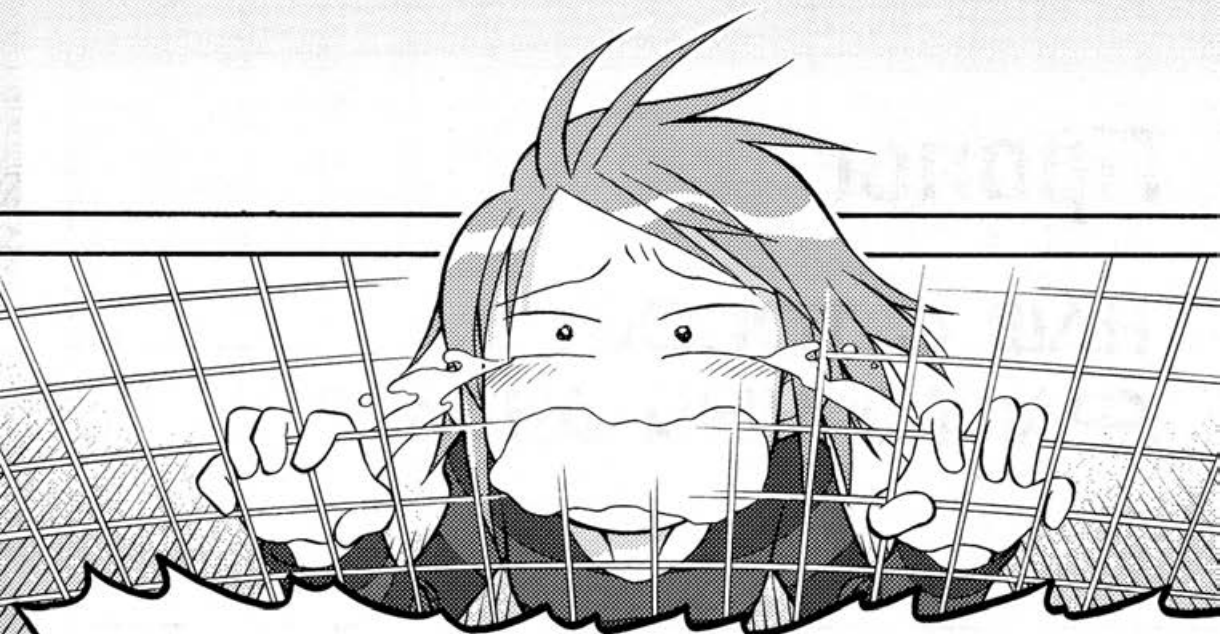
НАШ С ТОБОЮ
ФАКТОРНЫЙ АНАЛИЗ



ВУ-У-У



ВУ-У-У



ЯМАМОТО-САН!!!



ЯМАМОТО ДОЛЖЕН
ЕХАТЬ ЗА ГРАНИЦУ.
У НЕГО ТАМ ЕСТЬ
КОЕ-КАКАЯ РАБОТА.



РАЗ НАДО,
ЗНАЧИТ НАДО.
НЕ ТАК ЛИ,...

...РУЦ?

УА-А-А!



БЛАГОДАРЯ
ПОМОЩИ
ЯМАМОТО...

...ТЫ СМОГЛА
БЕЗ ПРОБЛЕМ
ПОСТУПИТЬ
В ИНСТИТУТ.



ТАК ПОЧЕМУ БЫ ТЕБЕ
НЕ СОСРЕДОТОЧИТЬСЯ
НА УЧЁБЕ ЗДЕСЬ,
В ЯПОНИИ?!

Да!



НЕ ТАК
ЛИ...?

ХМ!



Я ПОЕДУ ЗА НИМ
И БУДУ УЧИТЬСЯ ТАМ,
ЗА ГРАНИЦЕЙ!

ПРЕ...
ПРЕ-
КРАТЫ!



У Я... ЯМАМОТО
ТАМ ЕСТЬ ОЧЕНЬ
МНОГО РАБОТЫ!

НО
ВСЁ-ТАКИ!



ТЫ ЧТО, ХОЧЕШЬ ЕМУ
ТАМ МЕШАТЬ?

ОУ ОУ



НУ, ПОЙДЕМ!

ДА ЗАБУДЬ ТЫ...

...ПРО ЭТОГО
ЯМАМОТО!

НЕ МЕШАЙ.



ПАПА, А ЧТО ТАМ
ЗА РАБОТА
У ЯМАМОТО-САНА?

Ты ведь
должен
об этом
знать,
как его
начальник.
не так ли?

ЧТО?!



ЕГО РАБОТА
ЗА ГРАНИЦЕЙ,
РАЗУМЕЕТСЯ, СВЯЗАНА
С МАРКЕТИНГОМ!

НУ, НАПРИМЕР,
АНАЛИЗ АНКЕТ!



АНАЛИЗ АНКЕТ?
А ЧТО ЭТО ТАКОЕ?

Так...



НАПРИМЕР,
ФАКТОРНЫЙ
АНАЛИЗ.

ФАКТОРНЫЙ
АНАЛИЗ?

хлоп

НЕТ,
АЛЯ ТЕБЯ
ЭТО ЕЩЁ
СЛИШКОМ
СЛОЖНО,
ПРАВАА?



ФАКТОРНЫЙ
АНАЛИЗ...



ПУРШ

ЯМА-МОТО-САН!

РУЦ!

Я ИЗУЧИЛА
ФАКТОРНЫЙ
АНАЛИЗ,
ПОТОМУ ЧТО
ХОЧУ ПОМОГАТЬ
ТЕБЕ!

ОГО!
ЗНАЧИТ, ТЫ
УЖЕ ОСВОИЛА
ФАКТОРНЫЙ
АНАЛИЗ?!



ХОРОШО, ДАВАЙ
БУДЕМ РАБОТАТЬ
ВМЕСТЕ!

АА!



ВОТ ЧТО
Я ХОЧУ!

Хрусть

Ты что, Ручи?





АНКЕТА КАФЕ NORNS

Каждому заполнившему анкету — бесплатный купон на кассе!

Пол	муж / жен	Возраст	_____ лет
Род занятий		Годовой доход	_____ десятков тыс. йен

Как бы Вы охарактеризовали атмосферу нашего кафе?

Очень неприятная Неприятная Средняя Приятная Очень приятная

Как вы оцениваете одежду официанток и качество обслуживания?

Очень плохие Плохие Средние Хорошие Очень хорошие

Как Вы оцениваете наш чёрный чай?

Очень невкусный Невкусный Неоднозначно Вкусный Очень вкусный

Пожалуйста, оцените Ваше впечатление от кафе NORNS

Каково Ваше мнение о наших ценах?

Дёшево Средне Дорого

Проставьте цифры от 1 до 4 напротив каждого из видов чая в порядке убывания Ваших предпочтений.

Простой чай	место
Чай с лимоном	место
Чай с молоком	место
Чай с лепестками роз	место

Любите ли Вы бывать в кафе?

Да Нет





КАКОЙ ТУТ
ФАКТОРНЫЙ АНАЛИЗ?
ТАКУЮ АНКЕТУ
НИКТО ДАЖЕ
ЗАПОЛНЯТЬ
НЕ СТАНЕТ!

КАК?! ВЕДЬ
Я ТАК СТАРАЛАСЬ,
СОСТАВЛЯЛА!



ЭХ, РУИ... ТЫ ВООБЩЕ
ПРЕДСТАВЛЯЕШЬ СЕБЕ,
ЧТО ТАКОЕ
ФАКТОРНЫЙ АНАЛИЗ?



НУ,
НАЗВАНИЕ-ТО
Я ЗНАЮ...

хе-хе!

Я ВЕДЬ ТОЖЕ
НЕМНОЖКО
ИЗУЧАЛА
СТАТИСТИКУ.



ДЛЯ ПРОВЕДЕНИЯ
ФАКТОРНОГО АНАЛИЗА
НЕДОСТАТОЧНО ПРОСТО
СОБРАТЬ ДАННЫЕ АНКЕТ!

К ТОМУ ЖЕ,
ФАКТОРНЫЙ АНАЛИЗ —
НЕ ТАКАЯ ПРОСТАЯ
ВЕЩЬ!

АА, ПРАВАА?!



ТОГАА, РИСА,
ОБЪЯСНИ МНЕ ЕГО
ПОЖАЛУЙСТА!!

ТЫ ВЕДЬ ЗАНИМАЕШЬСЯ
СТАТИСТИКОЙ
В АСПИРАНТУРЕ, ЗНАЧИТ
ДОЛЖНА РАЗБИРАТЬСЯ.

КАК, ОПЯТЬ?!
История
повторяется.

ХВАТЬ

ХВАТЬ

Идея!



ХИ-ХИ-ХИ

ЕСТЬ БОЛЕЕ ПОХОДЯЩИЙ
ЧЕЛОВЕК, ЧЕМ Я.

ПРАВДА?
И КТО ЖЕ?



НУ-КА, МУУ!

ОБЪЯСНИ РУК
ФАКТОРНЫЙ АНАЛИЗ!

ЧТО, Я?!



А ЧТО, МУУ
ХОРОШО
РАЗБИРАЕТСЯ
В СТАТИСТИКЕ?

Неужели?!



КАК ТЫ МОЖЕШЬ
СОМНЕВАТЬСЯ?!
ВЕДЬ МУУ ОЧЕНЬ
СПОСОБНАЯ!

Но,
правда,
куда
мне
в этом
до Рису

И БРАТ У НЕЁ
РАБОТАЕТ
В МАРКЕТИНГОВОЙ
КОМПАНИИ.

К тому же она училась у меня!



МНОГОМЕРНЫЙ АНАЛИЗ — ЭТО, ТАК СКАЗАТЬ...

ОБЩЕЕ НАЗВАНИЕ ДЛЯ МЕТОДОВ АНАЛИЗА ДАННЫХ, СОСТОЯЩИХ ИЗ МНОГИХ ПЕРЕМЕННЫХ.

	Переменная 1	Переменная 2	Переменная 3	...	Переменная p
	↓	↓	↓		↓
	Пол	Возраст	Род занятий	...	Отношение к кафе
Респ. 1	Жен.	19 лет	Студент	...	Любит
Респ. 2	Жен.	27 лет	Домохозяйка	...	Не любит
⋮	⋮	⋮	⋮	...	⋮
Респ. n	Муж.	32 года	Служащий	...	Любит

ясно.

СУЩЕСТВУЕТ МНОГО ДРУГИХ ВИДОВ МНОГОМЕРНОГО АНАЛИЗА, ПОМИМО ФАКТОРНОГО.

Многомерный анализ

- Факторный анализ
- Множественная регрессия
- Логистическая регрессия
- Метод главных компонент
- Кластерный анализ
- Моделирование структурными уравнениями и другие

ДА?

ПРИ ЭТОМ,
ФАКТОРНЫЙ АНАЛИЗ
ЯВЛЯЕТСЯ МЕТОДОМ,
ПОЗВОЛЯЮЩИМ НАЙТИ
СКРЫТЫЕ ПРИЧИНЫ,
СТОЯЩИЕ ЗА ДАННЫМИ.

НАПРИМЕР,
"ПОТЕНЦИАЛЬНЫЕ
ВОЗМОЖНОСТИ",
"ПОДСОЗНАНИЕ"
И Т.П.

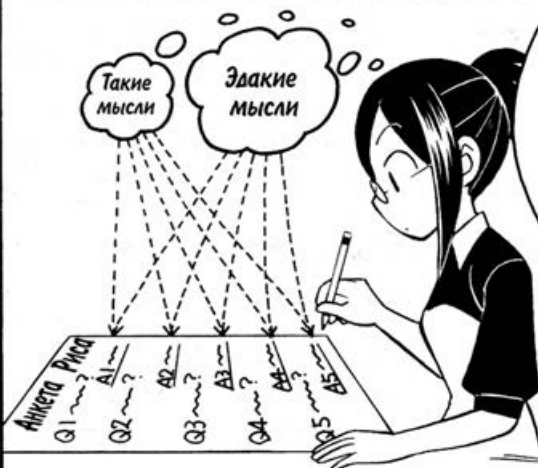
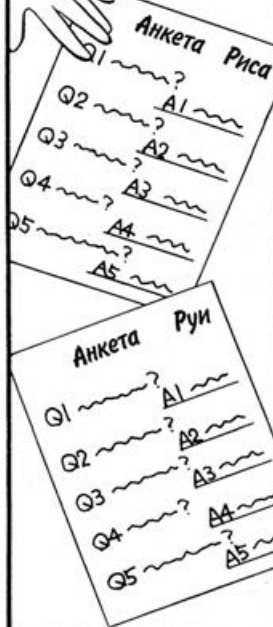
А ЧТО ЭТО
ЗНАЧИТ?



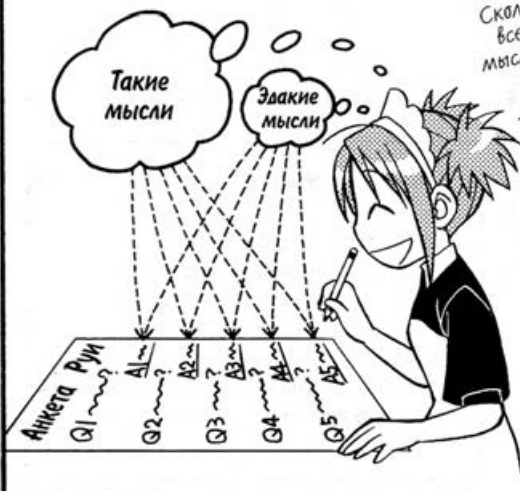
ПОПРОБУЮ
ОБЪЯСНИТЬ
НА ПАЛЬЦАХ.

ПУСТЬ У НАС
ЕСТЬ РЕЗУЛЬТАТЫ
АНКЕТИРОВАНИЯ.

В данном случае
число переменных
равно 5.



МОЖНО, НАВЕРНОЕ,
СКАЗАТЬ, ЧТО ЭТО
МЕТОД АНАЛИЗА,
ВЫЯВЛЯЮЩИЙ "МЫСЛИ"
АНКЕТИРУЕМЫХ,
СТОЯЩИЕ
ЗА ИХ ОТВЕТАМИ.



Сколько
всего
мыслей?

Какие
это
мысли?





Дя-дя

Я ПОНЯЛА!



Видишь,
миу-сэнсэй
тоже умеет
неплохо
объяснять.

Да!

Это я —
сэнсэй?



Миу тоже раньше
не разбиралась
в статистике.

Именно поэтому
она хорошо знает,
какие места труднее
для понимания.

Поэтому я думаю,
что она хорошо
подходит тебе
в учителя.



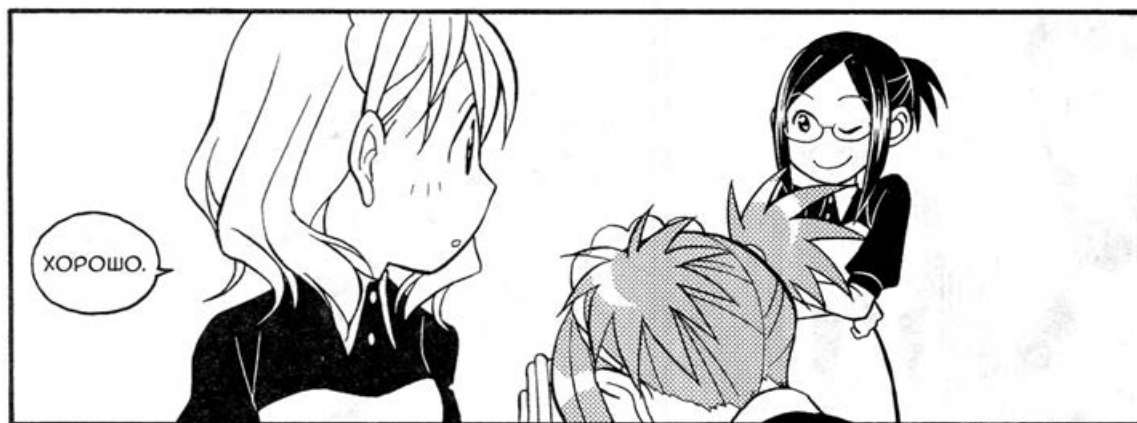
Пожалуйста,
миу, стань моим
учителем!

Объясни мне
факторный анализ!



Очень тебя
прошу!



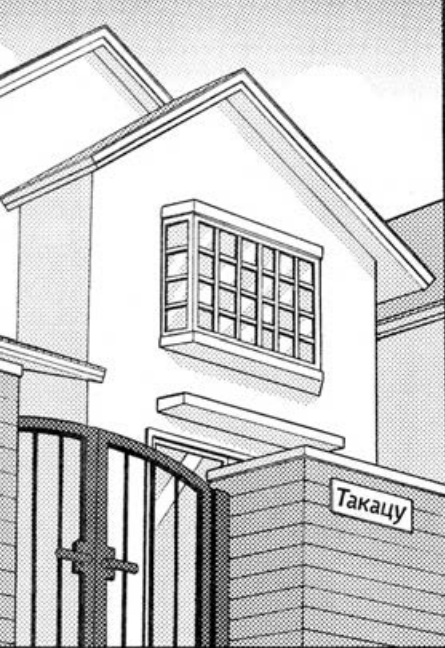


1

БАЗОВЫЕ ЗНАНИЯ ОБ АНКЕТИРОВАНИИ

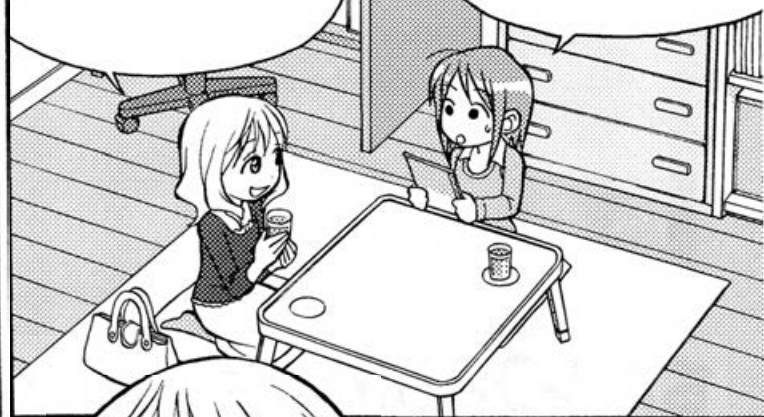
1. МЕТОДЫ ВЗЯТИЯ ВЫБОРОК
2. МЕТОДЫ ОПРОСА
3. КРИТЕРИИ РАЗМЕРА ВЫБОРКИ
4. МЕТОДЫ СЛУЧАЙНОГО ВЫБОРА
И МЕТОДЫ ПРЕДНАМЕРЕННОГО ВЫБОРА
5. КОЛИЧЕСТВЕННОЕ ИССЛЕДОВАНИЕ
И КАЧЕСТВЕННОЕ ИССЛЕДОВАНИЕ
6. ПОДХОДЫ К АНАЛИЗУ ДАННЫХ





ИТАК, СЕГОДНЯ
МЫ ИЗУЧИМ ОСНОВЫ
АНКЕТИРОВАНИЯ.

А НЕ ФАКТОРНЫЙ
АНАЛИЗ?



ДЕЛО В ТОМ,
ЧТО ОБЪЕКТОМ
ФАКТОРНОГО АНАЛИЗА
ЧАСТО ЯВЛЯЮТСЯ
ДАННЫЕ
АНКЕТИРОВАНИЯ.

ПОЭТОМУ ЛУЧШЕ
НАЧАТЬ ИМЕННО
С ЭТОГО.

А...
ВОТ
ОНО ЧТО.



ПОЭТОМУ ДАВАЙ
СЕГОДНЯ НАЧНЁМ
С МЕТОДОВ
ВЗЯТИЯ ВЫБОРОК
И МЕТОДОВ ОПРОСА!

**ВУХ-
ВУХ**

ХОРОШО!
Я БУДУ ДЕЛАТЬ,
КАК ТЫ
СКАЖЕШЬ!



МЕТОДЫ ВЗЯТИЯ ВЫБОРОК

ЗНАЕШЬ ЛИ ТЫ, ЧТО ТАКОЕ ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ И ВЫБОРКА?

РАЗУМЕЕТСЯ!

ГЕНЕРАЛЬНАЯ СОВОКУПНОСТЬ — ЭТО ГРУППА, КОТОРУЮ МЫ ДОЛЖНЫ НА САМОМ ДЕЛЕ ИССЛЕДОВАТЬ....

...А ВЫБОРКА — ЭТО ГРУППА, СОСТОЯЩАЯ ИЗ НЕКОТОРОГО ЧИСЛА ЭКЗЕМПЛЯРОВ, ВЗЯТЫХ ИЗ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ, ПРАВИЛЬНО?!

Генеральная совокупность

Взятие выборки

Выборка

ПРАВИЛЬНО!

Я исследую всё!

Генеральная совокупность

Гипотеза о целом на основе его части

Выборка

ПРИ ЭТОМ НЕОБХОДИМО ПОМНИТЬ О СЛЕДУЮЩЕМ...

КСТАТИ, ИССЛЕДОВАНИЕ ВСЕЙ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ НАЗЫВАЕТСЯ СПЛОШНЫМ ИССЛЕДОВАНИЕМ,

А ИССЛЕДОВАНИЕ ВЫБОРКИ НАЗЫВАЕТСЯ ВЫБОРОЧНЫМ ИССЛЕДОВАНИЕМ.

ПОНЯТНО.

?

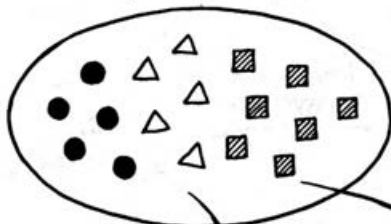
ПОНИМАЕШЬ ЛИ ТЫ,
ЧТО БЕССМЫСЛЕННО ПРОВОДИТЬ
ВЫБОРОЧНОЕ ИССЛЕДОВАНИЕ,
ЕСЛИ ВЫБОРКА НЕ ЯВЛЯЕТСЯ
"ТОЧНОЙ МИНИАТЮРОЙ
ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ"?



АА, ТЫ ПРАВА!
ВЕДЬ ТОГДА
РЕЗУЛЬТАТЫ
ОКАЖУТСЯ
ПРЕАВЗЯТЫМИ.



Генеральная совокупность



Взятие
выборки

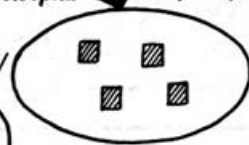
Выборка



Взятие
выборки

Выборка

Да тут
одни
квадратики.



ИЗБЕЖАТЬ ЭТОГО
НАМ ПОМОГУТ
МЕТОДЫ ВЗЯТИЯ
ВЫБОРОК!



МЕТОДЫ ВЗЯТИЯ
ВЫБОРОК — ЭТО
ОБЩЕЕ НАЗВАНИЕ
МЕТОДОВ ВЫБОРА
ЭКЗЕМПЛЯРОВ
ИЗ ГЕНЕРАЛЬНОЙ
СОВОКУПНОСТИ.



СЕГОДНЯ Я РАССКАЖУ
О ЧЕТЫРЁХ НАИБОЛЕЕ
ТИПИЧНЫХ МЕТОДАХ:
ПРОСТОМ СЛУЧАЙНОМ ВЫБОРЕ,
РАССЛОЕННОМ ВЫБОРЕ,
АВУХСТУПЕНЧАТОМ ВЫБОРЕ
И РАССЛОЕННОМ
АВУХСТУПЕНЧАТОМ ВЫБОРЕ.

ХОРОШО!

ПУСТЬ ПЕРЕД НАМИ
СТОИТ ВОТ
ТАКАЯ ЗАДАЧА.

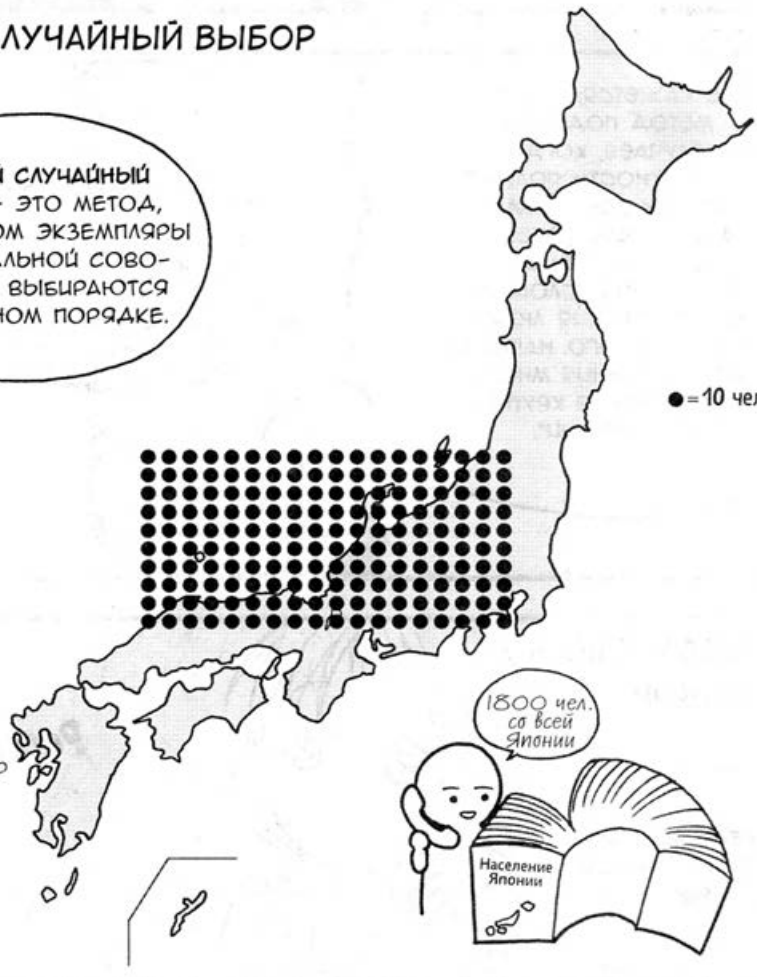


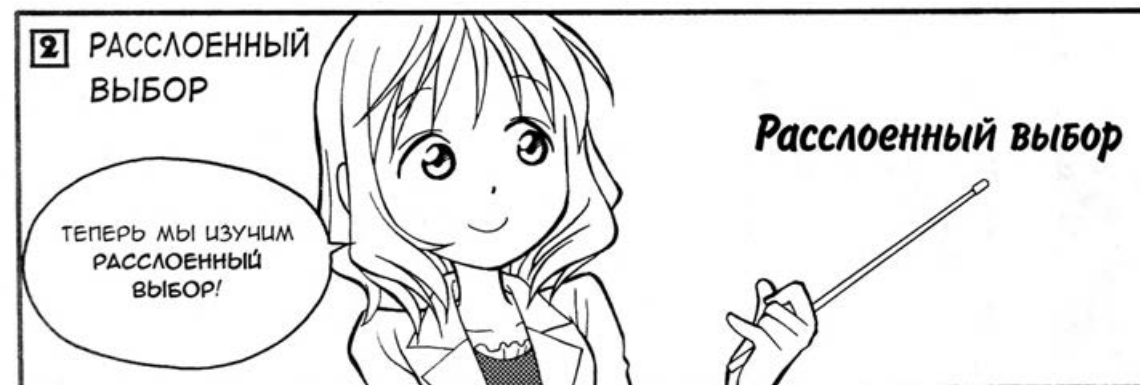
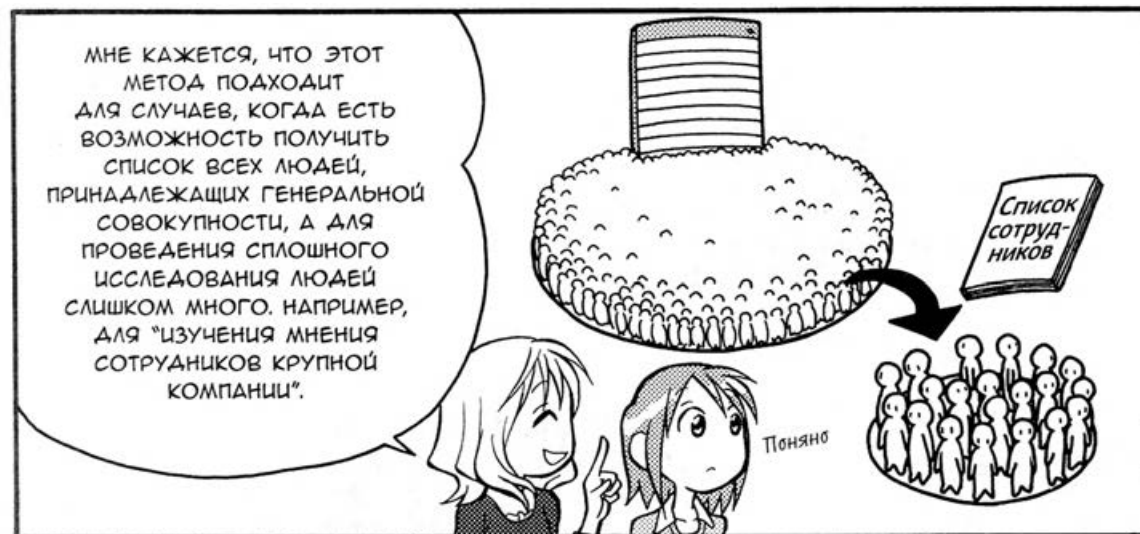
Ого!

Выбор
1800 человек
из всего
населения
Японии

1 ПРОСТОЙ СЛУЧАЙНЫЙ ВЫБОР

ПРОСТОЙ СЛУЧАЙНЫЙ
ВЫБОР — ЭТО МЕТОД,
ПРИ КОТОРОМ ЭКЗЕМПЛЯРЫ
ИЗ ГЕНЕРАЛЬНОЙ СОВО-
КУПНОСТИ ВЫБИРАЮТСЯ
В СЛУЧАЙНОМ ПОРЯДКЕ.

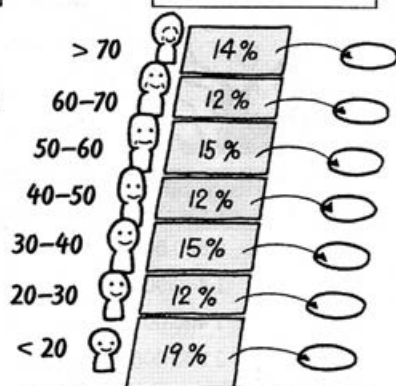




Слои по району проживания

Слои по возрасту

СНАЧАЛА МЫ ДЕЛИМ
ГЕНЕРАЛЬНУЮ
СОВОКУПНОСТЬ НА СЛОИ,
НАПРИМЕР, ПО
"РАЙОНУ ПРОЖИВАНИЯ",
"РОДУ ЗАНЯТИЙ" ИЛИ
"ПОЛУЧЕННОМУ
ОБРАЗОВАНИЮ",...



...А ЗАТЕМ ПРИМЕНЯЕМ
К КАЖДОМУ СЛОЮ МЕТОД
ПРОСТОГО СЛУЧАЙНОГО
ВЫБОРА

ЗНАЧИТ, МЫ
СНАЧАЛА ДЕЛИМ,
А ПОТОМ
ВЫБИРАЕМ, ТАК?



ДА. ИНОГДА
В ОТВЕТАХ,
ПОЛУЧЕННЫХ
ОТ РАЗНЫХ СЛОЕВ,
ЕСТЬ БОЛЬШАЯ
НЕРАВНОМЕРНОСТЬ.



Так,
«Поддерживаемая политическая партия»
может зависеть от района проживания,
а «Любимый артист» — от возраста
респондента.

ЯСНО.

ЭТОТ МЕТОД ПОДХОДИТ
ДЛЯ СЛУЧАЕВ, КОГДА
МЫ МОЖЕМ ЗАРАНЕЕ
ПРЕАВИАТЬ ТАКУЮ
НЕРАВНОМЕРНОСТЬ.



НУ, ДА.
ПОНЯТНО...

ДАВАЙ ТЕПЕРЬ ПРИМЕНИМ
МЕТОД РАССЛОЕННОГО
ВЫБОРА, ПРИНЯВ ЗА СЛОИ
"РАЙОН ПРОЖИВАНИЯ".



ДАВАЙ!



ЧИСЛО ЛЮДЕЙ, ВЫБИРАЕМЫХ
ИЗ КАЖДОГО РАЙОНА, СЛЕДУЮЩЕЕ.

Район проживания	Численность населения	Доля	Число выбираемых	Район проживания	Численность населения	Доля	Число выбираемых
Хоккайдо	5 600 000	0,045	80	Мие	1 800 000	0,014	30
Аомори	1 400 000	0,011	20	Сига	1 300 000	0,010	20
Иватэ	1 400 000	0,011	20	Киото	2 600 000	0,021	40
Мияги	2 300 000	0,018	30	Осака	8 700 000	0,070	130
Акита	1 100 000	0,009	20	Хёго	5 500 000	0,044	80
Ямагата	1 200 000	0,010	20	Нара	1 400 000	0,011	20
Фукусима	2 100 000	0,017	30	Вакаяма	1 000 000	0,008	10
Ибараки	2 900 000	0,023	40	Тоттори	600 000	0,005	10
Тотиги	2 000 000	0,016	30	Симанэ	700 000	0,006	10
Гумма	2 000 000	0,016	30	Окаяма	1 900 000	0,015	30
Саитама	6 900 000	0,055	100	Хиросима	2 800 000	0,022	40
Тиба	5 900 000	0,047	90	Ямагути	1 500 000	0,012	20
Токио	12 000 000	0,096	170	Токусима	800 000	0,006	10
Канагава	8 400 000	0,067	120	Кагава	1 000 000	0,008	10
Ниигата	2 400 000	0,019	30	Эхимэ	1 400 000	0,011	20
Тояма	1 100 000	0,009	20	Коти	800 000	0,006	10
Исикава	1 100 000	0,009	20	Фукуока	5 000 000	0,040	70
Фукуи	800 000	0,006	10	Сага	800 000	0,006	10
Яманаси	800 000	0,006	10	Нагасаки	1 500 000	0,012	20
Нагано	2 200 000	0,018	30	Кумамото	1 800 000	0,014	30
Гифу	2 100 000	0,017	30	Оита	1 200 000	0,010	20
Сидзуока	3 700 000	0,030	50	Миядзак	1 100 000	0,009	20
Айти	7 000 000	0,056	100	Кагосима	1 700 000	0,014	20
				Окинава	1 300 000	0,010	20
				Всего	124 600 000	1	1800

$$\frac{\text{Население данного района}}{\text{Общее население}} = \frac{7000000}{124600000} = 0,056$$

$$\begin{aligned} \text{Общее число выбираемых} \times \text{Доля} &= \\ &= 1800 \times 0,056 = \\ &= 101,1 \approx 100 \end{aligned}$$

* По данным Статистической службы Министерства внутренних дел и коммуникаций Японии за 2006 год

ЗНАЧИТ, МЫ ВЫБИРАЕМ
100 ЧЕЛОВЕК ИЗ АЙТИ,
170 ЧЕЛОВЕК ИЗ ТОКИО
И ТАК ДАЛЕЕ, ДА?



АГА!
ВЫБИРАЕМОЕ ЧИСЛО
СООТВЕТСТВУЕТ ДОЛЕ.



80 чел.
с Хаккайдо



30 чел.
из Мияги



40 чел. из
Хирасимы



30 чел. из
Кумамото



100 чел.
из Айти



170 чел.
из Токио



ПОДОЖДИ! ДЛЯ ЭТОГО
МЕТОДА ТОЖЕ ВЕДЬ
НУЖЕН "ПОЛНЫЙ
СПИСОК", ДА?

ДА, ТЫ
ПРАВА.

К ТОМУ ЖЕ,
В СПИСКЕ ДОЛЖНА
СОДЕРЖАТЬСЯ
ИНФОРМАЦИЯ
О СЛОЯХ.

ДА,
ПОНЯТНО!

Лет	Адрес
21	Канатава
23	Токио
19	Токио
27	Токио

Имя	Пол	Лет	Адрес
Такацу Руи	жен.	19	Токио
Игараси Миу	жен.	21	Токио
Мияно Риса	жен.	23	Токио
Ямамото М.	муж.	27	Токио
	муж.	48	Токио

3 ДВУХСТУПЕНЧАТЫЙ ВЫБОР

ДВУХСТУПЕНЧАТЫЙ
ВЫБОР —
ЭТО МЕТОД
ВЗЯТИЯ ВЫБОРКИ
В ДВА ЭТАПА.

В ДВА
ЭТАПА?



ИСПОЛЬЗУЯ НАШ ПРИМЕР,
НА ПЕРВОМ ЭТАПЕ
МЫ ВЫБИРАЕМ РАЙОНЫ,

*На самом деле, конечно,
это могут быть не только
районы*



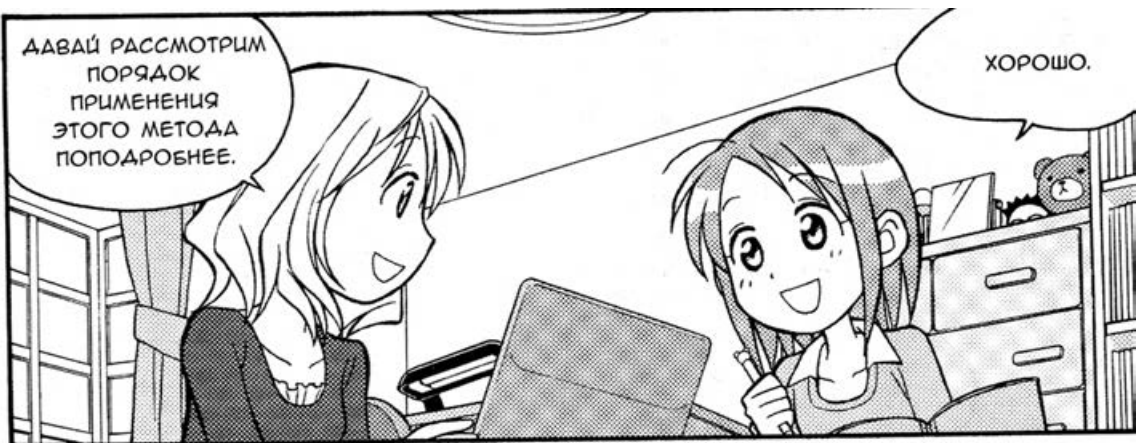
А НА ВТОРОМ ЭТАПЕ
ПРИМЕНЯЕМ К КАЖДОМУ
ИЗ ВЫБРАННЫХ РАЙОНОВ
МЕТОД ПРОСТОГО
СЛУЧАЙНОГО ВЫБОРА.



ЯСНО.

ДАВАЙ РАССМОТРИМ
ПОРЯДОК
ПРИМЕНЕНИЯ
ЭТОГО МЕТОДА
ПОПОДРОБНЕЕ.

ХОРОШО.



В НАСТОЯЩЕЕ ВРЕМЯ
НАСЕЛЕНИЕ 47 РАЙОНОВ
ЯПОНИИ ПРИМЕРНО
СЛЕДУЮЩЕЕ.



Название района	Численность населения*	Первый номер в общем списке	Последний номер в общем списке
Хоккайдо	5 600 000	1	5 600 000
Аомори	1 400 000	5 600 001	7 000 000
Иватэ	1 400 000	7 000 001	8 400 000
Мияги	2 300 000	8 400 001	10 700 000
Акита	1 100 000	10 700 001	11 800 000
Ямагата	1 200 000	11 800 001	13 000 000
Фукусима	2 100 000	13 000 001	15 100 000
Ибараки	2 900 000	15 100 001	18 000 000
Тотиги	2 000 000	18 000 001	20 000 000
Гумма	2 000 000	20 000 001	22 000 000
Саитама	6 900 000	22 000 001	28 900 000
Тиба	5 900 000	28 900 001	34 800 000
Токио	12 000 000	34 800 001	46 800 000
Канагава	8 400 000	46 800 001	55 200 000
Ниигата	2 400 000	55 200 001	57 600 000
Тояма	1 100 000	57 600 001	58 700 000
Исикава	1 100 000	58 700 001	59 800 000
Фукуи	800 000	59 800 001	60 600 000
Яманаси	800 000	60 600 001	61 400 000
Нагано	2 200 000	61 400 001	63 600 000
Гифу	2 100 000	63 600 001	65 700 000
Сидзуока	3 700 000	65 700 001	69 400 000
Айти	7 000 000	69 400 001	76 400 000
Мизэ	1 800 000	76 400 001	78 200 000

Название района	Численность населения*	Первый номер в общем списке	Последний номер в общем списке
Сига	1 300 000	78 200 001	79 500 000
Киото	2 600 000	79 500 001	82 100 000
Осака	8 700 000	82 100 001	90 800 000
Хёго	5 500 000	90 800 001	96 300 000
Нара	1 400 000	96 300 001	97 700 000
Вакаяма	1 000 000	97 700 001	98 700 000
Тоттори	600 000	98 700 001	99 300 000
Симанэ	700 000	99 300 001	100 000 000
Окаяма	1 900 000	100 000 001	101 900 000
Хиросима	2 800 000	101 900 001	104 700 000
Ямагути	1 500 000	104 700 001	106 200 000
Токусима	800 000	106 200 001	107 000 000
Кагава	1 000 000	107 000 001	108 000 000
Эхимэ	1 400 000	108 000 001	109 400 000
Коти	800 000	109 400 001	110 200 000
Фукуока	5 000 000	110 200 001	115 200 000
Сага	800 000	115 200 001	116 000 000
Нагасаки	1 500 000	116 000 001	117 500 000
Кумамото	1 800 000	117 500 001	119 300 000
Оита	1 200 000	119 300 001	120 500 000
Миядзаки	1 100 000	120 500 001	121 600 000
Кагосима	1 700 000	121 600 001	123 300 000
Окинава	1 300 000	123 300 001	124 600 000
Всего	124 600 000		

*По данным Статистической службы Министерства внутренних дел и коммуникаций Японии за 2006 год

Шаг 1

Получаем десять случайных чисел в интервале от 1 до 124 600 000 с помощью, например, генератора случайных чисел RAND программы Excel.

Случайное число 1	104 333 307
Случайное число 2	8 007 588
Случайное число 3	35 224 073
Случайное число 4	72 352 247
Случайное число 5	3 951 586
Случайное число 6	114 308 209
Случайное число 7	3 724 893
Случайное число 8	100 701 197
Случайное число 9	62 591 858
Случайное число 10	89 167 908

Шаг 2

По общему списку находим районы, диапазоны которых содержат случайные числа, полученные на Шаге 1.

Случ. число 1	104 333 307	→	Хиросима
Случ. число 2	8 007 588	→	Иватэ
Случ. число 3	35 224 073	→	Токио
Случ. число 4	72 352 247	→	Айти
Случ. число 5	3 951 586	→	Хоккайдо
Случ. число 6	114 308 209	→	Фукуока
Случ. число 7	3 724 893	→	Хоккайдо
Случ. число 8	100 701 197	→	Тояма
Случ. число 9	62 591 858	→	Нагано
Случ. число 10	89 167 908	→	Осака

ЧИСЛО 124 600 000 —
ЭТО ОБЩАЯ ЧИСЛЕННОСТЬ
НАСЕЛЕНИЯ ВСЕХ 47 РАЙОНОВ
ЯПОНИИ.



ХОККАЙДО
ОКАЗАЛСЯ
ВЫБРАН
ДВА РАЗА?!



НИЧЕГО НЕ
ПОДЕЛАЕШЬ —
МЫ ДОЛЖНЫ
ДЕЙСТВОВАТЬ
В СООТВЕТСТВИИ
С МЕТОДОМ.



Шаг 3

Выбираем в случайном порядке по 180 человек из каждого района, полученного на Шаге 2.

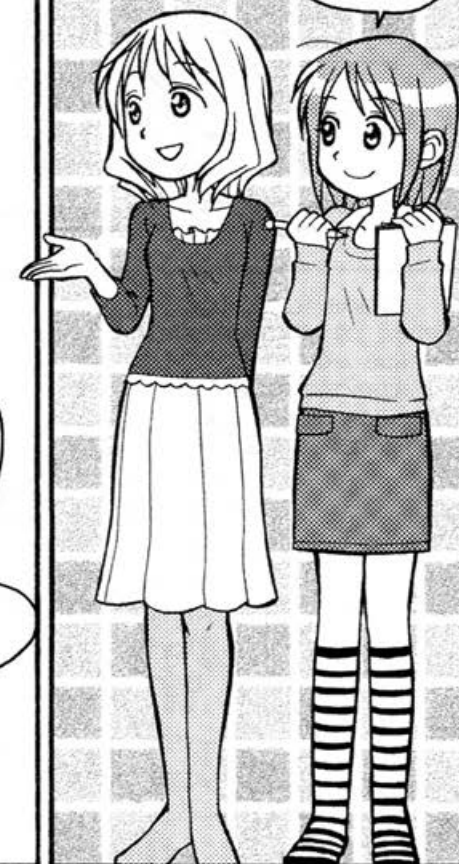


ЧТО КАСАЕТСЯ ХОККАЙДО, МЫ ДОЛЖНЫ ПРИМЕНИТЬ МЕТОД ПРОСТОГО СЛУЧАЙНОГО ВЫБОРА ОДИН РАЗ ДЛЯ 360 ЧЕЛОВЕК, А НЕ ДВА РАЗА ПО 180 ЧЕЛОВЕК.

ЯСНО.

В ЭТОМ ПРИМЕРЕ МЫ ВЫБИРАЛИ ПО 180 ЧЕЛОВЕК ИЗ 10 РАЙОНОВ, ОДНАКО В ДЕЙСТВИТЕЛЬНОСТИ "ВЫБИРАЕМОЕ ЧИСЛО РАЙОНОВ" И "ЧИСЛО ЧЕЛОВЕК ИЗ КАЖДОГО ВЫБРАННОГО РАЙОНА" ОТАДАНЫ НА УСМОТРЕНИЕ АНАЛИТИКА.

ПОНЯТНО.





ДВУХСТУПЕНЧАТЫЙ ВЫБОР БОЛЕЕ ПРАКТИЧЕН, ЧЕМ ПРОСТОЙ СЛУЧАЙНЫЙ ВЫБОР И РАССЛОЕННЫЙ ВЫБОР.

ПОЧЕМУ?



ВЕДЬ ДОСТАТОЧНО ОБЛАДАТЬ СПИСКАМИ НАСЕЛЕНИЯ ТОЛЬКО РАЙОНОВ, ВЫБРАННЫХ НА ПЕРВОМ ЭТАПЕ, НЕ ТАК ЛИ?

АХ, ДА!
КОНЕЧНО!



НО ЕСТЬ ТАКЖЕ НЕСКОЛЬКО ПРОБЛЕМ.

ВО-ПЕРВЫХ, НАМ НЕОБХОДИМО ЗНАТЬ ТОЧНУЮ ЧИСЛЕННОСТЬ НАСЕЛЕНИЯ КАЖДОГО РАЙОНА.

ДЛЯ ШАГА 1,
НЕ ТАК ЛИ?



КРОМЕ ТОГО, НАМ ПРИДЁТСЯ ИГНОРИРОВАТЬ МНЕНИЕ ЛЮДЕЙ, КОТОРЫЕ ЖИВУТ В РАЙОНАХ, НЕ ВЫБРАННЫХ НА ПЕРВОМ ЭТАПЕ.

И ВПРАВДУ.

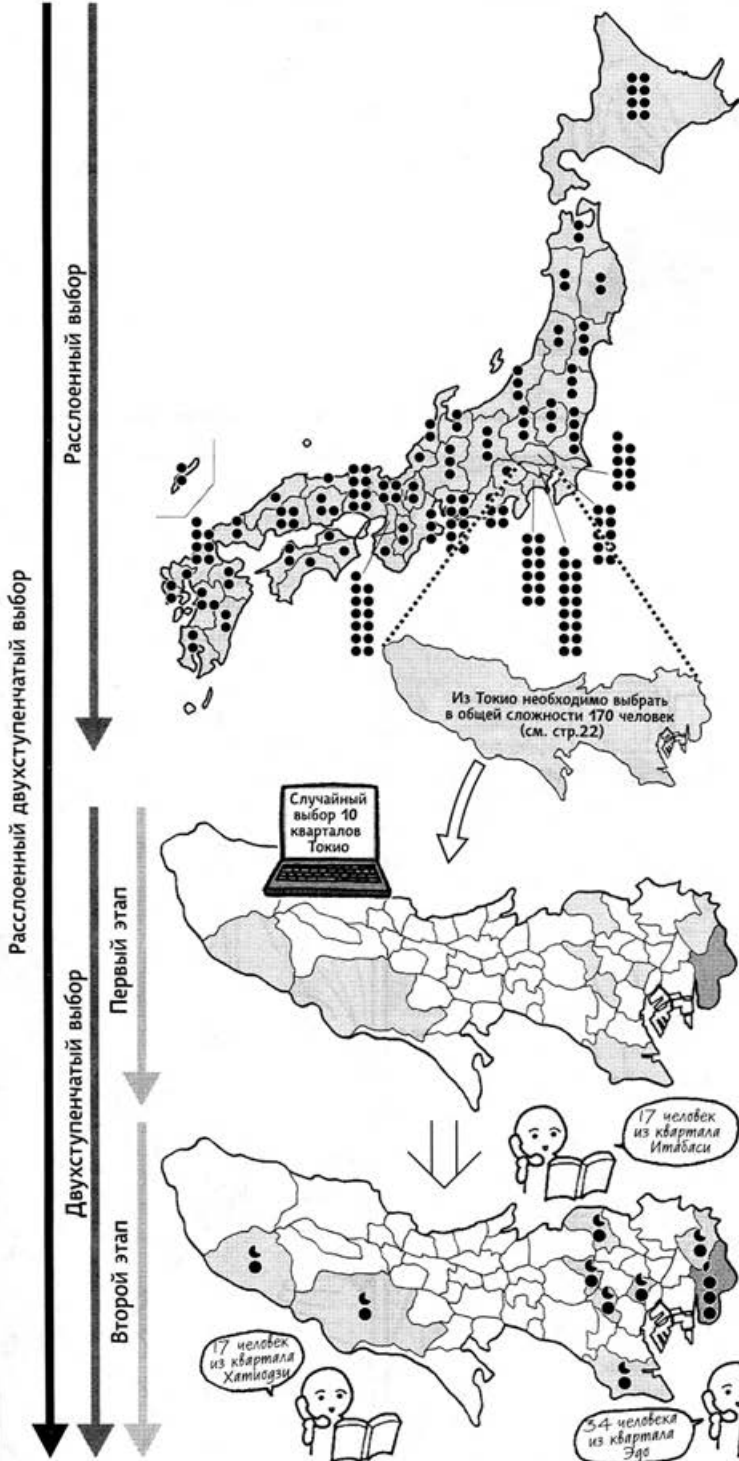
А как же районы региона Хокурику?

А как же районы острова Сикоку?

4 РАССЛОЕННЫЙ ДВУХСТУПЕНЧАТЫЙ ВЫБОР

И В ЗАВЕРШЕНИЕ У НАС — РАССЛОЕННЫЙ ДВУХСТУПЕНЧАТЫЙ ВЫБОР

Расслоенный двухступенчатый выбор
КАК ЯСНО ИЗ НАЗВАНИЯ, ОН ПРЕДСТАВЛЯЕТ СОБОЙ КОМБИНАЦИЮ МЕТОДОВ РАССЛОЕННОГО ВЫБОРА И ДВУХСТУПЕНЧАТОГО ВЫБОРА. Я ОБЪЯСНЮ ЕГО КРАТКО, С ПОМОЩЬЮ РИСУНКА.



ОГО! ЭТОТ МЕТОД ОБЕСПЕЧИВАЕТ РАВНОМЕРНОСТЬ, И, КРОМЕ ТОГО, НЕ НУЖНО ПОЛНОГО СПИСКА ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ, ДА?



ЭТОТ МЕТОД НАИБОЛЕЕ ПРАКТИЧЕН ДЛЯ ПРОВЕДЕНИЯ ИССЛЕДОВАНИЙ ОТНОСИТЕЛЬНО БОЛЬШОГО МАСШТАБА.

Расслоенный двухступенчатый выбор
 Расслоенный выбор
 Двухступенчатый выбор
 Первый этап
 Второй этап

2. МЕТОДЫ ОПРОСА

НА ЭТОМ МЫ ЗАКОНЧИМ ИЗУЧАТЬ МЕТОДЫ ВЗЯТИЯ ВЫБОРОК И ПЕРЕИДЁМ К МЕТОДАМ ОПРОСА.

ХОРОШО!

МЕТОДОВ ОПРОСА ТОЖЕ СУЩЕСТВУЕТ ВЕЛИКОЕ МНОЖЕСТВО.

ОГО!

Почтовый опрос

Отправляем анкеты по почте и получаем ответы.



Интернет-опрос



Используем сеть Интернет.

Опрос «до востребования»

Я приду опять такого-то числа.



До этого числа заполните, пожалуйста, эту анкету.

Интервью-опрос

Встречаемся и спрашиваем непосредственно.



Телефонный опрос









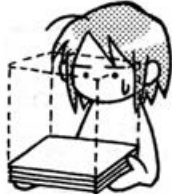



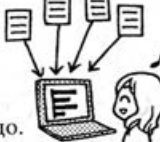
Обзваниваем номера по телефонному справочнику.



Обзваниваем номера, состоящие из цифр, полученных случайным методом.

КОНЕЧНО, ВСЕ МЕТОДЫ ОПРОСА ВАЖНЫ, ОДНАКО ЗАЕСЬ Я ПРЕДЛАГАЮ СРАВНИТЬ ОСОБЕННОСТИ ТОЛЬКО НАИБОЛЕЕ ТИПИЧНЫХ ИЗ НИХ — ПОЧТОВОГО ОПРОСА И ИНТЕРНЕТ-ОПРОСА.

ХОРОШО!

	<p>Почтовый опрос Метод рассылки респондентам анкет и получения от них ответов через какое-то время.</p> 	<p>Интернет-опрос Со свободным участием Метод получения ответов через Web от неограниченного числа респондентов. С группой наблюдения Метод получения ответов через Web от заранее выбранной группы наблюдения.</p> 
Территории и списки	<p>Позволяет опрашивать респондентов на обширной территории. Необходимы списки с адресами для отправки анкет.</p> 	<p>Позволяет опрашивать респондентов на обширной территории. Никакие списки не нужны, так как респонденты участвуют в опросе по собственной инициативе</p> 
Число вопросов	<p>Респондент имеет возможность отвечать на вопросы без спешки, откладывая часть вопросов «на завтра». Поэтому вопросов может быть много.</p> 	<p>От респондента требуется ответить на все вопросы сразу, отложить часть вопросов «на завтра» невозможно. Поэтому вопросов не может быть слишком много.</p> 
Уровень доверия к ответам	<p>Обычно возвращается только около 20 ... 30% разосланных анкет. Следовательно, даже если аналитик спланировал выборку, которую можно рассматривать как «точную миниатюру генеральной совокупности», и разослал анкеты выбранным респондентам, нет никакой гарантии, что распределение вернувшихся ответов можно будет рассматривать как «точную миниатюру генеральной совокупности».</p> 	<p>Со свободным участием Так как респонденты участвуют в опросе по собственной инициативе, аналитик не имеет возможности спланировать выборку на основе предположения о генеральной совокупности.</p>  <p>С группой наблюдения На первый взгляд кажется, что умелый выбор из полной группы наблюдения позволит сформировать выборку, которую можно рассматривать как «точную миниатюру генеральной совокупности». Однако полная группа наблюдения — это люди, у которых есть компьютер и которые по собственной воле стали членами группы наблюдения, поэтому очевидно, что для изучения мнения «обычных» людей они не очень-то и подходят.</p> 
Сроки сбора данных	Долго	Быстро
Ввод данных	<p>Производится самим аналитиком либо его помощниками.</p> 	<p>Производится респондентами. Лучше сказать, что ввод данных происходит автоматически одновременно с ответами респондентов, поэтому операции по вводу данных производить не надо.</p> 

* Анкеты (бланки анкет) называют также «опросными листами», или «опросниками».

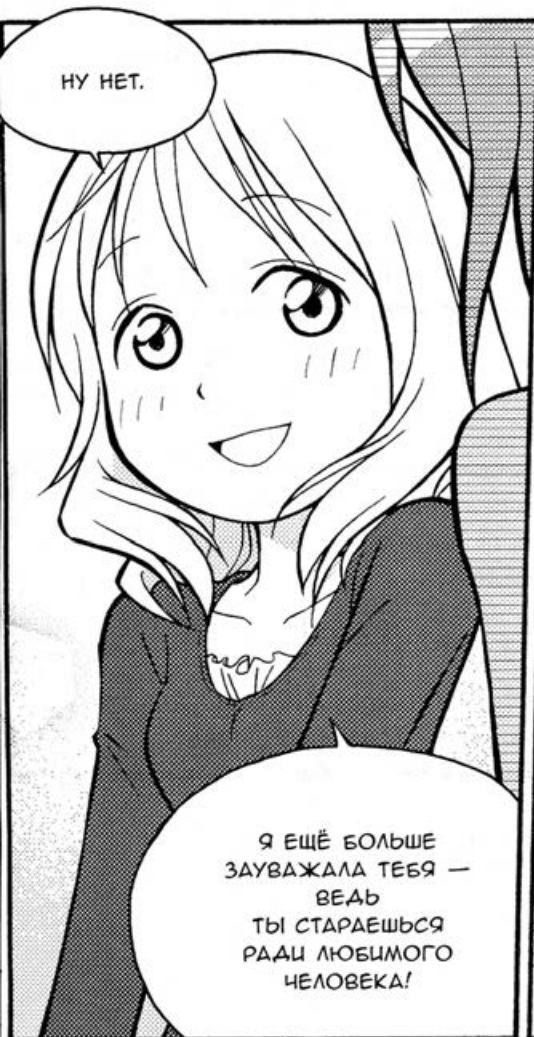






А, ВОТ ОНО ЧТО?
ТОГДА ПОНЯТНО.

ХЕ-ХЕ.
МОЖЕТ БЫТЬ,
ТЫ СЧИТАЕШЬ, ЧТО
МОЯ МОТИВАЦИЯ
НЕ ОЧЕНЬ ЧИСТА?



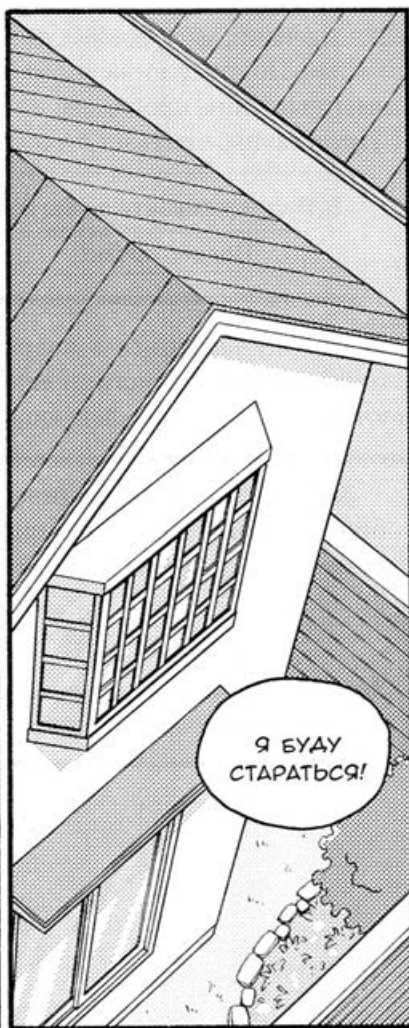
НУ НЕТ.

Я ЕЩЕ БОЛЬШЕ
ЗАУВАЖАЛА ТЕБЯ —
ВЕДЬ
ТЫ СТАРАЕШЬСЯ
РАДИ ЛЮБИМОГО
ЧЕЛОВЕКА!



ТАК ЧТО,
РУЦ,
ПРОДОЛЖАЙ
ПРИЛАГАТЬ
УСИЛИЯ!

СПАСИБО
ТЕБЕ,
МЦУ!



Я БУДУ
СТАРАТЬСЯ!

3. КРИТЕРИИ РАЗМЕРА ВЫБОРКИ

Число экземпляров, формирующих выборку, называют *размером выборки*.

Чем больше размер выборки, тем лучше она представляет генеральную совокупность, поэтому есть смысл в сборе как можно большего количества данных.

Однако увеличение размера выборки — не всегда простая задача ввиду таких проблем, как ограниченность сроков сбора данных или ограниченность финансирования. Было бы хорошо, если бы существовал некий статистический критерий минимального размера выборки, которая представляла бы собой уменьшенную копию генеральной совокупности. Однако, к сожалению, такого критерия не существует.

И хотя в мире анкетирования существует представление о том, что критерий минимального размера выборки равен «примерно 400», это значение не следует слепо принимать на веру — лучше просто принять его к сведению. Однако почему-то многие думают: «Это значение заслуживает доверия с точки зрения статистики, хотя я и не понимаю его смысла». Поэтому ниже я объясню метод получения этого значения, «примерно 400», и связанные с ним проблемы.

Представим, что газета А планирует через месяц провести следующее анкетирование:

В. Поддерживаете ли Вы кабинет министров? (поставьте только один кружок).

1. Поддерживаю

2. Не поддерживаю

Я не буду вдаваться в подробности, чтобы не усложнять объяснение, но скажу, что с точки зрения статистики ещё до проведения анкетирования с *доверительной вероятностью* 95% известно следующее:

Хотя, какие бы статистические методы не использовались, уровень поддержки кабинета министров P в генеральной совокупности «все избиратели» точно узнать невозможно, он, несомненно, находится в интервале

$$p - 1,96 \times \sqrt{\frac{p \times (1-p)}{n}} \leq P \leq p + 1,96 \times \sqrt{\frac{p \times (1-p)}{n}}.$$

Здесь P — уровень поддержки кабинета министров в генеральной совокупности, p — уровень поддержки кабинета министров в выборке, n — размер выборки.

Кстати, доверительная вероятность здесь означает «достоверность утверждения в рамке», её ещё называют *доверительным уровнем* или *коэффициентом доверия*.

Ещё раз внимательно взгляните на вышеприведённое утверждение в рамке. Чем меньше значение выражения

$$1,96 \times \sqrt{\frac{p \times (1-p)}{n}},$$

тем меньше ширина интервала, и тем более убедительным будет утверждение. Здесь от Вас требуется принять на веру следующие утверждения:

- Чтобы результаты были убедительны, необходимо, чтобы значение выражения $1,96 \times \sqrt{\frac{p \times (1-p)}{n}}$ было меньше, чем $0,1p$.
- Хотя анкетирование ещё не проведено и утверждать это нельзя, но p непременно должно находиться ровно посередине между 0 и 1. То есть, непременно $p = 0,5$.

Это позволит нам получить следующий результат:

$$1,96 \times \sqrt{\frac{0,5 \times (1-0,5)}{n}} \leq 0,5 \times \frac{1}{10}$$

$$1,96 \times \sqrt{\frac{0,5 \times (1-0,5)}{n}} \leq 0,05$$

$$\frac{1,96}{0,05} \times \sqrt{\frac{0,5 \times (1-0,5)}{n}} \leq 1$$

$$\left(\frac{1,96}{0,05}\right)^2 \times \frac{0,5 \times (1-0,5)}{n} \leq 1^2$$

$$\left(\frac{1,96}{0,05}\right)^2 \times 0,5 \times (1-0,5) \leq n$$

$$384,2 \leq n$$

Это значение 384,2 и является вышеупомянутым «магическим числом» — «примерно 400».

По поводу этого значения «примерно 400» необходимо помнить о четырёх моментах.

Во-первых, нельзя думать, что сбор данных примерно 400 человек является гарантией получения приемлемого результата.

Например, если спросить: «Как вы относитесь к 9-й статье конституции Японии» у 400 подписчиков газеты А, то может быть, вам удастся выяснить мнение подписчиков газеты А, но никак не мнение всего народа Японии.

Во-вторых, нельзя думать, что результаты анализа данных, взятых у менее чем 400 человек, не заслуживают доверия. Например, если вышеприведённый вопрос: «Как вы относитесь к 9-й статье конституции Японии», задать не 400 подписчикам газеты А, а 250 подписчикам газет А, В, С, D и Е, выбрав по 50 человек для каждой газеты, то результаты будут более убедительны.

В-третьих, вспомните процесс вывода этого значения «примерно 400». В анкете там был единственный вопрос: «поддерживаете ли вы кабинет министров», и только 2 варианта ответа: «поддерживаю», «не поддерживаю». Как вы, наверное, уже поняли, значение «примерно 400» соответствует случаю «одного вопроса с двумя вариантами ответа», но неприменимо для случая, например, «десяти вопросов с пятью вариантами ответа». Другими словами, это значение не универсально.

В-четвёртых, Вы, может быть, обратили внимание, что в рамке на предыдущей странице были некоторые значения ($1/10$ и $p=0,5$). На самом деле это значения, которые аналитик выбирает самостоятельно, они не являются чем-то абсолютным. Другими словами, исследователь может на своё усмотрение использовать $1/50$ вместо $1/10$ или $p = 0,273$

вместо $p = 0,5$. Поэтому, в зависимости от выбора этих величин, мы можем получить значение, совершенно отличное от «примерно 400». Точнее говоря, мы его непременно получим.

4. МЕТОДЫ СЛУЧАЙНОГО ВЫБОРА И МЕТОДЫ ПРЕДНАМЕРЕННОГО ВЫБОРА

Все методы взятия выборок можно разделить на *методы случайного выбора* и *методы преднамеренного выбора*. При использовании методов случайного выбора стремятся, чтобы все экземпляры, образующие выборку, выбирались из генеральной совокупности с равной вероятностью. Описанные на стр. 19-29 методы:

- метод простого случайного выбора;
- метод расслоенного выбора;
- метод двухступенчатого выбора;
- метод расслоенного двухступенчатого выбора

относятся к методам случайного выбора.

При использовании методов преднамеренного выбора нет необходимости обеспечивать равную вероятность выбора экземпляров из генеральной совокупности.

Примеры методов преднамеренного выбора приведены ниже в таблице.

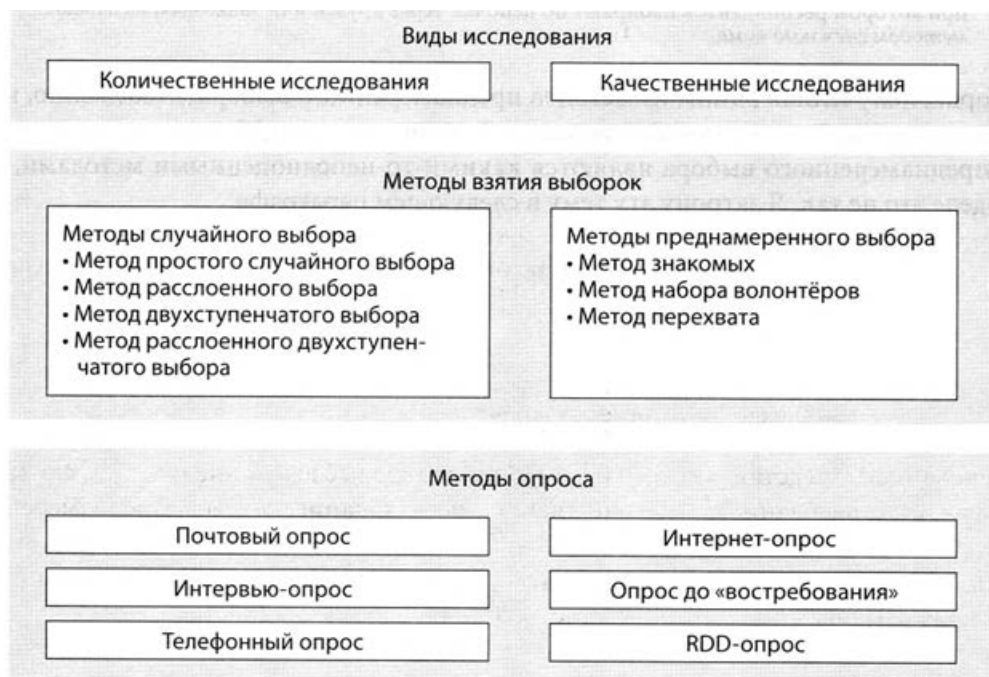
Метод знакомых ⁷	Выборка формируется из людей, готовых оказать содействие исследованию: знакомых, друзей и т.д.
Метод набора	Выборка формируется из людей, откликнувшихся на специальное приглашение, помещённое, например, в рекламной открытке книги.
Метод перехвата	Выборка формируется из людей, встреченных на улице и согласившихся принять участие в исследовании.

* *метод знакомых* также называется *методом связей* или *методом удобного случая*. Метод, при котором респондентов набирают по цепочке через друзей или знакомых, называют *методом снежного кома*.

Выборка, полученная одним из методов преднамеренного выбора, несомненно, никогда не будет «точной миниатюрой генеральной совокупности». Можно подумать, что методы преднамеренного выбора являются какими-то неполноценными методами, но на самом деле это не так. Я затрону эту тему в следующем параграфе.

5. КОЛИЧЕСТВЕННОЕ ИССЛЕДОВАНИЕ И КАЧЕСТВЕННОЕ ИССЛЕДОВАНИЕ

До этого мы обсуждали различные особенности методов взятия выборок и методов опроса, но прежде всего необходимо усвоить, что все исследования делятся на *количественные исследования* и *качественные исследования*.



■ Рис.1.1. Виды исследования, методы взятия выборок и методы опроса

При *количественном исследовании* объект изучается на основании «данных анкетирования», «статистических материалов государственных служб» и т.п. За исключением данной главы, в данной книге рассматриваются только такие исследования. При *качественном исследовании* объектом сбора данных является небольшая группа людей, по существу это интервью.

Количественное исследование даёт следующие преимущества:

- возможность получения объективных результатов;
- простота обобщения результатов;
- высокая воспроизводимость.

Недостатком количественного исследования является невозможность получения от каждого респондента информации глубокого содержания. Другими словами, возможно получения только информации о примерном положении дел в исследуемой группе. Качественное исследование в этом смысле обладает прямо противоположными свойствами. Так как это — сбор данных от небольшой группы людей, возможно получение информации глубокого содержания от каждого опрошиваемого респондента. Однако качественное исследование имеет следующие недостатки:

- трудность получения объективных результатов;
- трудность обобщения результатов;
- невысокая воспроизводимость.

На первый взгляд может показаться, что качественное исследование — не очень хороший метод. Однако это не так. Приведу доступный пример. Представьте, что вы — со-грудник компании, которая собирается провести в отношении своих товаров одно из следующих исследований.

Исследование 1	Анкетирование 1000 человек, выбранных случайным образом, то есть среди них многие совершенно безразличны к товарам нашей компании, на предмет отношения к нашим товарам.
Исследование 2	Проведение круглого стола с участием 10 человек, проявляющих горячий интерес к товарам нашей компании, с целью досконального выявления их мнения по поводу наших товаров.

Исследование 1 представляет собой «количественное исследование методом случайного выбора», а исследование 2 — «качественное исследование методом преднамеренного выбора». Чтобы дать однозначную оценку, нужно, конечно, учитывать также цели исследования, однако можете ли вы сказать, что исследование 1 является самым лучшим вариантом, а исследование 2 проводить не стоит?

Кроме того, количественное исследование можно проводить не только в отношении выборок, полученных методами случайного выбора, но также и в отношении выборок, полученных методами преднамеренного выбора. Хотя качественное исследование в подавляющем большинстве случаев проводится в отношении выборок, полученных методами преднамеренного выбора, ничто не запрещает проводить его в отношении выборок, полученных методом случайного выбора.

6. ПОДХОДЫ К АНАЛИЗУ ДАННЫХ

В данном параграфе я расскажу о важном моменте, который, однако, не связан с предыдущим содержанием.

Дело в том, что существует два подхода к анализу данных: *поисковый подход* и *подтверждающий подход*.

Процесс анализа данных при использовании поискового подхода

1. Под рукой оказались данные.
2. По принципу «попытка не пытка», попробовали применить к этим данным различные методы анализа.
3. Поняли «задним умом»: «Похоже, дела в этой сфере жизни обстоят так».
4. Опубликовали результаты анализа.

Процесс анализа данных при использовании подтверждающего подхода

1. Построили гипотезу.
2. С целью проверки гипотезы собрали данные и проанализировали их.
3. Установили верность или ошибочность гипотезы.
4. Опубликовали результаты анализа.

Преимущество поискового подхода заключается в возможности быстрой и лёгкой реализации, при условии наличия соответствующих данных, а недостаток — в возможности с помощью «заднего ума» получить произвольный результат: можно обрабатывать данные так, как вздумается, насильно притягивать связи между переменными и т.д. Один неверный шаг — и можно получить всё, что угодно. Поэтому существует вероятность, что окружающие не сочтут убедительными полученные вами результаты, когда вы их опубликуете.

Подтверждающий подход требует первоначального построения гипотезы, поэтому на практике он сложнее. С другой стороны, так как сбор и анализ данных проводятся после построения гипотезы, в случае её подтверждения результаты анализа будут выглядеть для публики очень убедительно. Даже если гипотеза не подтвердится, вы, по крайней мере, сможете установить тот факт, что данная гипотеза неверна, тем самым внося вклад в выбор направления дальнейших исследований.

2

БЛАНКИ АНКЕТ И ВОПРОСЫ

1. СТРУКТУРА БЛАНКА АНКЕТЫ
2. ТИПЫ ВОПРОСОВ
3. КАКИХ ВОПРОСОВ НУЖНО ИЗБЕГАТЬ
4. КАКИХ ВОПРОСОВ НУЖНО ИЗБЕГАТЬ (ПРОДОЛЖЕНИЕ)
5. НАЛИЧИЕ НЕЙТРАЛЬНОГО ВАРИАНТА ОТВЕТА





ПЛОП

Я ПОШЛА!



КАК, УЖЕ?

Ням-ням!



АА, НАША ДОЧЬ ВЫГЛЯДИТ БОДРО, НЕСМОТРЯ НА ТО, ЧТО НЕ МОЖЕТ ПОКА ВСТРЕЧАТЬСЯ С ЯМАМОТО.

ХВАТИТ ГОВОРИТЬ О НЁМ!

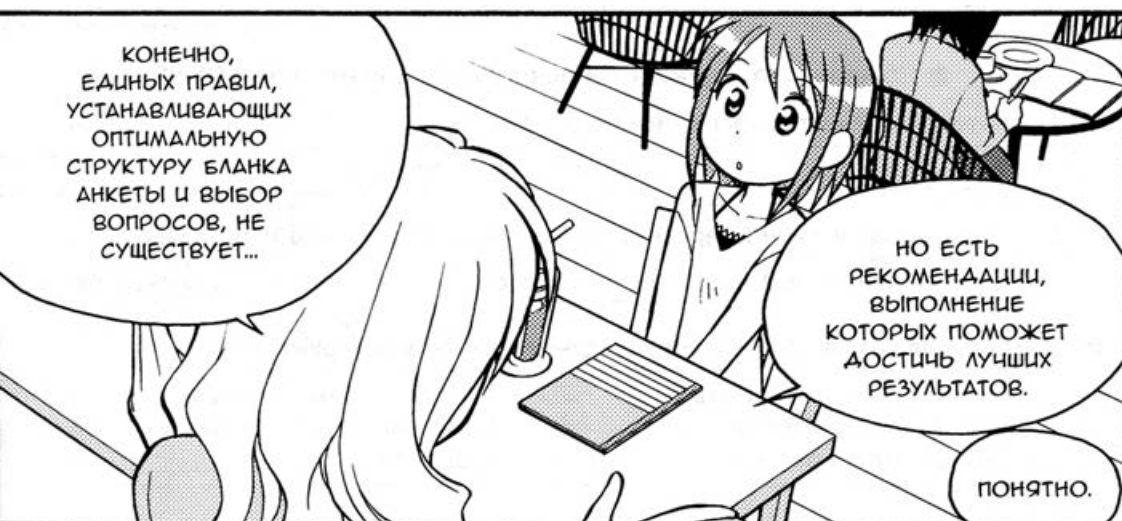


РУИ ЕЩЁ СТУДЕНТКА, ЕЙ НАДО КАК СЛЕДУЕТ ЗАНИМАТЬСЯ!

АА, АА, ТЫ ПРАВ.

Лучше бы ты скорее переодел пижаму.





1. СТРУКТУРА БЛАНКА АНКЕТЫ



АНКЕТА ДЛЯ ПОСЕТИТЕЛЕЙ

■ Пожалуйста, расскажите о себе

B1. Ваш пол (только 1 кружочек).

1. муж.

2. жен.

B2. Ваш возраст.

лет

B3. Ваш род занятий (только 1 кружочек).

1. Служащий компании

2. Предприниматель

3. Студент

4. Домохозяйка

5. Другое: _____

■ Пожалуйста, оцените Ваше впечатление от кафе NORNIS

B4. Оцените наше качество обслуживания (только 1 кружочек).

1. Очень плохое 2. Плохое 3. Среднее 4. Хорошее 5. Очень хорошее

B5. Оцените вкус наших кондитерских изделий и напитков (только 1 кружочек).

1. Очень невкусные 2. Невкусные 3. Неоднозначно 4. Вкусные 5. Очень вкусные

B6. Что побудило Вас посетить наше заведение (любое число кружочков)?

1. Реклама в газетах/журналах

2. Наша интернет-страница

3. Рассказ друга/знакомого

4. Просто проходил мимо

5. Понравилось наше здание

6. Другое: _____

Благодарим Вас за содействие!



АНКЕТА ДЛЯ ПОСЕТИТЕЛЕЙ

■ Пожалуйста, оцените Ваше впечатление от кафе NORNIS

B1. Оцените наше качество обслуживания (только 1 кружочек).

1. Очень плохое 2. Плохое 3. Среднее 4. Хорошее 5. Очень хорошее

B2. Оцените вкус наших кондитерских изделий и напитков (только 1 кружочек).

1. Очень невкусные 2. Невкусные 3. Неоднозначно 4. Вкусные 5. Очень вкусные

B3. Что побудило Вас посетить наше заведение (любое число кружочков)?

1. Реклама в газетах/журналах
2. Наша интернет-страница
3. Рассказ друга/знакомого
4. Просто проходил мимо
5. Понравилось наше здание
6. Другое: _____

■ Пожалуйста, расскажите о себе

B4. Ваш пол (только 1 кружочек).

1. муж. 2. жен.

B5. Ваш возраст.

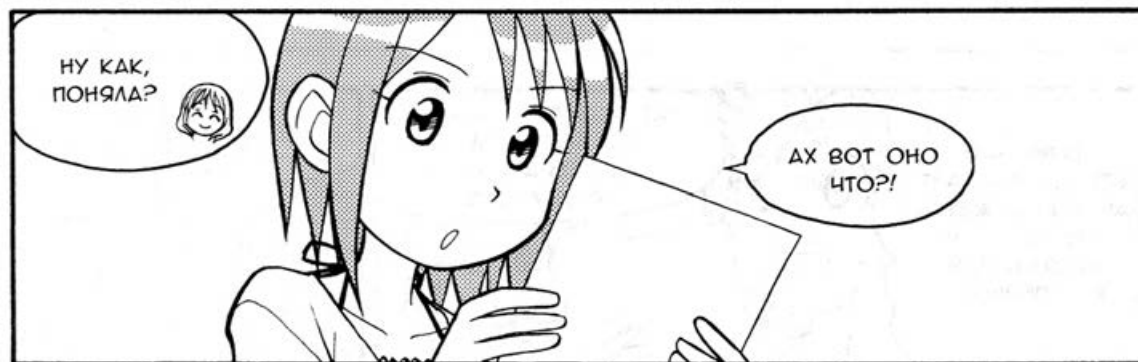
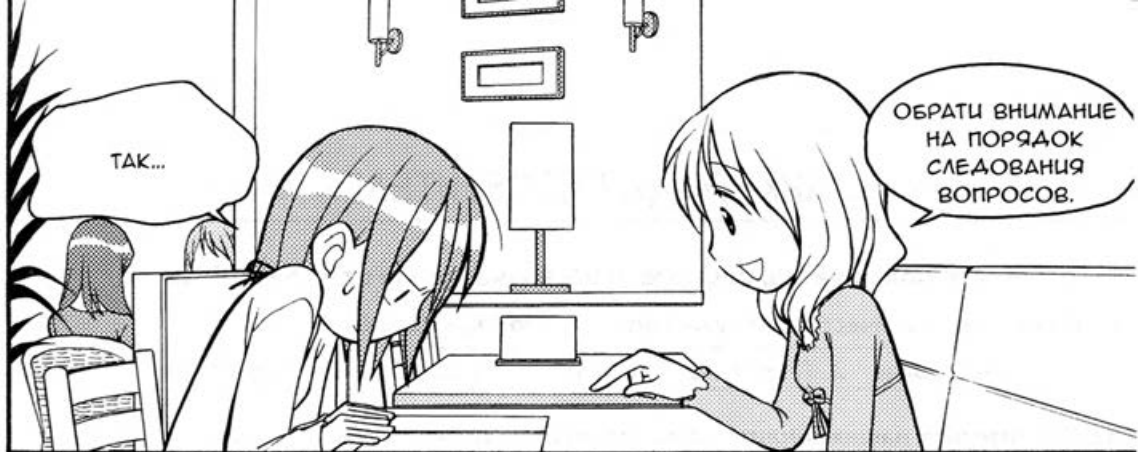
лет

B6. Ваш род занятий (только 1 кружочек).

1. Служащий компании 2. Предприниматель 3. Студент
4. Домохозяйка 5. Другое: _____

Благодарим Вас за содействие!





ВОПРОСЫ О СИТУАЦИИ

Приходилось ли Вам делать что-то, есть ли у Вас что-то, когда Вы купили что-то, где Вы купили что-то и т.п.



ВОПРОСЫ ОБ ОТНОШЕНИИ

Что Вам понравилось или не понравилось, почему Вам это понравилось или не понравилось, что Вы думаете об этом и т.п.



ВОПРОСЫ О ПРИЗНАКАХ

Пол, возраст, семейное положение, размер дохода, полученное образование, состав семьи и т.п.

БЕЗОПАСНЫМ БЛАНКОМ АНКЕТЫ ЯВЛЯЕТСЯ ТАКОЙ, В КОТОРОМ СНАЧАЛА РЕСПОНДЕНТА СПРАШИВАЮТ О "СИТУАЦИИ", ТО ЕСТЬ О ЕГО ДЕЙСТВИЯХ, ОПЫТЕ И Т.П.,

ЗАТЕМ СЛЕДУЮТ ВОПРОСЫ ОБ "ОТНОШЕНИИ К СИТУАЦИИ", ТО ЕСТЬ ОБ ОЩУЩЕНИЯХ И МЫСЛЯХ РЕСПОНДЕНТА,

А ВОПРОСЫ О "ПРИЗНАКАХ" РЕСПОНДЕНТА ЗАДАЮТСЯ В САМОМ КОНЦЕ АНКЕТЫ.



НАВЕРНОЕ, ЭТОТ БЛАНК АНКЕТЫ КАФЕ СОСТАВЛЯЛ КАКОЙ-НИБУДЬ НОВИЧОК.

Вроде меня.

Хе-хе-хе

НЕТ, ЭТО БЛАНК АНКЕТЫ ДЛЯ КАФЕ NORNS, КОТОРЫЙ КОГДА-ТО ДАВНО СОСТАВИЛА Я.

КХЕ...

2. ТИПЫ ВОПРОСОВ

ПРОСТИ, ЭТО ВЕДЬ
ЗА ТВОЙ СЧЁТ.

НИЧЕГО,
НИЧЕГО!
СЕГОДНЯ
УГОЩАЮ Я!

ТЕПЕРЬ
Я РАССКАЖУ
О ВОПРОСАХ,
КОТОРЫЕ
ПОМЕЩАЮТ
В БЛАНК
АНКЕТЫ.

*Хорошо,
тогда
я отвечаю.*

Я ГОТОВА,
НАЧИНАЙ.

ВСЕ ВОПРОСЫ АНКЕТ
ДЕЛЯТСЯ НА ВОПРОСЫ
С ЕДИНИЧНЫМ ВЫБОРОМ,
С МНОЖЕСТВЕННЫМ ВЫБОРОМ,
С КОЛИЧЕСТВЕННЫМ ОТВЕТОМ
И С ЛИТЕРАЛЬНЫМ ОТВЕТОМ.

Хм...

ПРИВЕДУ
ТИПИЧНЫЕ
ПРИМЕРЫ
КАЖДОГО
ИЗ ЭТИХ
ТИПОВ.

Вопросы с единичным выбором

СНАЧАЛА — О ВОПРОСАХ С ЕДИНИЧНЫМ ВЫБОРОМ. ЭТО ВОПРОСЫ, В КОТОРЫХ МОЖНО ВЫБРАТЬ ТОЛЬКО ОДИН ВАРИАНТ ОТВЕТА.



В. Какой из нижеприведённых тортов Вы любите больше всего? (обвести только 1 вариант).

1. Песочный торт 2. Творожный торт 3. Шоколадный торт 4. Монблан

В. Насколько важно для Вас удаление от близлежащей железнодорожной станции при выборе квартиры для одиночного проживания? (обвести только 1 вариант).

1. Совсем не важно 2. Не важно 3. Неоднозначно 4. Важно 5. Очень важно

ЕСЛИ В АНКЕТЕ ЕСТЬ МНОГО ВОПРОСОВ С ОДИНАКОВЫМИ ВАРИАНТАМИ ОТВЕТОВ, ТО В ЦЕЛЯХ ЭКОНОМИИ БУМАГИ Я РЕКОМЕНДУЮ ОФОРМИТЬ БЛАНК АНКЕТЫ В ВИДЕ ТАБЛИЦЫ.



В. Насколько важны для Вас нижеследующие условия при выборе квартиры для одиночного проживания? (в каждой строке — только по 1 варианту)

	Совсем не важно	Не важно	Неоднозначно	Важно	Очень важно
а. Удаление от ближайшей железнодорожной станции	1	2	3	4	5
б. Освещённость солнцем	1	2	3	4	5
в. Наличие кладовых	1	2	3	4	5
г. Наличие мебели	1	2	3	4	5

Вопросы с множественным выбором

ТЕПЕРЬ — О ВОПРОСАХ С МНОЖЕСТВЕННЫМ ВЫБОРОМ. ЭТО ВОПРОСЫ, В КОТОРЫХ МОЖНО ВЫБРАТЬ СРАЗУ НЕСКОЛЬКО ВАРИАНТОВ ОТВЕТА.



В. Какие условия важны для Вас при выборе квартиры для одиночного проживания? (можно поставить от 1 до 5 кружочков).

- | | |
|--|--|
| 1. Удаление от железнодорожной станции | 4. Наличие мебели |
| 2. Освещённость солнцем | 5. Наличие общественных заведений поблизости |
| 3. Наличие кладовых | |

СЛЕДУЮЩАЯ ФОРМУЛИРОВКА ВОПРОСА ДОПУСТИМА, НО ИСПОЛЬЗОВАТЬ ЕЁ Я НЕ РЕКОМЕНДУЮ.



В. Какие условия важны для Вас при выборе квартиры для одиночного проживания? (поставьте не более 2 кружочков).

- | | |
|--|--|
| 1. Удаление от железнодорожной станции | 4. Наличие мебели |
| 2. Освещённость солнцем | 5. Наличие общественных заведений поблизости |
| 3. Наличие кладовых | |

Внимание!

ПОЧЕМУ?

В ОТЛИЧИЕ ОТ СЛУЧАЯ "ОТ 1 ДО 5 КРУЖОЧКОВ", В ДАННОМ СЛУЧАЕ РЕСПОНДЕНТ ДОЛЖЕН БУДЕТ ПЕРЕД ОТВЕТОМ ПРОБЕЖАТЬ ГЛАЗАМИ ВСЕ ВАРИАНТЫ, А ЗНАЧИТ НАГРУЗКА НА НЕГО УВЕЛИЧИТСЯ.



Вопросы с количественным ответом

ТЕПЕРЬ — О ВОПРОСАХ С КОЛИЧЕСТВЕННЫМ ОТВЕТОМ. ЭТО ВОПРОСЫ, КОТОРЫЕ ТРЕБУЮТ ОТВЕТА В ВИДЕ ТОЧНОГО ЧИСЛОВОГО ЗНАЧЕНИЯ. ЧТОБЫ ИЗБЕЖАТЬ ОШИБОК ПРИ ЗАПОЛНЕНИИ, РЕКОМЕНДУЕТСЯ ВСТАВИТЬ В БЛАНК АНКЕТЫ СЕТКУ С РАЗРЯДАМИ.



В. Сколько денег дают Вам родители в месяц на карманные расходы?

						йен
--	--	--	--	--	--	-----

Вопросы с литеральным ответом

И В ЗАКЛЮЧЕНИЕ — ВОПРОСЫ С ЛИТЕРАЛЬНЫМ ОТВЕТОМ. ЭТО ВОПРОСЫ, НА КОТОРЫЕ ОТВЕЧАЮТ СВОБОДНЫМ ТЕКСТОМ, А НЕ ВЫБОРОМ ЗАРАНЕЕ ПОДГОТОВЛЕННЫХ ВАРИАНТОВ ОТВЕТА.



В. Напишите имя одного Вашего самого любимого артиста?

Графа для ответа

В. Если у Вас есть замечания или предложения к нашему заведению, напишите их ниже.

--

АА,

Я ПОНЯЛА!

3. КАКИХ ВОПРОСОВ НУЖНО ИЗБЕГАТЬ



АНКЕТА КАФЕ NORN'S

Каждому заполнившему анкету — бесплатный купон на кассе!

Пол	муж / жен	Возраст	лет
Род занятий		Годовой доход	десятков тыс. йен

В. Как бы Вы охарактеризовали атмосферу нашего кафе?

- Очень неприятная Неприятная Средняя Приятная Очень приятная

В. Как вы оцениваете одежду официанток и качество обслуживания?

- Очень плохие Плохие Средние Хорошие Очень хорошие

В. Как вы оцениваете наш чёрный чай?

- Очень невкусный Невкусный Средний Вкусный Очень вкусный

В. Каково Ваше мнение о наших ценах?

- Дешёвые Средние Дорогие

В. Проставьте цифры от 1 до 4 напротив каждого из видов чая в порядке убывания Ваших предпочтений.

Простой чай	место
Чай с лимоном	место
Чай с молоком	место
Чай с лепестками роз	место

В. Любите ли Вы бывать в кафе?

- Да Нет

Ой... ТЕПЕРЬ Я ПОНЯЛ СВОЮ ОШЕБКУ — МОЙ ПЕРВЫЙ ВОПРОС КАСАЛСЯ ПРИЗНАКОВ...

АА, О ПРИЗНАКАХ ЛУЧШЕ СПРАШИВАТЬ В КОНЦЕ АНКЕТЫ!



Возраст	/ лет / / / /
Годовой доход	десятков тыс. йен

КРОМЕ ТОГО, МНЕ НЕ ОЧЕНЬ НРАВИТСЯ ТАКАЯ ФОРМА ВОПРОСА О ГОДОВОМ ДОХОДЕ.

?

× Слишком конкретные вопросы о довольно личных обстоятельствах.

ДЕЙСТВИТЕЛЬНО, БЫВАЮТ СЛУЧАИ, КОГДА НЕОБХОДИМО УЗНАТЬ КОНКРЕТНЫЙ РАЗМЕР ДОХОДА...

...НО В ОСТАЛЬНЫХ СЛУЧАЯХ ЛУЧШЕ ЗАДАВАТЬ ЭТОТ ВОПРОС ВОТ В ТАКОЙ ФОРМЕ. ТОГДА РЕСПОНДЕНТУ БУДЕТ ЛЕГЧЕ ОТВЕТИТЬ НА НЕГО.

ПОНЯТНО.

В. Каков Ваш годовой доход? (только 1 кружок)

1. Менее 4 млн. йен
2. От 4 до 6 млн. йен
3. От 6 до 8 млн. йен
4. От 8 до 10 млн. йен
5. Более 10 млн. йен

А МНЕ СЕЙЧАС ВООООЩЕ КАЖЕТСЯ, ЧТО НЕ БЫЛО НИКАКОЙ НЕОБХОДИМОСТИ СПРАШИВАТЬ В ЭТОЙ АНКЕТЕ О ГОДОВОМ ДОХОДЕ.

АА, Я ТОЖЕ ТАК СЧИТАЮ.

В. Как бы Вы охарактеризовали атмосферу нашего кафе?

Очень неприятная Неприятная Средняя Приятная Очень приятная

В. Как вы оцениваете одежду официанток и качество обслуживания?

Очень плохие Плохие Средние Хорошие Очень хорошие

В. Как Вы оцениваете наш чёрный чай?

Очень невкусный Невкусный Средний Вкусный Очень вкусный

В. Каково Ваше мнение о наших ценах?

Дешёвые Средние Дорогие



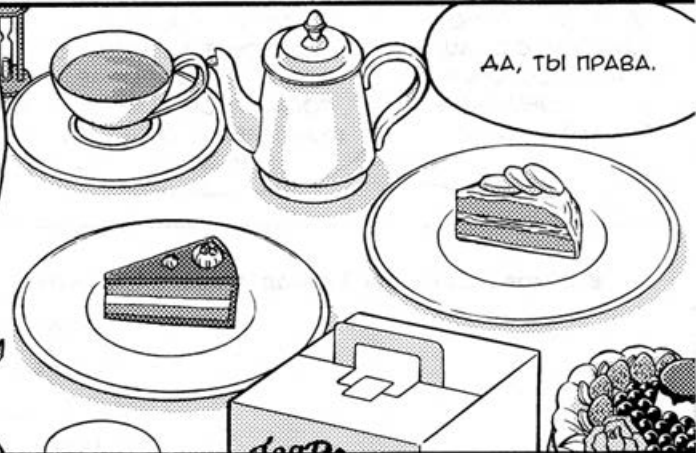
Расплывчато
сформулированные вопросы.

А ФОРМУЛИРОВКА
ЭТОГО ВОПРОСА МНЕ
КАЖЕТСЯ НЕМНОГО
РАСПЛЫВЧАТОЙ.



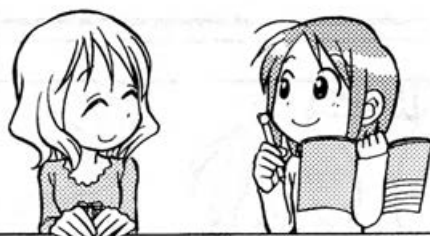
АХ..
ТОЧНО...

ТЫ СПРАШИВАЕШЬ: "КАКОВО
ВАШЕ МНЕНИЕ О НАШИХ
ЦЕНАХ?", НО РЕСПОНДЕНТУ
БУДЕТ НЕПОНЯТНО,
КАКИЕ ИМЕННО ЦЕНЫ
ИМЕЮТСЯ В ВИДУ:
ЦЕНЫ НА ЧЁРНЫЙ ЧАЙ ИЛИ
ЦЕНЫ НА КОНДИТЕРСКИЕ
ИЗДЕЛИЯ.



ДА, ТЫ ПРАВА.

ВОПРОСЫ ДОЛЖНЫ
БЫТЬ СФОРМУЛИРОВАНЫ
КАК МОЖНО
БОЛЕЕ КОНКРЕТНО.



ЯСНО!

В. Каково Ваше мнение о цене нашего чёрного чая (только 1 кружочек)

1. Дёшево

2. Средне

3. Дорого

В. Как вы оцениваете одежду официанток и качество обслуживания?

- Очень плохие Плохие Средние
 Хорошие Очень хорошие

× Неоднозначные вопросы.

ДАЛЕЕ, НЕ КАЖЕТСЯ ЛИ
ТЕБЕ ЭТОТ ВОПРОС
НЕОДНОЗНАЧНЫМ?



АХ!

НАПРИМЕР, РЕСПОНДЕНТЫ,
КОТОРЫМ ПОКАЖЕТСЯ, ЧТО
"ОБСЛУЖИВАНИЕ ХОРОШЕЕ,
НО ОДЕЖДА ОФИЦИАНТОК
ПЛОХАЯ", ЗАТРУДНЯТСЯ
С ОТВЕТОМ, НЕ ТАК ЛИ?

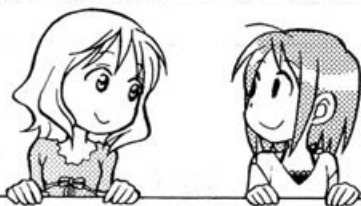
ДА,
НАВЕРНОЕ...

Что это ты
на себя
напялила?!



Ай!

КАЖАБЫЙ ВОПРОС
ЛУЧШЕ СФОРМУЛИРОВАТЬ
ОДНОЗНАЧНО.



ПОНЯТНО!

В. Как вы оцениваете одежду официанток? (только 1 кружочек)

1. Очень плохая 2. Плохая 3. Средняя 4. Хорошая 5. Очень хорошая

В. Как вы оцениваете качество обслуживания? (только 1 кружочек)

1. Очень плохое 2. Плохое 3. Среднее 4. Хорошее 5. Очень хорошее

В. Проставьте цифры от 1 до 4 напротив каждого из видов чая в порядке убывания Ваших предпочтений.

Простой чай	место
Чай с лимоном	место
Чай с молоком	место
Чай с лепестками роз	место



Вопросы, требующие назначения иерархии.

А ЗДЕСЬ РЕСПОНДЕНТ ДОЛЖЕН НАЗНАЧИТЬ ИЕРАРХИЮ.

КАК?! ЭТО ТОЖЕ НЕЛЬЗЯ?!

НЕТ, Я НЕ ГОВОРЮ, ЧТО СОВСЕМ НЕЛЬЗЯ.

СУЩЕСТВУЕТ ДАЖЕ ТАКОЙ МЕТОД АНАЛИЗА, КАК **СОВМЕСТНЫЙ АНАЛИЗ (CONJOINT ANALYSIS)**, В ОСНОВЕ КОТОРОГО ЛЕЖИТ КАК РАЗ НАЗНАЧЕНИЕ ИЕРАРХИИ.

ТОГДА ЧТО В ЭТОМ ПЛОХОГО?

ДЕЛО В ТОМ, ЧТО НЕКОТОРЫМ РЕСПОНДЕНТАМ МОГУТ ОДИНАКОВО ПРАВИТЬСЯ ДВА ВИДА ЧАЯ, ПОЭТОМУ ИМ БУДЕТ ТРУДНО ВЫБРАТЬ, КАКОЙ ЧАЙ ПОСТАВИТЬ НА ПЕРВОЕ МЕСТО.

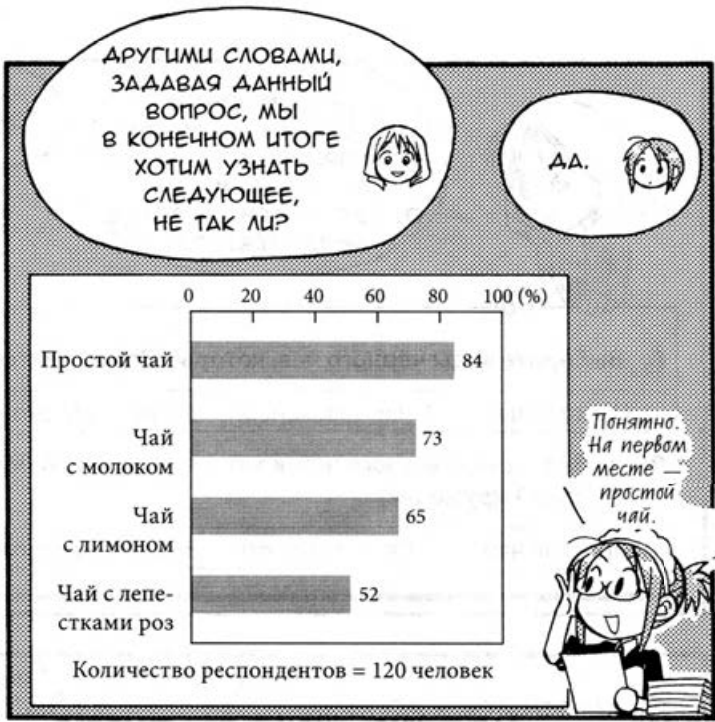
АХ, ВОТ В ЧЁМ ДЕЛО.

КРОМЕ ТОГО, ТЕ РЕСПОНДЕНТЫ, КОТОРЫЕ ДО ЭТОГО ДАЖЕ НЕ ЗНАЛИ О СУЩЕСТВОВАНИИ, НАПРИМЕР, "ЧАЯ С ЛЕПЕСТКАМИ РОЗ", ЗАТРУДНЯЮТСЯ С ПРИСВОЕНИЕМ ЕМУ МЕСТА.

Чай с лепестками роз?

ДА, ТЫ ПРАВА...

Да. На какое место нам бы его поставить?



ПОЭТОМУ НЕ ОБЯЗАТЕЛЬНО ТРЕБОВАТЬ ОТ РЕСПОНДЕНТА, ЧТОБЫ ОН НАЗНАЧИЛ ИЕРАРХИЮ.

ВОПРОС ЛУЧШЕ ПОСТАВИТЬ ВОТ ТАК:

В. Выберите виды чёрного чая, которые Вам нравятся (от 1-го до 4-х кружочков)

1. Простой чай 2. Чай с лимоном 3. Чай с молоком 4. Чай с лепестками роз

	Простой чай	Чай с лимоном	Чай с молоком	Чай с лепестками роз
Респондент А.	1	0	1	0
Респондент В.	1	1	0	0
Респондент С.	1	1	0	0
	:	:	:	:
	↓	↓	↓	↓
Доля (%)	84	65	73	52

ПРОСУММИРОВАВ ОТВЕТЫ, МЫ ПОЛУЧИМ ТО, ЧТО ХОТИМ, НЕ ТАК ЛИ?

ТОЧНО!

ЕСЛИ ЖЕ ТЫ ВСЁ-ТАКИ ХОЧЕШЬ, ЧТОБЫ РЕСПОНДЕНТЫ УКАЗЫВАЛИ ИЕРАРХИЮ ПРЕАПОЧТЕНИЙ...

...ТО
МОЖНО ПОСТУПИТЬ
СЛЕДУЮЩИМ ОБРАЗОМ.
ХОТЯ ПРИСВАИВАТЬ МЕСТА
ВСЕМ ВИДАМ ЧАЯ
ОТ РЕСПОНДЕНТОВ ЗАДЕСЬ
НЕ ПОТРЕБУЕТСЯ.

ПОНЯТНО.

В. Выберите виды чёрного чая, которые Вам нравятся (от 1-го до 4-х кружочков)

1. Простой чай 2. Чай с лимоном 3. Чай с молоком 4. Чай с лепестками роз

В. Выберите вид чёрного чая, который нравится Вам больше всего (только 1 кружочек)

1. Простой чай 2. Чай с лимоном 3. Чай с молоком 4. Чай с лепестками роз

И НАПОСЛЕДОК Я МОГЛА БЫ
ПОСОВЕТОВАТЬ ТЕБЕ
ПРОНУМЕРОВАТЬ ВОПРОСЫ
И ВАРИАНТЫ ОТВЕТОВ,
А ТАКЖЕ ДОПОЛНИТЬ ТЕКСТЫ
ВОПРОСОВ ПОЯСНЕНИЯМИ ТИПА
"ТОЛЬКО 1 КРУЖОЧЕК"...

Ведь тогда респонденту будет легче
отвечать, не так ли?

Пол муж.
Род зан-й неприятна

1
В. Как бы Вы охарактеризовали атмосферу нашей компании? неоднозначно

В. Как вы оцениваете работу менеджера? (только 1 кружочек)

Как вы оцениваете работу менеджера? (только 1 кружочек)

ТЕПЕРЬ МНЕ
ВСЁ ЯСНО!

АА, НЕ ПРОСТО ВСЁ
ЭТО — ТАК МНОГО
МОМЕНТОВ,
НА КОТОРЫЕ
НАДО ОБРАЩАТЬ
ВНИМАНИЕ.

Уфф...

НЕТ, ТЫ
ПРЕУВЕЛИЧИВАЕШЬ.

НА САМОМ ДЕЛЕ ВСЁ ОЧЕНЬ
ПРОСТО — НАДО ТОЛЬКО
ПОСМОТРЕТЬ НА АНКЕТУ
"ГЛАЗАМИ РЕСПОНДЕНТА!"

ГЛАЗАМИ
РЕСПОН-
ДЕНТА...

АА!



МНЕ КАЖЕТСЯ,
ЧТО НЕОБХОДИМО
РАЗМЫШЛЯТЬ ВОТ ТАК:

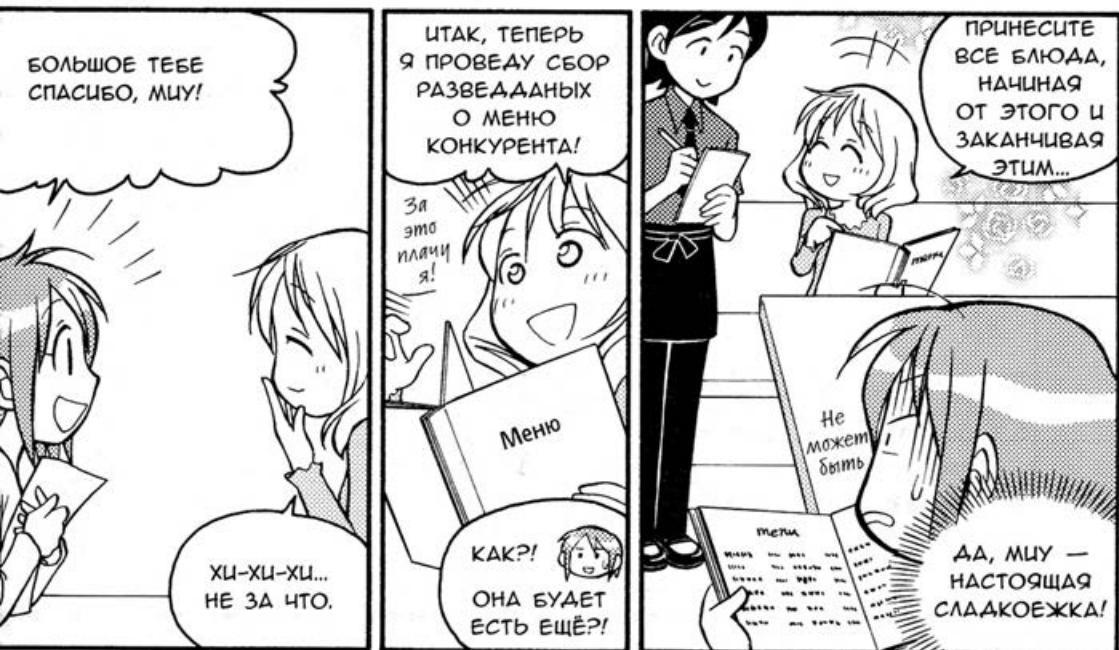
КАКИЕ ЧУВСТВА ВЫЗВАЛ БЫ
У МЕНЯ ЭТОТ БЛАНК АНКЕТЫ,
ЕСЛИ БЫ Я САМА БЫЛА
РЕСПОНДЕНТОМ?

СОГЛАСЯТСЯ ЛИ МОИ
БЛИЗКИЕ И ДРУЗЬЯ
ЗАПОЛНИТЬ ТАКУЮ
АНКЕТУ?



АХ, ВОТ ОНО
ЧТО!

КАЖЕТСЯ, ТЕПЕРЬ
Я ДЕЙСТВИТЕЛЬНО
ВСЁ ПОНЯЛА!



БОЛЬШОЕ ТЕБЕ
СПАСИБО, МУУ!

ИТАК, ТЕПЕРЬ
Я ПРОВЕДУ СБОР
РАЗВЕДАТЕЛЬНЫХ
О МЕНЮ
КОНКУРЕНТА!

За
это
плачу
я!

Меню

КАК?!
ОНА БУДЕТ
ЕСТЬ ЕЩЁ?!

ПРИНЕСИТЕ
ВСЕ БЛЮДА,
НАЧИНАЯ
ОТ ЭТОГО И
ЗАКАНЧИВАЯ
ЭТИМ...

Не
может
быть

АА, МУУ —
НАСТОЯЩАЯ
СЛАДКОЕЖКА!

4. КАКИХ ВОПРОСОВ НУЖНО ИЗБЕГАТЬ (ПРОДОЛЖЕНИЕ)

Выше были описаны 4 типа вопросов, которые следует избегать:

- конкретные вопросы, касающиеся довольно личных обстоятельств;
- расплывчато сформулированные вопросы;
- неоднозначные вопросы;
- вопросы, требующие присвоения рангов.

Однако существуют и другие типы вопросов, которые лучше не задавать. Приведу здесь несколько из них.

■ ПРОВОКАЦИОННЫЕ ВОПРОСЫ

В. Считается, что важность научно-технического образования для Японии в 21 веке будет расти, так как Япония не обладает большими природными ресурсами. В этой связи, что Вы думаете об уровне преподавания естественно-научных дисциплин в средней школе (только 1 кружочек).

1. Лучше повысить

2. Можно оставить прежним

Большинство людей, скорее всего, выберет первый вариант ответа.

■ ВОПРОСЫ СО СЛИШКОМ БОЛЬШИМ ЧИСЛОМ ГРАДАЦИЙ ОТВЕТОВ

В. Какое значение Вы придаёте нижеследующим качествам компаний при выборе места работы (в каждой строке — только по 1 кружку).

	В высшей степени не важно	Совершенно не важно	Не важно	Не очень важно	Нельзя сказать определёнno	В некоторой степени важно	Важно	Очень важно	В высшей степени важно
а. Известность компании	1	2	3	4	5	6	7	8	9
б. Хорошая система обучения новых сотрудников	1	2	3	4	5	6	7	8	9
в. Активно поручает работу молодым сотрудникам	1	2	3	4	5	6	7	8	9
:	:	:	:	:	:	:	:	:	:

Скорее всего, составлявший эту анкету исследователь решил, что 5 градаций ответов недостаточно для выражения всех нюансов психологии респондента. Я могу понять этого исследователя и считаю, что вполне допустимо поместить в анкету несколько вопросов с большим числом градаций ответов. Однако если таких вопросов будет слишком много, то респондент в процессе заполнения устанет и подумает, например: «Да какая разница, что выбрать: 7 или 8!»

Осмелюсь здесь высказать своё личное мнение, не имеющее под собой научной основы: градаций ответов должно быть не более 7.

■ ВОПРОСЫ, ТРЕБУЮЩИЕ ПРИСВОЕНИЯ ОЦЕНКИ В БАЛЛАХ

В. Оцените вкус этого печенья в баллах от 0 до 10.

баллов

Без добавления графической шкалы, подобной нижеприведённой, респонденту будет трудно представить себе интервалы между баллами, что вызовет затруднения при ответе на вопрос.



Кроме того, можно обратить внимание, что в действительности непонятно, как расценивать значение «0 баллов»: то ли это «невкусно», то ли «нельзя сказать определённо», что также вызовет у респондента затруднения при ответе.

■ ВОПРОСЫ, ТРЕБУЮЩИЕ ОТВЕТ В СВОБОДНОЙ ФОРМЕ

В. Назовите одного Вашего самого любимого артиста.

Графа для ответа

Вопросы, требующие ответа в свободной форме, обладают следующими преимуществами.

- Дают респонденту возможность ответить свободно в буквальном смысле.
- Дают исследователю возможность получить неожиданный и интересный ответ.

На первый взгляд кажется, что у этой формы вопроса одни преимущества, но в действительности она не так уж замечательна.

Попробуйте поставить себя на место респондента. Сможете ли вы сразу ответить на вопрос: «Кто Ваш любимый артист?», если вы, конечно, не ученик начальной школы. Другими словами, сможете ли вы сразу ответить на вопрос по теме, к которой у вас нет особого интереса. Теперь представьте, что вы исследователь. В случае почтового опроса, именно вам придётся вводить в таблицу ответы в свободной форме.

Осознаёте ли вы, что за возможность получить разнообразные ответы вам придётся заплатить довольно долгим и кропотливым трудом?

Я считаю, вопросы, требующие свободной формы ответа, лучше использовать для предварительного опроса, а затем провести основной опрос, используя в качестве вариантов ответа «пятерку лучших», например, следующим образом:

В. Назовите одного Вашего самого любимого артиста (только 1 кружочек).

1. Иносэ Юкито 2. Мацуяма Арису 3. Яно Аири 4. Намики Кадзуро 4. Хиракава Сакура

5. НАЛИЧИЕ НЕЙТРАЛЬНОГО ВАРИАНТА ОТВЕТА

Варианты ответа с градиентной оценкой могут включать или не включать нейтральный (неопределённый) вариант ответа, то есть что-то вроде «нельзя сказать однозначно».

■ ВАРИАНТЫ ОТВЕТА ВКЛЮЧАЮТ НЕЙТРАЛЬНЫЙ ВАРИАНТ

	Не важно	Не очень важно	Нельзя сказать однозначно	В некоторой степени важно	Важно
а. Удаление от железнодорожной станции	1	2	3	4	5
б. Освещённость солнцем	1	2	3	4	5

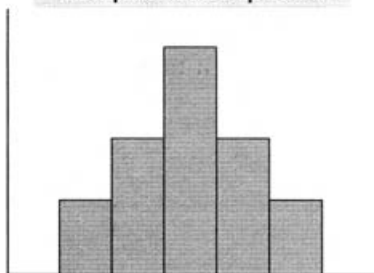
ВАРИАНТЫ ОТВЕТА НЕ ВКЛЮЧАЮТ НЕЙТРАЛЬНЫЙ ВАРИАНТ

	Не важно	Не очень важно	В некоторой степени важно	Важно
а. Удаление от железнодорожной станции	1	2	3	4
б. Освещённость солнцем	1	2	3	4

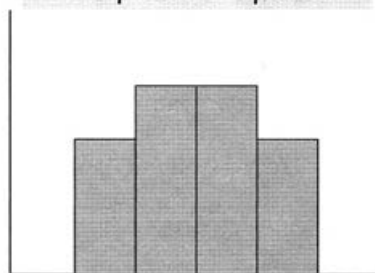
Включать или не включать нейтральный вариант — это дело исследователя. Однако отсутствие нейтрального варианта влечёт за собой:

- увеличение психологической нагрузки на респондента, который должен обязательно дать чёткий ответ: «да» или «нет»;
- уменьшение возможности того, что форма гистограммы будет близка к нормальному распределению, по сравнению со случаем присутствия нейтрального варианта.

Гистограмма при наличии нейтрального варианта



Гистограмма при отсутствии нейтрального варианта



Если вы не знаете, стоит или нет включать нейтральный вариант, то я рекомендую попробовать провести подготовительный опрос, например, своих друзей или знакомых, и принять решение на основе его результатов.

3

БАЗОВЫЕ ЗНАНИЯ ПО МАТЕМАТИКЕ

1. МАТРИЦЫ КОРРЕЛЯЦИИ
2. ЕДИНИЧНЫЕ МАТРИЦЫ
3. ВРАЩЕНИЕ
4. СОБСТВЕННЫЕ ЗНАЧЕНИЯ И СОБСТВЕННЫЕ ВЕКТОРЫ
5. СИММЕТРИЧНЫЕ МАТРИЦЫ
6. ДОПОЛНИТЕЛЬНЫЕ СВЕДЕНИЯ О МАТРИЦАХ
7. СУММА КВАДРАТОВ ОТКЛОНЕНИЙ, ДИСПЕРСИЯ
СЛУЧАЙНОЙ ВЕЛИЧИНЫ, КВАДРАТИЧНОЕ ОТКЛОНЕНИЕ



ДИНЬ-ДОН

ЗАРАВСТВУЙТЕ

А, ЭТО ВЫ!
ПРОХОДИТЕ,
ПОЖАЛУЙСТА.

ХЕ-ХЕ...
НАКОНЕЦ-ТО
НАСТАЛ ДЕНЬ
УЧЁБЫ!

ХИ-ХИ...
НУ ТЫ ДАЁШЬ,
МОЯ МИЛАЯ
РУЧЬ...

ИТАК, ДАВАЙ ПОСВЯТИМ
СЕГОДНЯШНЕЕ ЗАНЯТИЕ
БАЗОВЫМ
МАТЕМАТИЧЕСКИМ
ЗНАНИЯМ, НЕОБХОДИМЫМ
ДЛЯ ПОНИМАНИЯ
ФАКТОРНОГО АНАЛИЗА.

ДА, ПОЖАЛУЙСТА!

ТАК...
СЕГОДНЯ Я НЕ БУДУ
ПОДРОБНО ОБЪЯСНЯТЬ
ПРОЦЕССЫ ВЫЧИСЛЕНИЙ
И ПРИВОДИТЬ
ДОКАЗАТЕЛЬСТВА
УТВЕРЖДЕНИЙ.
Я СОБИРАЮСЬ ПРОСТО
ОБЪЯСНИТЬ ТЕБЕ
ОСНОВНЫЕ
ОПРЕДЕЛЕНИЯ.

ДУМАЮ, ЧТО ДЛЯ НАЧАЛА
ТЕБЕ НУЖНО ПРОСТО
ПОЛУЧИТЬ ПРЕДСТАВЛЕНИЕ
О ПРЕДМЕРЕ.

ХОРОШО!

1. МАТРИЦЫ КОРРЕЛЯЦИИ

НАЧНУ Я С МАТРИЦ КОРРЕЛЯЦИИ.

МАТРИЦЫ КОРРЕЛЯЦИИ СОСТОЯТ ИЗ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ.

	Родной язык	Обществоведение	Естествознание	Иностран. язык	Математика
Ученик 1	92	83	77	156	38
Ученик 2	97	82	68	114	33
Ученик 3	100	100	93	176	44
Ученик 4	89	77	100	158	46
Ученик 5	95	79	75	140	37
Ученик 6	99	96	84	174	42
Ученик 7	97	87	98	190	49
Ученик 8	93	77	73	132	35
Ученик 9	89	75	72	132	35
Ученик 10	98	93	70	186	37

НАПРИМЕР, ДЛЯ ЭТИХ ДАННЫХ МАТРИЦА КОРРЕЛЯЦИИ...

Коэффициенты корреляции

	Р	О	Е	И	М
Р	r_{RR}	r_{RO}	r_{RE}	r_{RI}	r_{RM}
О	r_{OR}	r_{OO}	r_{OE}	r_{OI}	r_{OM}
Е	r_{ER}	r_{EO}	r_{EE}	r_{EI}	r_{EM}
И	r_{IR}	r_{IO}	r_{IE}	r_{II}	r_{IM}
М	r_{MR}	r_{MO}	r_{ME}	r_{MI}	r_{MM}

Р – родной язык
 О – обществоведение
 Е – естествознание
 И – иностранный
 М – математика

...БУДЕТ ВЫГЛЯДЕТЬ ВОТ ТАК.

АГА, ПОНЯТНО.

$$\begin{pmatrix} r_{11} & r_{12} & r_{13} & r_{14} & r_{15} \\ r_{21} & r_{22} & r_{23} & r_{24} & r_{25} \\ r_{31} & r_{32} & r_{33} & r_{34} & r_{35} \\ r_{41} & r_{42} & r_{43} & r_{44} & r_{45} \\ r_{51} & r_{52} & r_{53} & r_{54} & r_{55} \end{pmatrix} = \begin{pmatrix} 1 & r_{12} & r_{13} & r_{14} & r_{15} \\ r_{21} & 1 & r_{23} & r_{24} & r_{25} \\ r_{31} & r_{32} & 1 & r_{34} & r_{35} \\ r_{41} & r_{42} & r_{43} & 1 & r_{45} \\ r_{51} & r_{52} & r_{53} & r_{54} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0.86 & 0.03 & 0.46 & 0.17 \\ 0.86 & 1 & 0.26 & 0.71 & 0.38 \\ 0.03 & 0.26 & 1 & 0.59 & 0.97 \\ 0.46 & 0.71 & 0.59 & 1 & 0.75 \\ 0.17 & 0.38 & 0.97 & 0.75 & 1 \end{pmatrix}$$

r_{ij} и r_{ji} равны

r_{ii} равны 1 при любых i

ВЫЧИСЛЕНИЯ ДАДУТ СЛЕДУЮЩЕЕ.

ЯСНО.

2. ЕДИНИЧНЫЕ МАТРИЦЫ



ТЕПЕРЬ —
О ЕДИНИЧНЫХ
МАТРИЦАХ.



ЕДИНИЧНЫЕ
МАТРИЦЫ
ПРЕДСТАВЛЯЮТ
ИЗ СЕБЯ
СЛЕДУ-
ЮЩЕЕ.

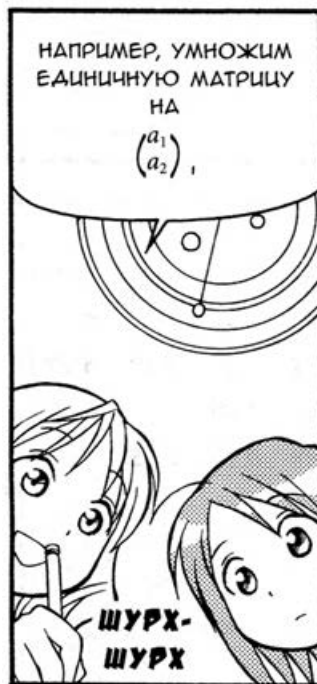
ХМ...
ВСЕ ЗНАЧЕНИЯ
НА ГЛАВНОЙ
ДИАГОНАЛИ
РАВНЫ ЕДИНИЦЕ,
А ОСТАЛЬНЫЕ
ЗНАЧЕНИЯ — НУЛИ.



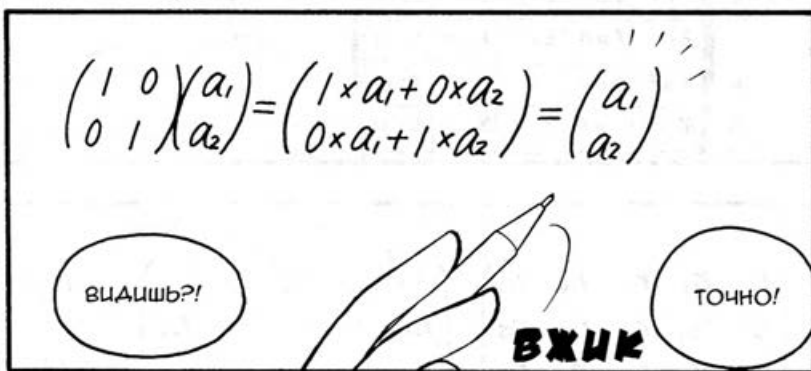
ПРИ ПЕРЕМНОЖЕНИИ
ЛЮБОЙ
ИСХОДНОЙ МАТРИЦЫ
И ЕДИНИЧНОЙ МАТРИЦЫ
ВИД ИСХОДНОЙ
МАТРИЦЫ
НЕ ИЗМЕНЯЕТСЯ.



ЧТО ЭТО
ЗНАЧИТ?



НАПРИМЕР, УМНОЖИМ
ЕДИНИЧНУЮ МАТРИЦУ
НА
 $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$,



ВДАИШЬ?!

ТОЧНО!

ВЖИК



ШУРХ-
ШУРХ



В ОБЩЕМ,
ЭТО МАТРИЦА
ВЕДЕТ СЕБЯ
КАК ЧИСЛО "1".

$$1 \times 7 = 7$$

$$1 \times A = A$$

ЯСНО.

ДАВАЙ РАССМОТРИМ ТАКЖЕ ДРУГИЕ
ПРИМЕРЫ ЕДИНИЧНЫХ МАТРИЦ.



$$\begin{matrix} 2 \times 2 & 2 \times 1 & & 2 \times 1 \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 1 \times a_1 + 0 \times a_2 \\ 0 \times a_1 + 1 \times a_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \end{matrix}$$

$$\begin{matrix} p \times p & p \times 1 & & p \times 1 \\ \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = \begin{pmatrix} 1 \times a_1 + 0 \times a_2 + \dots + 0 \times a_p \\ 0 \times a_1 + 1 \times a_2 + \dots + 0 \times a_p \\ \vdots \\ 0 \times a_1 + 0 \times a_2 + \dots + 1 \times a_p \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} \end{matrix}$$

$$\begin{matrix} 2 \times 2 & & 2 \times p & & 2 \times p \\ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \dots & a_{p1} \\ a_{12} & a_{22} & \dots & a_{p2} \end{pmatrix} = \begin{pmatrix} 1 \times a_{11} + 0 \times a_{12} & 1 \times a_{21} + 0 \times a_{22} & \dots & 1 \times a_{p1} + 0 \times a_{p2} \\ 0 \times a_{11} + 1 \times a_{12} & 0 \times a_{21} + 1 \times a_{22} & \dots & 0 \times a_{p1} + 1 \times a_{p2} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{p1} \\ a_{12} & a_{22} & \dots & a_{p2} \end{pmatrix} \end{matrix}$$

$$\begin{matrix} p \times 2 & 2 \times 2 & & p \times 2 \\ \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{p1} & a_{p2} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a_{11} \times 1 + a_{12} \times 0 & a_{11} \times 0 + a_{12} \times 1 \\ a_{21} \times 1 + a_{22} \times 0 & a_{21} \times 0 + a_{22} \times 1 \\ \vdots & \vdots \\ a_{p1} \times 1 + a_{p2} \times 0 & a_{p1} \times 0 + a_{p2} \times 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{p1} & a_{p2} \end{pmatrix} \end{matrix}$$



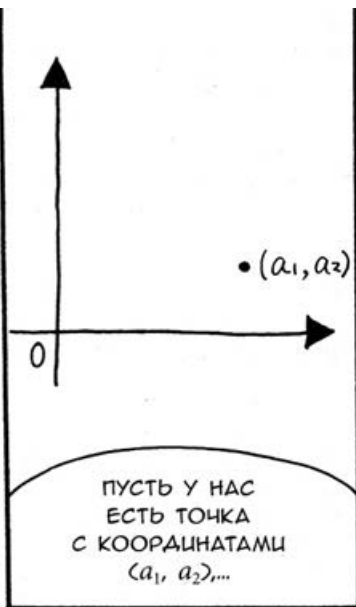
В САМОМ ДЕЛЕ, ЕДИНИЧНАЯ МАТРИЦА
НЕ ОКАЗЫВАЕТ НА ИСХОДНУЮ МАТРИЦУ
НИКАКОГО ВЛИЯНИЯ!

3. ВРАЩЕНИЕ

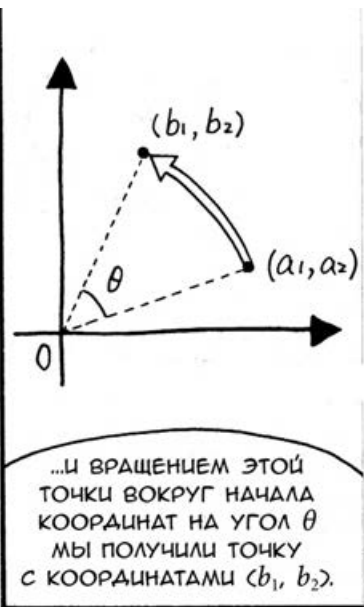
СЛЕДУЮЩАЯ ТЕМА —
ВРАЩЕНИЕ КООРДИНАТ.



ВРАЩЕНИЕ?

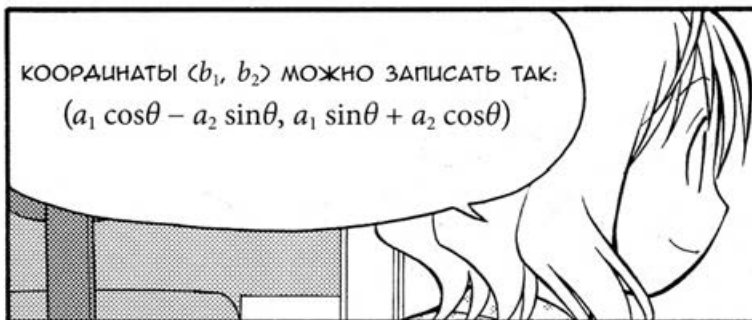


ПУСТЬ У НАС
ЕСТЬ ТОЧКА
С КООРДИНАТАМИ
 $(a_1, a_2), \dots$



...И ВРАЩЕНИЕМ ЭТОЙ
ТОЧКИ ВОКРУГ НАЧАЛА
КООРДИНАТ НА УГОЛ θ
МЫ ПОЛУЧИЛИ ТОЧКУ
С КООРДИНАТАМИ (b_1, b_2) .

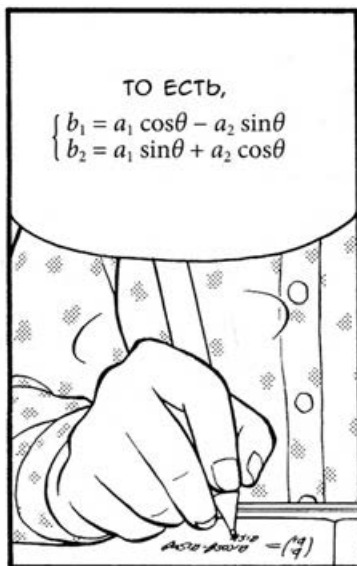
КООРДИНАТЫ (b_1, b_2) МОЖНО ЗАПИСАТЬ ТАК:
 $(a_1 \cos \theta - a_2 \sin \theta, a_1 \sin \theta + a_2 \cos \theta)$



АА?!

ТО ЕСТЬ,

$$\begin{cases} b_1 = a_1 \cos \theta - a_2 \sin \theta \\ b_2 = a_1 \sin \theta + a_2 \cos \theta \end{cases}$$



$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 \cos \theta - a_2 \sin \theta \\ a_1 \sin \theta + a_2 \cos \theta \end{pmatrix} = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

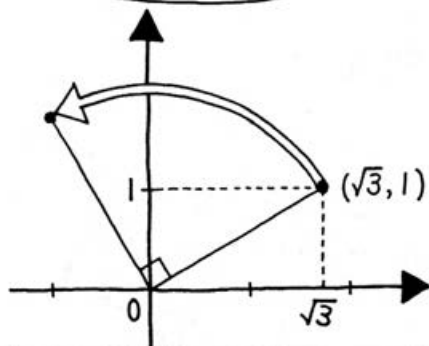
ЗНАЧИТ,
МЫ МОЖЕМ
ПЕРЕПИСАТЬ ЭТО
ВОТ ТАКИМ
ОБРАЗОМ.

В САМОМ
ДЕЛЕ.



ДАВАЙ ТЕПЕРЬ
РАССМОТРИМ
ПРИМЕР.

ПОПРОБУЙ РАССЧИТАТЬ ТОЧКУ
(b_1, b_2), ПОЛУЧЕННУЮ ВРАЩЕНИЕМ
ТОЧКИ ($\sqrt{3}, 1$) НА УГОЛ 90° .



А Я УЖЕ ЗНАЮ!
ПОЛУЧИТСЯ ТОЧКА
($-1, \sqrt{3}$)!

НЕТ, ТЫ
РАССЧИТАЙ.

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \cos 90^\circ & -\sin 90^\circ \\ \sin 90^\circ & \cos 90^\circ \end{pmatrix} \begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix}$$

ТАК...

ПРАВИЛЬНО?

$$\begin{aligned} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} &= \begin{pmatrix} \cos 90^\circ & -\sin 90^\circ \\ \sin 90^\circ & \cos 90^\circ \end{pmatrix} \begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{3} \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 \times \sqrt{3} + (-1) \times 1 \\ 1 \times \sqrt{3} + 0 \times 1 \end{pmatrix} \\ &= \begin{pmatrix} -1 \\ \sqrt{3} \end{pmatrix} \end{aligned}$$

АА, МОЛОДЕЦ!

Мы изучили, что точка (b_1, b_2) , полученная вращением точки (a_1, a_2) на угол θ , записывается как $(a_1 \cos \theta - a_2 \sin \theta, a_1 \sin \theta + a_2 \cos \theta)$. Очевидно также, что при вращении точки (b_1, b_2) на угол $-\theta$ мы опять получим точку (a_1, a_2) . В матричной форме это записывается так:

$$\begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$



Если записать процесс вычисления выражения

$$\begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

более подробно, то мы получим следующее:

$$\begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} \begin{pmatrix} a_1 \cos \theta - a_2 \sin \theta \\ a_1 \sin \theta + a_2 \cos \theta \end{pmatrix} =$$

$$= \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$$

Подставляем
выражение
для точки (b_1, b_2)

См. объяснение
на стр. 70

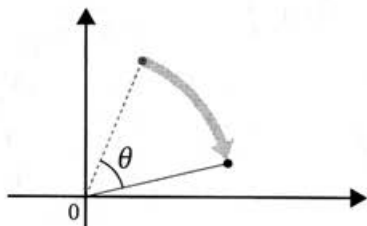
Посмотри внимательно на вышеприведённое выражение. Из него логично предположить, что:

$$\begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \text{ не так ли?}$$

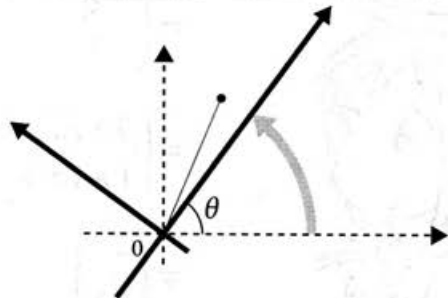
Кстати, два нижеприведённых графика имеют одинаковый смысл.



Вращение точки (b_1, b_2) на угол $-\theta$



Вращение горизонтальной и вертикальной осей на угол θ



СОБСТВЕННЫЕ ЗНАЧЕНИЯ И СОБСТВЕННЫЕ ВЕКТОРЫ



У КАЖДОЙ МАТРИЦЫ ЕСТЬ СОБСТВЕННЫЕ ЗНАЧЕНИЯ И СОБСТВЕННЫЕ ВЕКТОРЫ, СООТВЕТСТВУЮЩИЕ ДРУГ ДРУГУ.

ИТАК, ПЕРЕХОДИМ К СЛЕДУЮЩЕЙ ТЕМЕ.

АА?



НАПРИМЕР...

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \lambda \begin{pmatrix} a_1 \\ a_2 \end{pmatrix}, \text{ то есть } \begin{cases} a_1 + 2a_2 = \lambda a_1 \\ 3a_1 + 4a_2 = \lambda a_2 \end{cases}$$

...СОБСТВЕННОЕ ЗНАЧЕНИЕ МАТРИЦЫ

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \text{ ЭТО...}$$



...ТАКОЕ ЗНАЧЕНИЕ λ (ЛЯМБДА), КОТОРОЕ УДОВЛЕТВОРЯЕТ ВОТ ЭТОМУ УРАВНЕНИЮ.

Лямбда.



А ВЕКТОР $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$, КОТОРЫЙ СООТВЕТСТВУЕТ λ , НАЗЫВАЕТСЯ СОБСТВЕННЫМ ВЕКТОРОМ МАТРИЦЫ $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$.

ЭТИ ПОНЯТИЯ ПОНАДОБЯТСЯ НАМ ПРИ ИЗУЧЕНИИ ФАКТОРНОГО АНАЛИЗА.

ВОТ КАК?

КАК БЫ ТАМ НИ БЫЛО, ДАВАЙ РАССМОТРИМ ПРИМЕР.



ШУРХ

А КАКОЙ ВО ВСЁМ ЭТОМ СМЫСЛ?

ТЕПЕРЬ ПРИВЕДУ ПРИМЕРЫ. ПРОЦЕСС РАСЧЁТОВ
 ДОВОЛЬНО СЛОЖНЫЙ И ГРОМОЗАКИЙ, ПОЭТОМУ
 ПРИВОЖУ ТОЛЬКО РЕЗУЛЬТАТЫ.



Пример 1.

$$\bullet \begin{pmatrix} -10 & 6 \\ -18 & 11 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -10 \times 1 + 6 \times 2 \\ -18 \times 1 + 11 \times 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 4 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$\bullet \begin{pmatrix} -10 & 6 \\ -18 & 11 \end{pmatrix} \begin{pmatrix} 2 \\ 3 \end{pmatrix} = \begin{pmatrix} -10 \times 2 + 6 \times 3 \\ -18 \times 2 + 11 \times 3 \end{pmatrix} = \begin{pmatrix} -2 \\ -3 \end{pmatrix} = -1 \begin{pmatrix} 2 \\ 3 \end{pmatrix}$$

Следовательно, 2 и -1 являются собственными значениями для $\begin{pmatrix} -10 & 6 \\ -18 & 11 \end{pmatrix}$, собственным вектором для 2 является $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$, а собственным вектором для -1 является $\begin{pmatrix} 2 \\ 3 \end{pmatrix}$.

Пример 2.

$$\bullet \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \times 1 + 0 \times 0 + 0 \times 0 \\ 0 \times 1 + 4 \times 0 + 0 \times 0 \\ 0 \times 1 + 0 \times 0 + 6 \times 0 \end{pmatrix} = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} = 2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\bullet \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 2 \times 0 + 0 \times 1 + 0 \times 0 \\ 0 \times 0 + 4 \times 1 + 0 \times 0 \\ 0 \times 0 + 0 \times 1 + 6 \times 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix} = 4 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

$$\bullet \begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 2 \times 0 + 0 \times 0 + 0 \times 1 \\ 0 \times 0 + 4 \times 0 + 0 \times 1 \\ 0 \times 0 + 0 \times 0 + 6 \times 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 6 \end{pmatrix} = 6 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Следовательно, 2, 4 и 6 являются собственными значениями для $\begin{pmatrix} 2 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 6 \end{pmatrix}$, собственным вектором для 2 является $\begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$, собственным вектором для 4 является $\begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, а собственным вектором для 6 является $\begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$.



МАТРИЦА, СОСТОЯЩАЯ ИЗ p СТРОК И p СТОЛБЦОВ,
 КАК ПРАВИЛО, ИМЕЕТ p ПАР СОБСТВЕННЫХ ЗНАЧЕНИЙ
 И СОБСТВЕННЫХ ВЕКТОРОВ.



НО КАК ЖЕ ИХ
ВЫЧИСЛЯЮТ?

НА САМОМ ДЕЛЕ,
ВЫЧИСЛЯТЬ ИХ
НА БУМАГЕ —
ДОВОЛЬНО
СЛОЖНОЕ ЗАНЯТИЕ.

А, ПОНЯТНО!
С ПОМОЩЬЮ
EXCEL, АА?

К СОЖАЛЕНИЮ,
В EXCEL НЕТ
СООТВЕТСТВУЮЩИХ
ФУНКЦИЙ.

КАК?!
ЧТО ЖЕ ТОГДА
ДЕЛАТЬ?!

ПОЭТОМУ
ПРИХОДИТСЯ ЛИБО,
НАБРАВШИЕСЬ
ТЕРПЕНИЯ, БРАТЬ
В РУКИ КАРАНДАШ,
ЛИБО ИСПОЛЬЗОВАТЬ
СПЕЦИАЛЬНЫЕ
ПРОГРАММЫ
АНАЛИЗА ДАННЫХ.

АА?

5. СИММЕТРИЧНЫЕ МАТРИЦЫ

БУДЕТ НЕМНОЖКО
ГРОМОЗДАКОЕ
ОБЪЯСНЕНИЕ, ПОЭТОМУ
СЛУШАЙ ВНИМАТЕЛЬНО.

НАПОСЛЕДОК
Я РАССКАЖУ
О СИММЕТРИЧНЫХ
МАТРИЦАХ.

Хотя
на самом
деле тут
нет ничего
сложного.

ХОРОШО!

СИММЕТРИЧНАЯ
МАТРИЦА — ЭТО...

...ТАКАЯ МАТРИЦА, ЗНАЧЕНИЯ
КОТОРОЙ СИММЕТРИЧНЫ
ОТНОСИТЕЛЬНО ГЛАВНОЙ
ДИАГОНАЛИ.

МНЕ ПОМНИТСЯ, ЧТО
ИМЕННО ТАКИМИ БЫЛИ
МАТРИЦЫ КОРРЕЛЯЦИИ
И ЕДИНИЧНЫЕ
МАТРИЦЫ!

А СЕЙЧАС
Я РАССКАЖУ
О ВАЖНОМ
МОМЕНТЕ.

ДЕЛО В ТОМ, ЧТО
СИММЕТРИЧНУЮ
МАТРИЦУ,...

ШУРХ-ШУРХ

...НАПРИМЕР, ВОТ ЭТУ:

$$\begin{pmatrix} 1 & 0,7 & 0,8 \\ 0,7 & 1 & 0,9 \\ 0,8 & 0,9 & 1 \end{pmatrix} \dots$$

$$\begin{pmatrix} 0.7 & 0.8 \\ 1 & 0.9 \\ 0.9 & 1 \end{pmatrix} = \begin{pmatrix} \sqrt{2.60} \times 0.55 & \sqrt{0.31} \times 0.80 & \sqrt{0.08} \times 0.23 \\ \sqrt{2.60} \times 0.58 & \sqrt{0.31} \times (-0.57) & \sqrt{0.08} \times 0.59 \\ \sqrt{2.60} \times 0.60 & \sqrt{0.31} \times (-0.19) & \sqrt{0.08} \times (-0.78) \end{pmatrix} \begin{pmatrix} \sqrt{2.60} \times 0.55 & \sqrt{2.60} \times 0.58 & \sqrt{2.60} \times 0.60 \\ \sqrt{0.31} \times 0.80 & \sqrt{0.31} \times (-0.57) & \sqrt{0.31} \times (-0.19) \\ \sqrt{0.08} \times 0.23 & \sqrt{0.08} \times 0.59 & \sqrt{0.08} \times (-0.78) \end{pmatrix}$$

Собственный вектор, соответствующий максимальному собственному значению

Собственный вектор, соответствующий 2-му по величине собственному значению

Собственный вектор, соответствующий 3-му по величине собственному значению

Матрица, полученная заменой строк на столбцы матрицы слева



...МОЖНО ПЕРЕПИСАТЬ ВОТ ТАКИМ ОБРАЗОМ.



Да, это выглядит грандиозно.



ДАЛЕЕ, ЕСЛИ, КАК В ЭТОМ ПРИМЕРЕ, ТРЕТЬЕ ПО ВЕЛИЧИНЕ СОБСТВЕННОЕ ЗНАЧЕНИЕ БЛИЗКО К НУЛЮ...

$$\begin{pmatrix} 1 & 0.7 & 0.8 \\ 0.7 & 1 & 0.9 \\ 0.8 & 0.9 & 1 \end{pmatrix} \approx \begin{pmatrix} \sqrt{2.60} \times 0.55 & \sqrt{0.31} \times 0.80 \\ \sqrt{2.60} \times 0.58 & \sqrt{0.31} \times (-0.57) \\ \sqrt{2.60} \times 0.60 & \sqrt{0.31} \times (-0.19) \end{pmatrix} \begin{pmatrix} \sqrt{2.60} \times 0.55 & \sqrt{2.60} \times 0.58 & \sqrt{2.60} \times 0.60 \\ \sqrt{0.31} \times 0.80 & \sqrt{0.31} \times (-0.57) & \sqrt{0.31} \times (-0.19) \end{pmatrix}$$

...ТО ВЫПОЛНЯЕТСЯ ВОТ ТАКОЕ ПРИБЛИЖЁННОЕ РАВЕНСТВО!

ОГО! ОДНА СТРОКА И СТОЛБЕЦ ИСЧЕЗЛИ, ДА?

ДАЛЕЕ, ХОТЯ ДЛЯ ДАННОГО ПРИМЕРА
ЭТО И НЕВЕРНО, НО ЕСЛИ И ВТОРОЕ,
И ТРЕТЬЕ ПО ВЕЛИЧИНЕ СОБСТВЕННЫЕ
ЗНАЧЕНИЯ БЫЛИ БЫ БЛИЗКИ К НУЛЮ,...

$$\begin{pmatrix} 1 & 0.7 & 0.8 \\ 0.7 & 1 & 0.9 \\ 0.8 & 0.9 & 1 \end{pmatrix} \approx \begin{pmatrix} \sqrt{2.60} \times 0.55 \\ \sqrt{2.60} \times 0.58 \\ \sqrt{2.60} \times 0.60 \end{pmatrix} \begin{pmatrix} \sqrt{2.60} \times 0.55 & \sqrt{2.60} \times 0.58 & \sqrt{2.60} \times 0.60 \end{pmatrix}$$



...ТО
ВЫПОЛНЯЛОСЬ БЫ
ВОТ ТАКОЕ
ПРИБЛИЖЁННОЕ
РАВЕНСТВО!

АА?!
КАК
ИНТЕРЕСНО!



ТО, О ЧЁМ Я СЕЙЧАС
РАССКАЗАЛА, ОСОБЕННО
ВАЖНО ДЛЯ ПОНИМАНИЯ
МЕТОДИКИ ВЫЧИСЛЕНИЙ
ФАКТОРНОГО АНАЛИЗА,
ПОЭТОМУ ЗАПОМНИ
ХОРОШЕНЬКО.

ХОРОШО!

ДАВАЙ
СЕГОДНЯ
НА ЭТОМ
ЗАКОНЧИМ.

БОЛЬШОЕ
СПАСИБО!

СОВСЕМ ЗАБЫЛА.
РУИ, ТЫ НЕ ПРОТИВ,
ЕСЛИ СЛЕДУЮЩИЙ
УРОК МЫ ПРОВЕДЕМ
У МЕНЯ ДОМА?
УГОЩУ ТЕБЯ ВКУС-
НЫМ ЧАЕМ.

ТОГАА ДО
СЛЕДУЮЩЕЙ
НЕДЕЛИ!

ДО
СВИДАНИЯ!

хлоп

АА? Я
С УАОВОЛЬСТВОМ
К ТЕБЕ ПРИДУ!

ИТАК,
СЕГОДНЯ
Я ЕЩЁ
НА ОДИН
ШАГ...

...СМОГЛА
ПРИБЛИЗИТЬСЯ
К ЯМАМОТО!

Займись-ка
я повторением.

РУИ...

...ПОЧЕМУ ТЫ
ТАК УСЕРАНО
ЗАНИМАЕШЬСЯ?!

6. ДОПОЛНИТЕЛЬНЫЕ СВЕДЕНИЯ О МАТРИЦАХ

6.1. ПРАВИЛА ЗАПИСИ МАТРИЦ

Начнём с правил записи матриц. Например, система $\begin{cases} x_1 + 2x_2 = -1 \\ 3x_1 + 4x_2 = 5 \end{cases}$ записывается как $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \times \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 5 \end{pmatrix}$, а её левая часть $\begin{cases} x_1 + 2x_2 \\ 3x_1 + 4x_2 \end{cases}$ записывается как $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \times \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$.

Общий случай

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1q}x_q = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2q}x_q = b_2 \\ \dots\dots\dots \\ a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pq}x_q = b_p \end{cases} \rightarrow \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1q} \\ a_{21} & a_{22} & \dots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pq} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix}.$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \dots + a_{1q}x_q \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2q}x_q \\ \dots\dots\dots \\ a_{p1}x_1 + a_{p2}x_2 + \dots + a_{pq}x_q \end{cases} \rightarrow \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1q} \\ a_{21} & a_{22} & \dots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pq} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{pmatrix}.$$

6.2. СЛОЖЕНИЕ МАТРИЦ

Теперь рассмотрим сумму матриц. Например, сумма матриц $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ и $\begin{pmatrix} 4 & 5 \\ -2 & 4 \end{pmatrix}$ записывается как $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} + \begin{pmatrix} 4 & 5 \\ -2 & 4 \end{pmatrix}$ и вычисляется как $\begin{pmatrix} 1+4 & 2+5 \\ 3+(-2) & 4+4 \end{pmatrix}$.

Общий случай

$$\text{Сумма матриц } \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{pmatrix} \text{ и } \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ b_{21} & b_{22} & \cdots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pq} \end{pmatrix}$$

$$\text{записывается как } \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1q} \\ b_{21} & b_{22} & \cdots & b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pq} \end{pmatrix}$$

$$\text{и означает матрицу } \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1q} + b_{1q} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2q} + b_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} + b_{p1} & a_{p2} + b_{p2} & \cdots & a_{pq} + b_{pq} \end{pmatrix}.$$

6.3. ПРОИЗВЕДЕНИЕ МАТРИЦ

Найдём произведение матриц. Например, произведение матриц $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ и $\begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}$, записываемое в виде $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 & y_1 \\ x_2 & y_2 \end{pmatrix}$, хотя и называется «произведением», на самом деле является всего лишь совмещённой записью двух выражений:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \text{ и } \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix},$$

то есть это сокращённая запись двух систем уравнений:

$$\begin{cases} x_1 + 2x_2 \\ 3x_1 + 4x_2 \end{cases} \text{ и } \begin{cases} y_1 + 2y_2 \\ 3y_1 + 4y_2 \end{cases}.$$

Рассмотрим ещё один пример:

$$\text{Произведение} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \\ 13 & 14 & 15 \end{pmatrix} \begin{pmatrix} k_1 & l_1 & m_1 & n_1 \\ k_2 & l_2 & m_2 & n_2 \\ k_3 & l_3 & m_3 & n_3 \end{pmatrix}$$

является совмещённой формой записи четырёх выражений:

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \\ 13 & 14 & 15 \end{pmatrix} \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix} = \begin{pmatrix} k_1 + 2k_2 + 3k_3 \\ 4k_1 + 5k_2 + 6k_3 \\ 7k_1 + 8k_2 + 9k_3 \\ 10k_1 + 11k_2 + 12k_3 \\ 13k_1 + 14k_2 + 15k_3 \end{pmatrix};$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \\ 13 & 14 & 15 \end{pmatrix} \begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} = \begin{pmatrix} l_1 + 2l_2 + 3l_3 \\ 4l_1 + 5l_2 + 6l_3 \\ 7l_1 + 8l_2 + 9l_3 \\ 10l_1 + 11l_2 + 12l_3 \\ 13l_1 + 14l_2 + 15l_3 \end{pmatrix};$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \\ 13 & 14 & 15 \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} = \begin{pmatrix} m_1 + 2m_2 + 3m_3 \\ 4m_1 + 5m_2 + 6m_3 \\ 7m_1 + 8m_2 + 9m_3 \\ 10m_1 + 11m_2 + 12m_3 \\ 13m_1 + 14m_2 + 15m_3 \end{pmatrix};$$

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \\ 13 & 14 & 15 \end{pmatrix} \begin{pmatrix} n_1 \\ n_2 \\ n_3 \end{pmatrix} = \begin{pmatrix} n_1 + 2n_2 + 3n_3 \\ 4n_1 + 5n_2 + 6n_3 \\ 7n_1 + 8n_2 + 9n_3 \\ 10n_1 + 11n_2 + 12n_3 \\ 13n_1 + 14n_2 + 15n_3 \end{pmatrix}.$$

Поэтому данное произведение равно

$$\begin{pmatrix} k_1 + 2k_2 + 3k_3 & l_1 + 2l_2 + 3l_3 & m_1 + 2m_2 + 3m_3 & n_1 + 2n_2 + 3n_3 \\ 4k_1 + 5k_2 + 6k_3 & 4l_1 + 5l_2 + 6l_3 & 4m_1 + 5m_2 + 6m_3 & 4n_1 + 5n_2 + 6n_3 \\ 7k_1 + 8k_2 + 9k_3 & 7l_1 + 8l_2 + 9l_3 & 7m_1 + 8m_2 + 9m_3 & 7n_1 + 8n_2 + 9n_3 \\ 10k_1 + 11k_2 + 12k_3 & 10l_1 + 11l_2 + 12l_3 & 10m_1 + 11m_2 + 12m_3 & 10n_1 + 11n_2 + 12n_3 \\ 13k_1 + 14k_2 + 15k_3 & 13l_1 + 14l_2 + 15l_3 & 13m_1 + 14m_2 + 15m_3 & 13n_1 + 14n_2 + 15n_3 \end{pmatrix}.$$

Общий случай

Произведение матриц $\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{pmatrix}$ и $\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1r} \\ x_{21} & x_{22} & \cdots & x_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ x_{q1} & x_{q2} & \cdots & x_{qr} \end{pmatrix}$

записывается как $\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1r} \\ x_{21} & x_{22} & \cdots & x_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ x_{q1} & x_{q2} & \cdots & x_{qr} \end{pmatrix},$

и означает совмещённую запись следующих выражений:

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{q1} \end{pmatrix}, \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{pmatrix} \begin{pmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{q2} \end{pmatrix}, \dots, \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{pmatrix} \begin{pmatrix} x_{1r} \\ x_{2r} \\ \vdots \\ x_{qr} \end{pmatrix},$$

или следующих систем:

$$\begin{cases} a_{11}x_{11} + a_{12}x_{21} + \cdots + a_{1q}x_{q1} \\ a_{21}x_{11} + a_{22}x_{21} + \cdots + a_{2q}x_{q1} \\ \dots \\ a_{p1}x_{11} + a_{p2}x_{21} + \cdots + a_{pq}x_{q1} \end{cases}, \begin{cases} a_{11}x_{12} + a_{12}x_{22} + \cdots + a_{1q}x_{q2} \\ a_{21}x_{12} + a_{22}x_{22} + \cdots + a_{2q}x_{q2} \\ \dots \\ a_{p1}x_{12} + a_{p2}x_{22} + \cdots + a_{pq}x_{q2} \end{cases}, \dots, \begin{cases} a_{11}x_{1r} + a_{12}x_{2r} + \cdots + a_{1q}x_{qr} \\ a_{21}x_{1r} + a_{22}x_{2r} + \cdots + a_{2q}x_{qr} \\ \dots \\ a_{p1}x_{1r} + a_{p2}x_{2r} + \cdots + a_{pq}x_{qr} \end{cases}$$

6.4. ОБРАТНЫЕ МАТРИЦЫ

Теперь обратимся к обратной матрице. Обратная матрица — это такая матрица, которая при умножении на исходную даёт в результате единичную матрицу.

Обозначают обратную матрицу как возведение в степень (-1), например $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^{-1}$.

Например, обратной матрицей для $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$ будет матрица $\begin{pmatrix} -2 & 1 \\ 1,5 & -0,5 \end{pmatrix}$. Проверим это:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} -2 & 1 \\ 1,5 & -0,5 \end{pmatrix} = \begin{pmatrix} 1 \times (-2) + 2 \times 1,5 & 1 \times 1 + 2 \times (-0,5) \\ 3 \times (-2) + 4 \times 1,5 & 3 \times 1 + 4 \times (-0,5) \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Следовательно можно записать:

$$\begin{pmatrix} -2 & 1 \\ 1,5 & -0,5 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}^{-1}.$$

Общий случай

Обратная матрица для матрицы $\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{pmatrix}$

обозначается как $\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{pmatrix}^{-1}$

и удовлетворяет условию $\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p} \\ a_{21} & a_{22} & \dots & a_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pp} \end{pmatrix}^{-1} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$

6.5. ТРАНСПОНИРОВАННЫЕ МАТРИЦЫ

Перейдём к транспонированию матриц. При транспонировании матриц строки и столбцы меняются местами. Например матрицей, транспонированной к матрице $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$, будет матрица $\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$.

Транспонированную матрицу принято обозначать буквой t в верхнем индексе, размещённой перед или после матрицы, либо апострофом после матрицы:

$$\begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} = {}^t \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}' = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}'.$$

Рассмотрим ещё примеры.

Матрицей, транспонированной к $\begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \\ 10 & 11 & 12 \\ 13 & 14 & 15 \end{pmatrix}$, будет матрица $\begin{pmatrix} 1 & 4 & 7 & 10 & 13 \\ 2 & 5 & 8 & 11 & 14 \\ 3 & 6 & 9 & 12 & 15 \end{pmatrix}$.

Матрицей, транспонированной к $(-3 \ 0 \ 8 \ -7)$, будет матрица $\begin{pmatrix} -3 \\ 0 \\ 8 \\ -7 \end{pmatrix}$.

Общий случай

Транспонированная матрица для матрицы $\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1q} \\ a_{21} & a_{22} & \dots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pq} \end{pmatrix}$

обозначается как $\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1q} \\ a_{21} & a_{22} & \dots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pq} \end{pmatrix}'$ или $\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1q} \\ a_{21} & a_{22} & \dots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pq} \end{pmatrix}'$

и получается заменой строк на столбцы: $\begin{pmatrix} a_{11} & a_{21} & \dots & a_{p1} \\ a_{12} & a_{22} & \dots & a_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1q} & a_{2q} & \dots & a_{pq} \end{pmatrix}$.

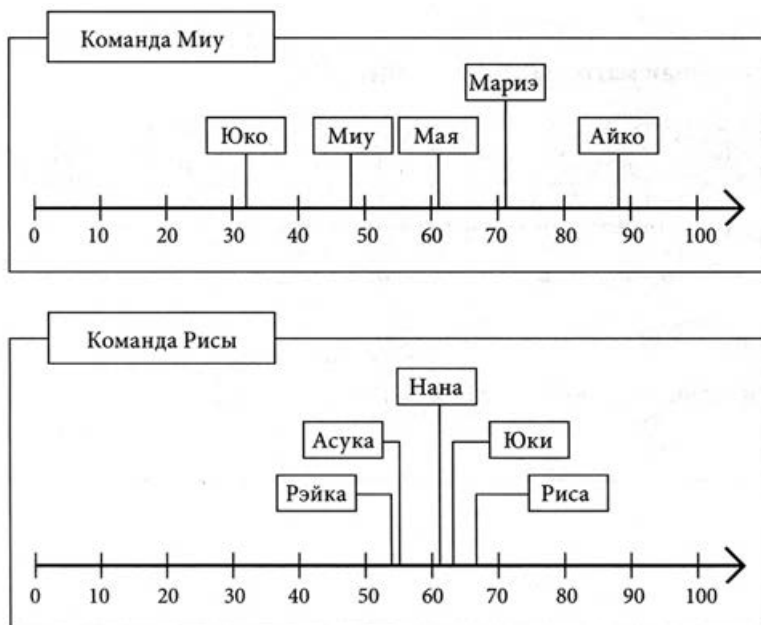
7. СУММА КВАДРАТОВ ОТКЛОНЕНИЙ, ДИСПЕРСИЯ СЛУЧАЙНОЙ ВЕЛИЧИНЫ, КВАДРАТИЧНОЕ ОТКЛОНЕНИЕ

Миу и Риса вместе с коллегами по работе пошли в караоке, где они разделились на 2 команды по 5 человек в каждой и стали состязаться в пении. Оценки, полученные каждым участником, приведены в табл. 3.1.

◆ Таблица 3.1. Результаты состязания в караоке

Команда Миу (баллов)		Команда Рисы (баллов)	
Миу	48	Риса	67
Юко	32	Асука	55
Айко	88	Нана	61
Мая	61	Юки	63
Мариэ	71	Рэйка	54
Средний балл	60	Средний балл	60

Ниже приведены графики, построенные на основе данных об оценках состязания.



■ Рис.3.1. Результаты состязания в караоке

Средний балл как в команде Миу, так и в команде Рисы оказался одинаковым — 60 баллов. Однако общая картина совершенно отличалась. У команды Миу наблюдалась, так сказать, большая неравномерность баллов, или, другими словами, «степень разброса» данных была выше.

В качестве показателя «степени разброса» данных существуют такие вещи, как *сумма квадратов отклонений*, *дисперсия* и *среднеквадратичное отклонение* (*квадратичное отклонение*). Все эти показатели характеризуются следующими свойствами:

- минимальное значение равно 0;
- значение тем больше, чем выше «степень разброса» данных.

Сумма квадратов отклонений часто используется для различных методов анализа, начиная с регрессионного анализа, и находится как сумма квадратов разностей каждого из значений данных и среднего арифметического значения. Этот показатель имеет один роковой недостаток — чем больше будет данных, тем больше будет его значение, поэтому в качестве показателя разброса данных он используется довольно редко.

Дисперсия — это показатель, который исправляет недостаток показателя *сумма квадратов отклонений*. Он вычисляется¹ как:

$$\frac{\text{Сумма квадратов отклонений}}{\text{Число данных}}$$

Среднеквадратичное отклонение (*квадратичное отклонение*), имеющее аналогичный дисперсии смысл, вычисляется по формуле:

$$\sqrt{\text{Дисперсия}}$$

Давайте попробуем найти сумму квадратов отклонений, дисперсию и среднеквадратичное отклонение для обеих команд Миу и Рисы.

◆ **Таблица 3.2.** Сумма квадратов отклонений, дисперсия и среднеквадратичное отклонение для команд Миу и Рисы

	Команда Миу	Команда Рисы
Сумма квадратов отклонений	$(48-60)^2+(32-60)^2+(88-60)^2+(61-60)^2+(71-60)^2 = (-12)^2+(-28)^2+28^2+1^2+11^2$ = 1834	$(67-60)^2+(55-60)^2+(61-60)^2+(63-60)^2+(54-60)^2 = 7^2+(-5)^2+1^2+3^2+(-6)^2$ = 120
Дисперсия	$\frac{1834}{5} = 366,8$	$\frac{120}{5} = 24$
Квадратичное отклонение	$\sqrt{366,8} = 19,2$	$\sqrt{24} = 4,9$

¹ Существует также *несмещённая дисперсия*, у которой в знаменателе находится не «число данных», а «число данных – 1». Чтобы не перегружать материал данной книги, здесь не рассматривается различие этих двух типов дисперсий.

4

МЕТОД ГЛАВНЫХ КОМПОНЕНТ

1. ЧТО ТАКОЕ МЕТОД ГЛАВНЫХ КОМПОНЕНТ
2. ВАЖНЫЕ МОМЕНТЫ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ
3. КОНКРЕТНЫЙ ПРИМЕР ИСПОЛЬЗОВАНИЯ
МЕТОДА ГЛАВНЫХ КОМПОНЕНТ
4. ВЫБОР ПЕРЕМЕННЫХ И ПЕРВАЯ ГЛАВНАЯ КОМПОНЕНТА
5. ПЕРВАЯ ГЛАВНАЯ КОМПОНЕНТА И ОБЩИЙ ПОКАЗАТЕЛЬ
6. КРИТЕРИЙ СУММАРНОГО ВКЛАДА
7. ГЛАВНЫЕ КОМПОНЕНТЫ, НАЧИНАЯ СО ВТОРОЙ
8. ДИСПЕРСИЯ И СОБСТВЕННЫЕ ЗНАЧЕНИЯ





ДА... КАК
Я И ОЖИДАЛА, ТВОЯ
КОМНАТА, МИУ, ОЧЕНЬ
МИЛО ОБСТАВЛЕНА..

ДА,
ПРАВДА?



ОГО!
ПИАНИНО!

ОГО!
Милые вещички!

ОЙ, А ЭТО ЧТО?



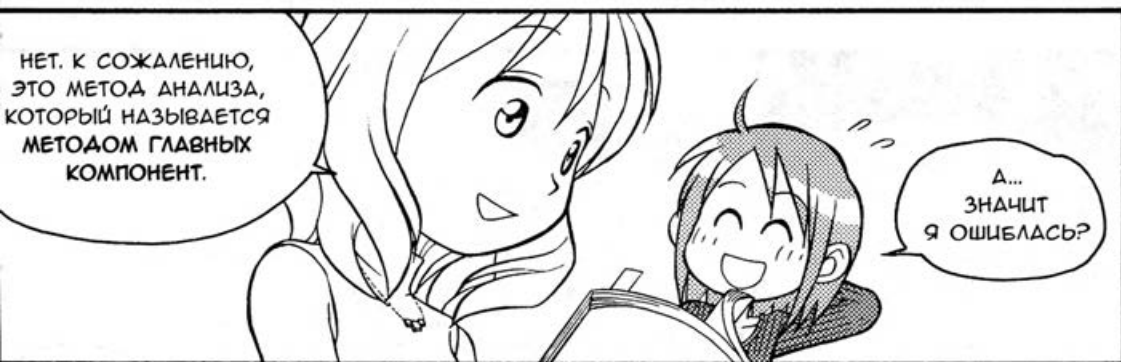
Нем!



КХЕ

ИТАК, Я НАЧИНАЮ
ЗАНЯТИЕ
ПО ФАКТОРНОМУ
АНАЛИЗУ.

ХОРОШО!





РУЦ, МОЖЕТ БЫТЬ
ТЕБЕ ПОПРОБОВАТЬ
ИЗУЧИТЬ МЕТОД
ГЛАВНЫХ
КОМПОНЕНТ?

ЧТО?

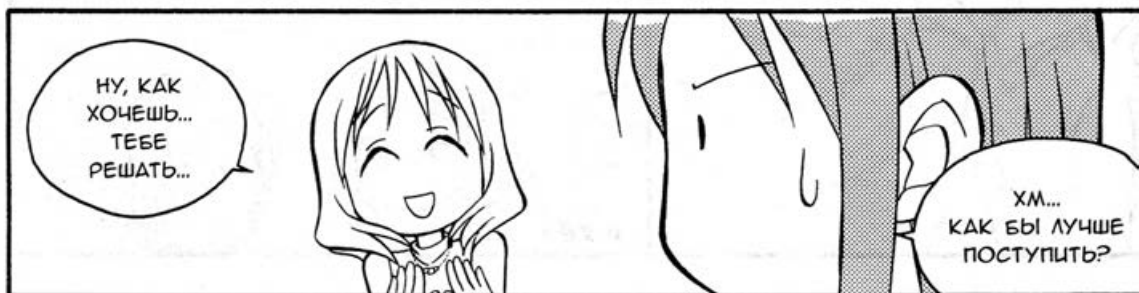


СЧИТАЕТСЯ, ЧТО
МЕТОД ГЛАВНЫХ
КОМПОНЕНТ ПОХОЖ
НА ФАКТОРНЫЙ
АНАЛИЗ, И МНОГИЕ
ЛЮДИ ПУТАЮТ
ЭТИ ДВА МЕТОДА.



Я НЕ ХОЧУ, РУЦ, ЧТОБЫ ТЫ
ПУТАЛА ИХ, ПОЭТОМУ
ПОДУМАЛА, ЧТО ПО ЭТОМУ
СЛУЧАЮ ТЕБЕ ЛУЧШЕ ИЗУЧИТЬ
МЕТОД ГЛАВНЫХ КОМПОНЕНТ.

ВОТ ОНО
ЧТО...



НУ, КАК
ХОЧЕШЬ...
ТЕБЕ
РЕШАТЬ...

ХМ...
КАК БЫ ЛУЧШЕ
ПОСТУПИТЬ?



КОНЕЧНО,
Я ХОЧУ КАК
МОЖНО СКОРЕЕ
ПРИСТУПИТЬ
К ФАКТОРНОМУ
АНАЛИЗУ...



Ого!
Ты понимаешь
даже разницу
между этими
методами?!

Разумеется!

...НО ТВОЁ
ПРЕДЛОЖЕНИЕ
ТОЖЕ ЗВУЧИТ
ЗАМАНЧИВО...



СОГЛА-
СИТЬСЯ, ЧТО
ЛИ?

ЗНАЧИТ, ТЫ
СОГЛАСНА?

Хи-хи-хи

1. ЧТО ТАКОЕ МЕТОД ГЛАВНЫХ КОМПОНЕНТ

ИТАК, МЫ МЕНЯЕМ
НАШИ ПЛАНЫ
ИЗУЧАЕМ СЕГОДНЯ
МЕТОД ГЛАВНЫХ
КОМПОНЕНТ!



АА,
ПОЖАЛУЙСТА!



КСТАТИ, МЕТОД
ГЛАВНЫХ КОМПОНЕНТ
С ОПРЕДЕЛЁННЫМИ
ОГОВОРКАМИ
ДЕЙСТВИТЕЛЬНО МОЖНО
СЧИТАТЬ АНАЛОГИЧНЫМ
ФАКТОРНОМУ АНАЛИЗУ.

ПРАВАА?
НЕУЖЕЛИ ОНИ
НАСТОЛЬКО
ПОХОЖИ?

НО С ДРУГОЙ СТОРОНЫ,
БЕЗ ЭТИХ ОГОВОРК
МЕТОД ГЛАВНЫХ
КОМПОНЕНТ И ФАКТОРНЫЙ
АНАЛИЗ — ЭТО РАЗНЫЕ
МЕТОДЫ.

МЕТОД
ГЛАВНЫХ
КОМПОНЕНТ, НА
САМОМ
ДЕЛЕ...

...ПРЕСТАВЛЯЕТ СОБОЙ
МЕТОД АНАЛИЗА
ДЛЯ ОПРЕДЕЛЕНИЯ
"РЕКОРАСМЕНОВ
ПО ОБЩИМ
ПОКАЗАТЕЛЯМ".

Победа!

Чемпионка по общему показателю

ПОНЯТНО...
ТАК ЧТО ЖЕ
ЭТО ЗА
МЕТОД?

ЧТО ТЫ
ИМЕЕШЬ
В ВИДУ?

В НОМЕРЕ
ЖУРНАЛА О КИНО
ЗА ПРОШЛЫЙ МЕСЯЦ
КАК РАЗ ЗАТРОНУТА
ПОДХОДЯЩАЯ ТЕМА.

РУЦ, ТЫ ЧАСТО
СМОТРИШЬ
ФИЛЬМЫ?

Сказание о кампусе,
в котором растут дыни

НУ, ИЗРЕДКА
Я ХОЖУ
В КИНОТЕАТР,
ИНОГДА
СМОТРЮ DVD.

Число прозанных DVD

Число посетителей кинотеатров

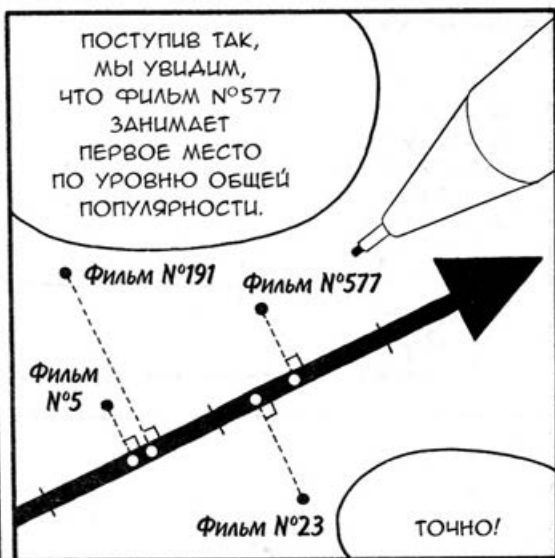
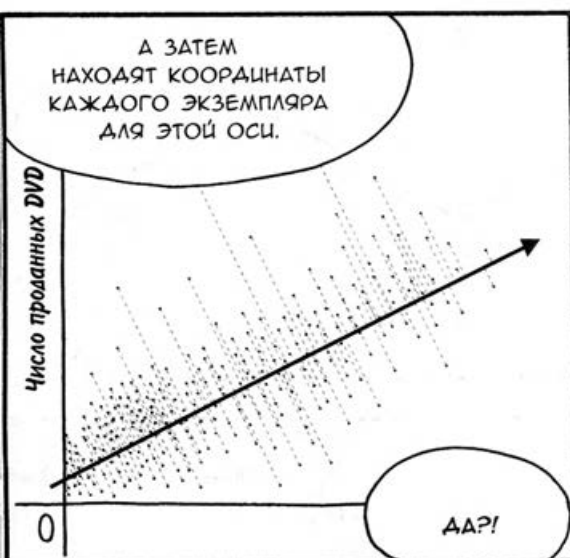
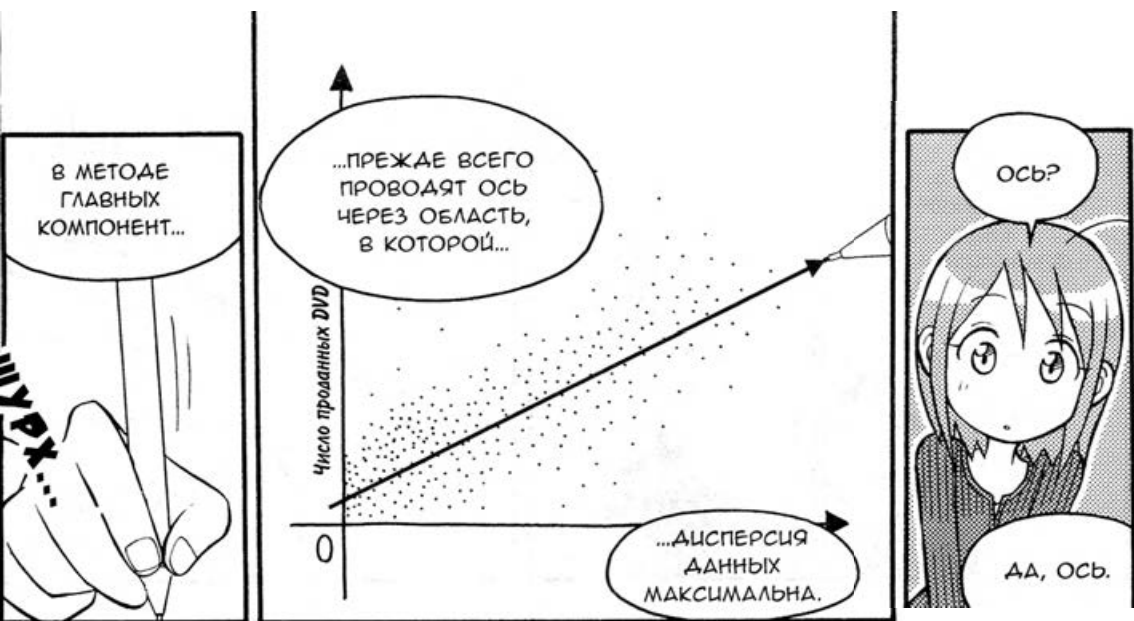
КАК ТЫ ДУМАЕШЬ,
КАКОЙ ФИЛЬМ
ИЗ НИХ ЗАНИМАЕТ
ПЕРВОЕ МЕСТО
ПО УРОВНЮ ОБЩЕЙ
ПОПУЛЯРНОСТИ?

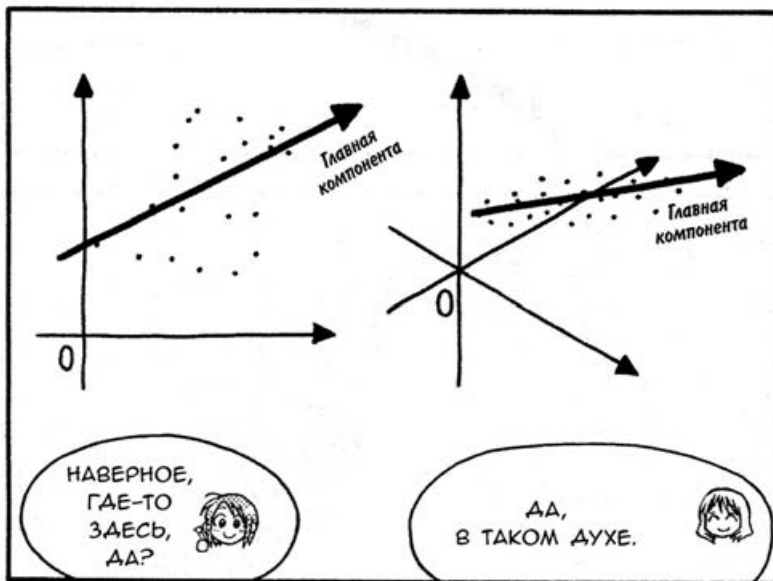
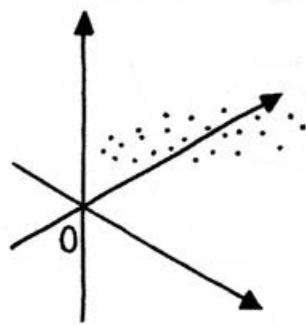
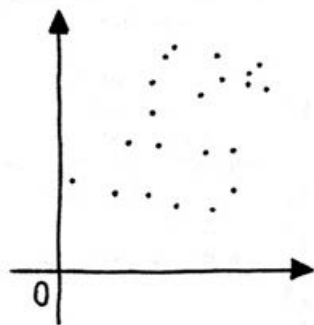
ЧТО?

НАВЕРНОЕ, ЕМУ
СООТВЕТСТВУЕТ
ОДНА ИЗ ЭТИХ
ТОЧЕК, ДА?

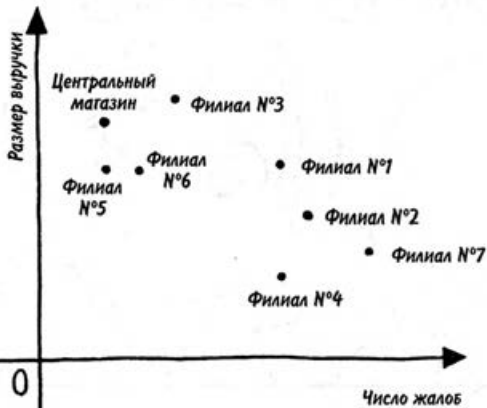
ЭТО ГРАФИК "ЧИСЛА
ПОСЕТИТЕЛЕЙ КИНОТЕАТРОВ"
И "ОБЪЁМА ПРОДАЖ DVD"
ДЛЯ ВСЕХ ФИЛЬМОВ,
ВЫШЕДШИХ В ПРОШЛОМ
ГОДУ.

ДА, ТЫ
ПРАВА.





ПУСТЬ ОДНА
ТОРГОВАЯ СЕТЬ
ПРОВЕРИЛА "ЧИСЛО
ЖАЛОБ" И "РАЗМЕР
ВЫРУЧКИ".



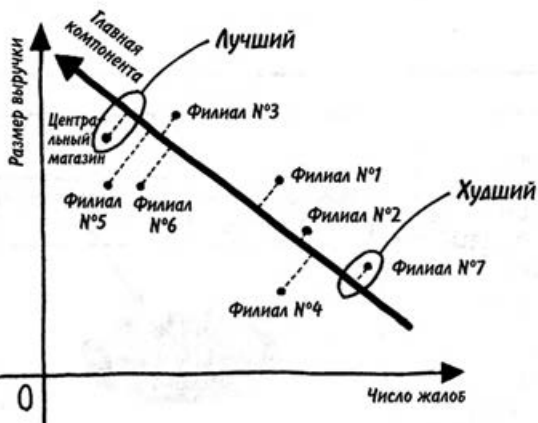
Ага...
Наклон
отрицательный,
да?



СМОЖЕШЬ ЛИ ТЫ
НАЗВАТЬ ЛУЧШИЙ
И ХУДАШИЙ МАГАЗИНЫ
В ЭТОМ ПРИМЕРЕ?



МНЕ
КАЖЕТСЯ...



...ЧТО БУДЕТ
ВОТ ТАК!

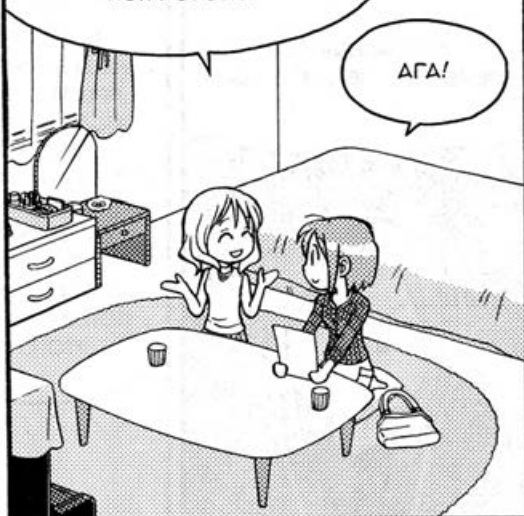


АА,
ВЕРНО!

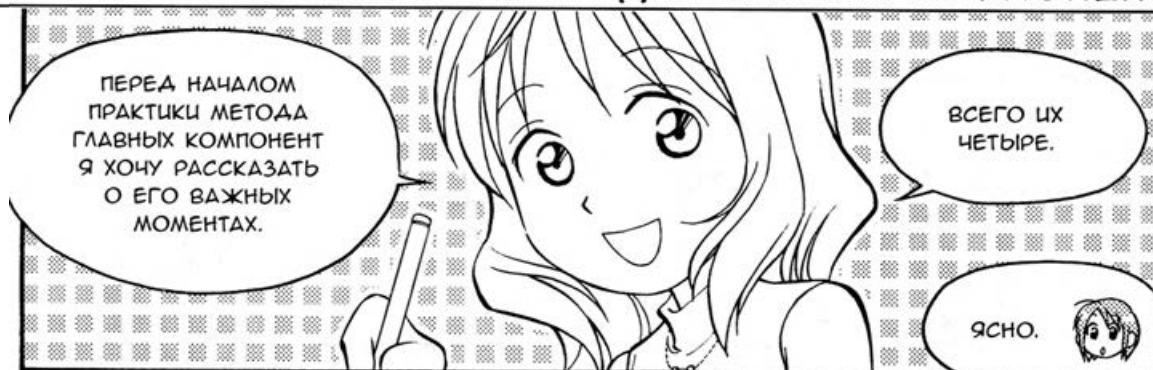


НУ КАК, СМОГЛА ЛИ ТЫ
ПОЛУЧИТЬ ОБЩЕЕ
ПРЕДСТАВЛЕНИЕ
О МЕТОДЕ ГЛАВНЫХ
КОМПОНЕНТ?

АГА!



2. ВАЖНЫЕ МОМЕНТЫ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ



НО ОБРАТИ
ВНИМАНИЕ!



В МЕТОДЕ ГЛАВНЫХ
КОМПОНЕНТ ОТКЛИК
ЯВЛЯЕТСЯ "ПЛОДОМ
ВООБРАЖЕНИЯ",
А НЕ РЕАЛЬНО
СУЩЕСТВУЮЩЕЙ
ПЕРЕМЕННОЙ.



ДАВАЙ ПРИМЕМ, ЧТО
ГЛАВНАЯ КОМПОНЕНТА,
СОСТОЯЩАЯ
ИЗ "ЧИСЛА ПОСЕТИТЕЛЕЙ
КИНОТЕАТРОВ" И "ЧИСЛА
ПРОДАНЫХ DVD", ЯВЛЯ-
ЕТСЯ "ОБЩИМ УРОВНЕМ
ПОПУЛЯРНОСТИ"!



ВОТ ОНО ЧТО...?

В САМОМ ДЕЛЕ,
ТАКИХ ДАННЫХ, КАК
"ОБЩИЙ УРОВЕНЬ
ПОПУЛЯРНОСТИ",
НАВЕРНОЕ,
НЕ СУЩЕСТВУЕТ.



Метод
главных
компонент
Важный
момент **2**

ТЕПЕРЬ —
ВТОРОЙ ВАЖНЫЙ
МОМЕНТ...





В МЕТОДЕ ГЛАВНЫХ КОМПОНЕНТ СУЩЕСТВУЕТ ДВА СПОСОБА ВЫЧИСЛЕНИЙ:...

...С НОРМИРОВКОЙ И БЕЗ НОРМИРОВКИ ДАННЫХ.

Без нормировки данных

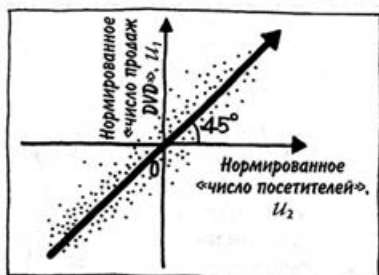
	Число посетителей кинотеатров	Число проданных DVD
Фильм №1	980	90
⋮	⋮	⋮
Фильм №742	770	110
Ср. арифм.	660	90
СКО	120	17

С нормировкой данных

	Нормирован. «число посетителей», U_2	Нормированное «число продаж DVD», U_1
Фильм №1	2.7	0
⋮	⋮	⋮
Фильм №742	0.9	1.2
Ср. арифм.	0	0
СКО	1	1



В случае нормировки получается ось, проходящая через начало координат под углом 45° да!



ЭТО, НАВЕРНОЕ, ИЗ-ЗА ТОГО, ЧТО «ЧИСЛО ПОСЕТИТЕЛЕЙ КИНОТЕАТРОВ» И «ЧИСЛО ПРОДАНЫХ DVD» ПРИВОДЯТСЯ К НОРМИРОВАННЫМ ЕДИНИЦАМ ИЗМЕРЕНИЯ?



— А Я ЗНАЮ!

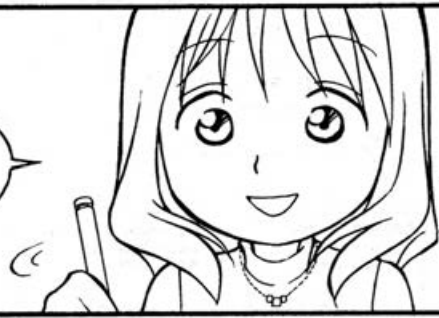
ДА.

ПОХОЖЕ, ЧТО БОЛЬШИНСТВО ИССЛЕДОВАТЕЛЕЙ ПРЕПОЧТИТАЮТ ПРОВОДИТЬ НОРМИРОВКУ ПЕРЕД ИСПОЛЬЗОВАНИЕМ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ.

ПОЭТОМУ СЕГОДНЯ Я РАССКАЖУ О МЕТОДЕ С НОРМИРОВКОЙ.

ХОРОШО!

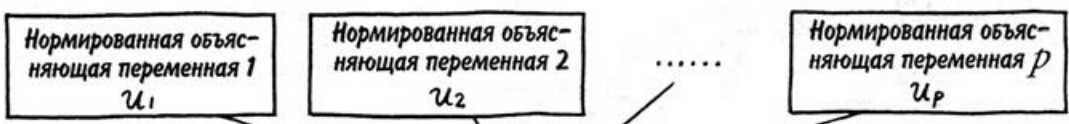
ТРЕТИЙ ВАЖНЫЙ МОМЕНТ.



ВЫРАЗИВ СТРУКТУРУ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ С ПОМОЩЬЮ ФОРМУЛЫ И СХЕМЫ,...

$$z = a_1 u_1 + a_2 u_2 + \dots + a_p u_p$$

↑
 Главная компонента Нормированная объясняющая переменная 1 Нормированная объясняющая переменная 2 Нормированная объясняющая переменная p



a_1 a_2 a_p

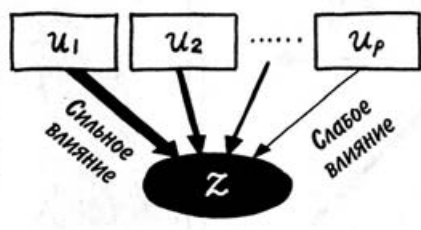
Главная компонента z

...МЫ ПОЛУЧИМ СЛЕДУЮЩЕЕ.

А ЧТО ОЗНАЧАЮТ ЭТИ a_1, a_2 И ТАК ДАЛЕЕ?

ЭТО СТЕПЕНИ ВЛИЯНИЯ ОБЪЯСНЯЮЩИХ ПЕРЕМЕННЫХ НА ГЛАВНУЮ КОМПОНЕНТУ.

ЯСНО.



ПОХОЖЕ, ДЛЯ НИХ НЕТ ОБЩЕПРИНЯТОГО НАЗВАНИЯ*.



АА?

$$z = a_1 u_1 + \dots$$

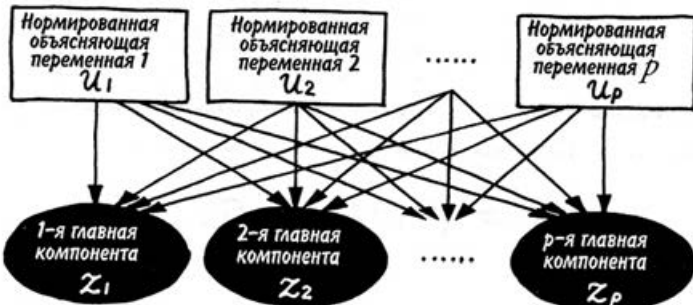
* Коэффициенты a_1, a_2, \dots называются «весами», или «факторными нагрузками» главной компоненты z на соответствующие переменные (прим. перев.)

ЧЕТВЕРТЫЙ ВАЖНЫЙ МОМЕНТ КАСАЕТСЯ ЧИСЛА ГЛАВНЫХ КОМПОНЕНТ.



ЧИСЛА? НО ВЕАЬ ЕСТЬ ТОЛЬКО ОДНА ОСЬ, ПРОХОДЯЩАЯ ЧЕРЕЗ ОБЛАСТЬ С МАКСИМАЛЬНОЙ ДИСПЕРСИЕЙ?

НУ-У, НА САМОМ ДЕЛЕ ГЛАВНЫХ КОМПОНЕНТ СТОЛЬКО, СКОЛЬКО ЕСТЬ ОБЪЯСНЯЮЩИХ ПЕРЕМЕННЫХ.

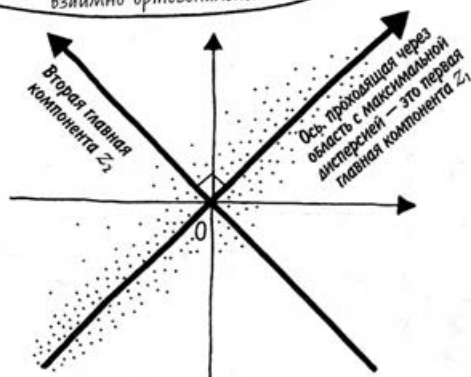


ЕСЛИ ЧИСЛО ОБЪЯСНЯЮЩИХ ПЕРЕМЕННЫХ РАВНО p , ТО СУЩЕСТВУЕТ p ГЛАВНЫХ КОМПОНЕНТ:

$$\begin{cases} z_1 = a_{11}u_1 + a_{12}u_2 + \dots + a_{1p}u_p \\ z_2 = a_{21}u_1 + a_{22}u_2 + \dots + a_{2p}u_p \\ \dots \\ z_p = a_{p1}u_1 + a_{p2}u_2 + \dots + a_{pp}u_p \end{cases}$$

ПРИ ЭТОМ "ОБЩЕМУ ПОКАЗАТЕЛЮ" СООТВЕТСТВУЕТ ПЕРВАЯ ГЛАВНАЯ КОМПОНЕНТА.

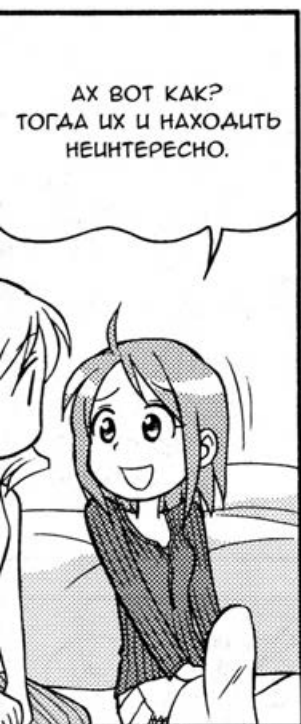
Кстати, все главные компоненты взаимно ортогональны.



А ЧТО ВЫРАЖАЮТ ДРУГИЕ ГЛАВНЫЕ КОМПОНЕНТЫ?



ОСОБОГО СМЫСЛА ОНИ НЕ ИМЕЮТ, ТАК КАК НАХОДЯТСЯ АВТОМАТИЧЕСКИ (МАТЕМАТИЧЕСКИ), ПОМИМО ВОЛИ АНАЛИТИКА.



АХ ВОТ КАК?
ТОГДА ИХ И НАХОДИТЬ
НЕИНТЕРЕСНО.

ОДНАКО, В НЕКОТОРЫХ СЛУЧАЯХ
ЕСТЬ ВОЗМОЖНОСТЬ ПРИДАТЬ
СМЫСЛ АПОСТЕРИОРНО.
НАПРИМЕР, ПОЛОЖИТЕЛЬНУЮ
ПОЛУОСЬ ВТОРОЙ ГЛАВНОЙ
КОМПОНЕНТЫ ЧИНОГДА МОГУТ
РАССМАТРИВАТЬ КАК
"ДОРОГОЙ ТИП",
А ОТРИЦАТЕЛЬНУЮ ПОЛУОСЬ —
"КАК ДЕШЕВЫЙ ТИП", ИЛИ
ПОЛОЖИТЕЛЬНУЮ ПОЛУОСЬ —
КАК "ПОДВИЖНЫЙ ТИП",
А ОТРИЦАТЕЛЬНУЮ ПОЛУОСЬ —
КАК "ЖАДУЩИЙ ТИП."

ТО ЕСТЬ, МЫ СНАЧАЛА
ПОЛУЧАЕМ РЕЗУЛЬТАТЫ,
А ПОТОМ СМОТРИМ НА НИХ
И ИНТЕРПРЕТИРУЕМ
НА СВОЁ УСМОТРЕНИЕ, ДА?



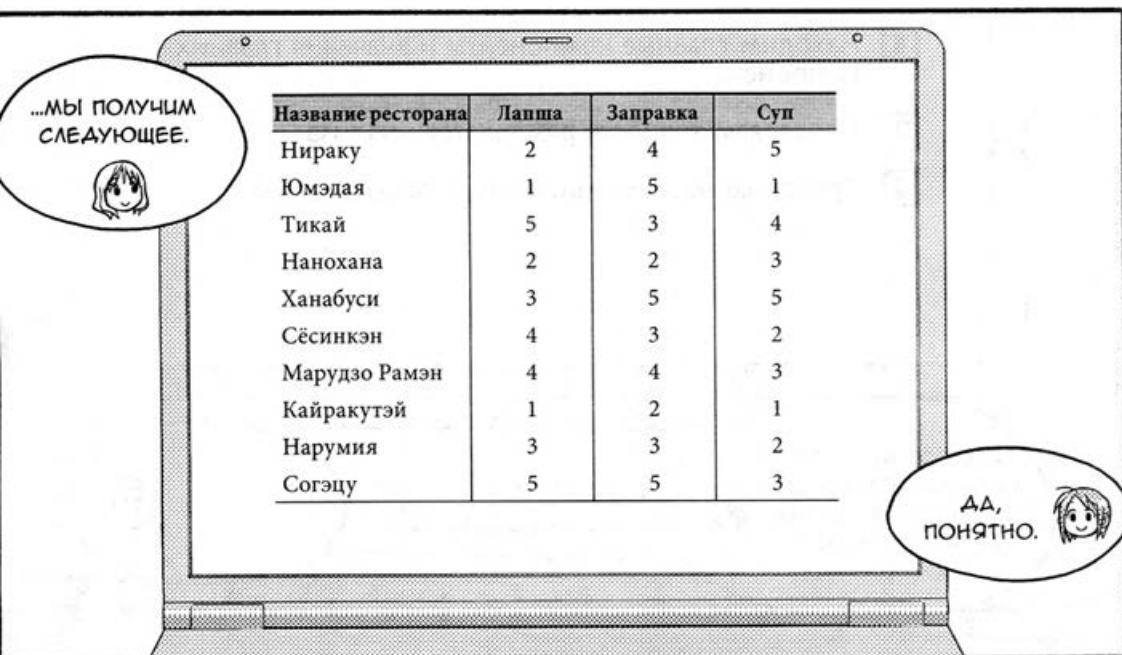
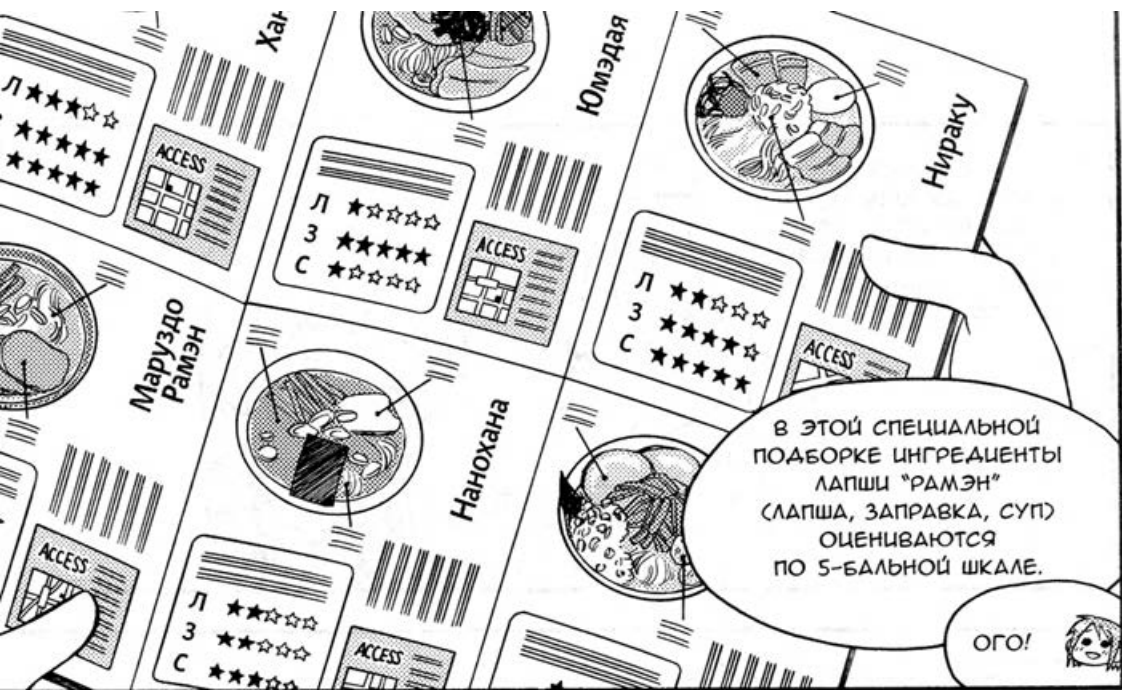
В ЭТОЙ СВЯЗИ
Я ХОЧУ СКАЗАТЬ
ВОТ О ЧЁМ:

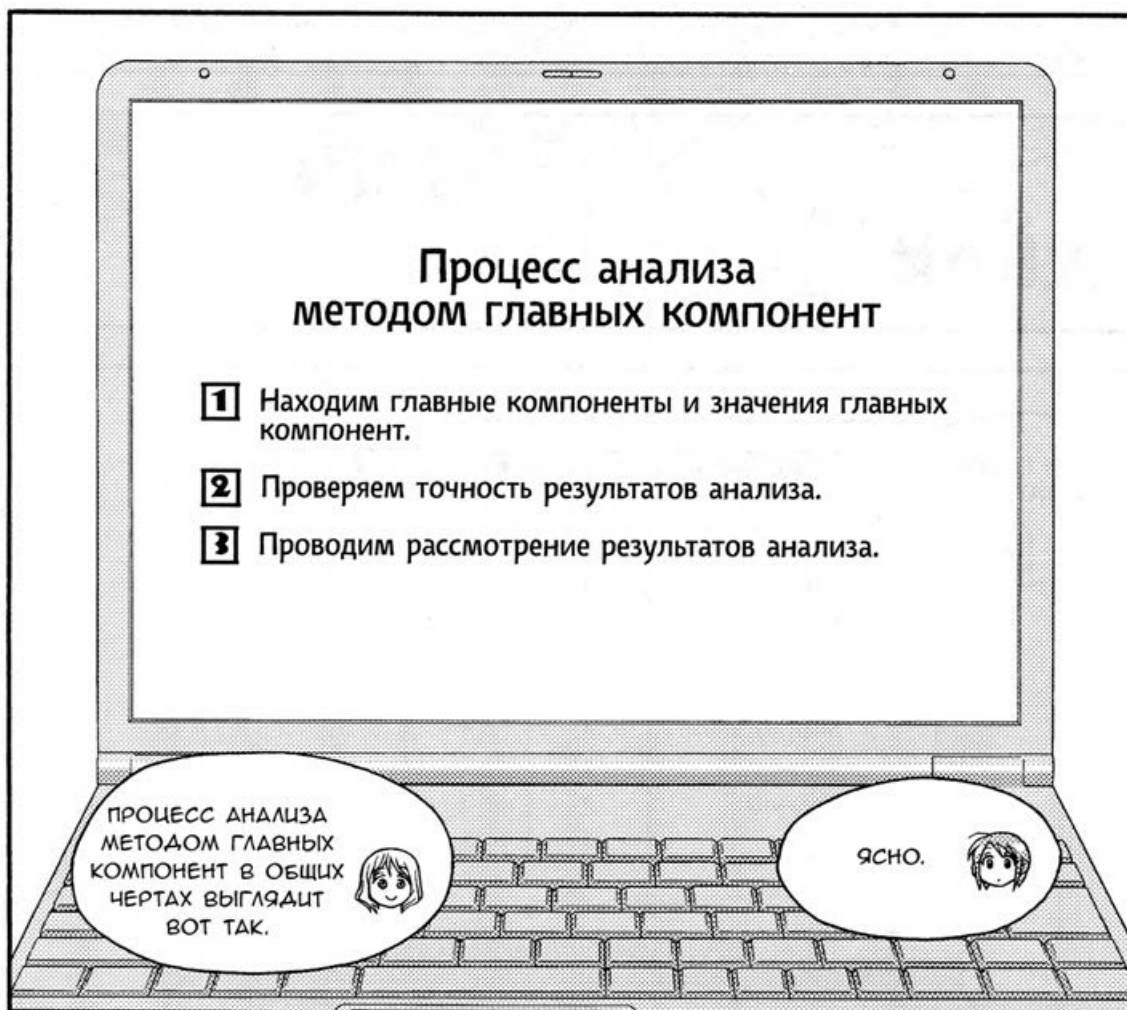
ХОТЯ ГЛАВНЫХ
КОМПОНЕНТ МОЖНО
НАЙТИ СТОЛЬКО, СКОЛЬКО
ИМЕЕТСЯ ОБЪЯСНЯЮЩИХ
ПЕРЕМЕННЫХ, В МЕТОДЕ
ГЛАВНЫХ КОМПОНЕНТ
ОБЫЧНО НАХОДЯТ
ТОЛЬКО ПЕРВУЮ И
ВТОРУЮ ГЛАВНЫЕ
КОМПОНЕНТЫ
И ПРЕДСТАВЛЯЮТ ИХ
В ВИДЕ ДВУХМЕРНОГО
ТОЧЕЧНОГО ГРАФИКА.

ПОНЯТНО. ЗНАЧИТ
ДО ВТОРОЙ ГЛАВНОЙ
КОМПОНЕНТЫ ИСКАТЬ
В ЛЮБОМ СЛУЧАЕ
ПРИДЁТСЯ, ДА?

3. КОНКРЕТНЫЙ ПРИМЕР ИСПОЛЬЗОВАНИЯ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ







1 НАХОДИМ ГЛАВНЫЕ КОМПОНЕНТЫ И ЗНАЧЕНИЯ ГЛАВНЫХ КОМПОНЕНТ

НАХОЖДЕНИЕ ГЛАВНЫХ КОМПОНЕНТ И ЗНАЧЕНИЙ ГЛАВНЫХ КОМПОНЕНТ ПРОИЗВОДИТСЯ В 7 ШАГОВ. Я ОПУЩУ СЛОЖНЫЕ МАТЕМАТИЧЕСКИЕ ПОДРОБНОСТИ, НО ЗАМЕЧУ, ЧТО В ПРОЦЕССЕ ВЫЧИСЛЕНИЙ ИСПОЛЬЗУЕТСЯ МЕТОД МНОЖИТЕЛЕЙ ЛАГРАНЖА. ЗАПОМНИТЕ ХОТЯ БЫ ЕГО НАЗВАНИЕ.



Шаг 1

Проводим нормировку по каждой из переменных.

(нп — нормированная переменная, СА — среднее арифметическое, СКО — среднееквадратичное отклонение)

Название ресторана	Лапша	Заправка	Суп	Название ресторана	Лапша (нп) u_1	Заправка (нп) u_2	Суп (нп) u_3
Нираку	2	4	5	Нираку	-0,7	0,3	1,4
Юмэдая	1	5	1	Юмэдая	-1,3	1,2	-1,3
Тикай	5	3	4	Тикай	1,3	-0,5	0,8
Нанохана	2	2	3	Нанохана	-0,7	-1,4	0,1
Ханабуси	3	5	5	Ханабуси	0,0	1,2	1,4
Сёсинкэн	4	3	2	Сёсинкэн	0,7	-0,5	-0,6
Марудзо Рамэн	4	4	3	Марудзо Рамэн	0,7	0,3	0,1
Кайракутэй	1	2	1	Кайракутэй	-1,3	-1,4	-1,3
Нарумия	3	3	2	Нарумия	0,0	-0,5	-0,6
Согэцу	5	5	3	Согэцу	1,3	1,2	0,1
СА	3,0	3,6	2,9	СА	0	0	0
СКО	1,5	1,2	1,4	СКО	1	1	1

$$\sqrt{\frac{(2-3,0)^2 + \dots + (5-3,0)^2}{10-1}} = 1,5$$

$$\frac{3-2,9}{1,4} = 0,1$$

ЗНАМЕНАТЕЛЬ СРЕДНЕКВАДРАТИЧНОГО ОТКЛОНЕНИЯ, ИСПОЛЬЗУЕМОГО ПРИ НОРМИРОВКЕ В МЕТОДЕ ГЛАВНЫХ КОМПОНЕНТ, ПРИНЯТО БРАТЬ КАК "РАЗМЕР ВЫБОРКИ МИНУС 1".



Шаг 2

Находим матрицу корреляции.

	Лапша	Заправка	Суп
Лапша	1	0,19	0,36
Заправка	0,19	1	0,30
Суп	0,36	0,30	1

Шаг 3

Находим собственные значения λ и собственные векторы $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$, удовлетворяющие

$$\text{соотношению } \begin{pmatrix} 1 & 0,19 & 0,36 \\ 0,19 & 1 & 0,30 \\ 0,36 & 0,30 & 1 \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \lambda \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}.$$

Для каждого собственного вектора обеспечиваем выполнение условия:

$$a_1^2 + a_2^2 + a_3^2 = 1.$$

С помощью программы анализа данных должны быть получены следующие результаты.

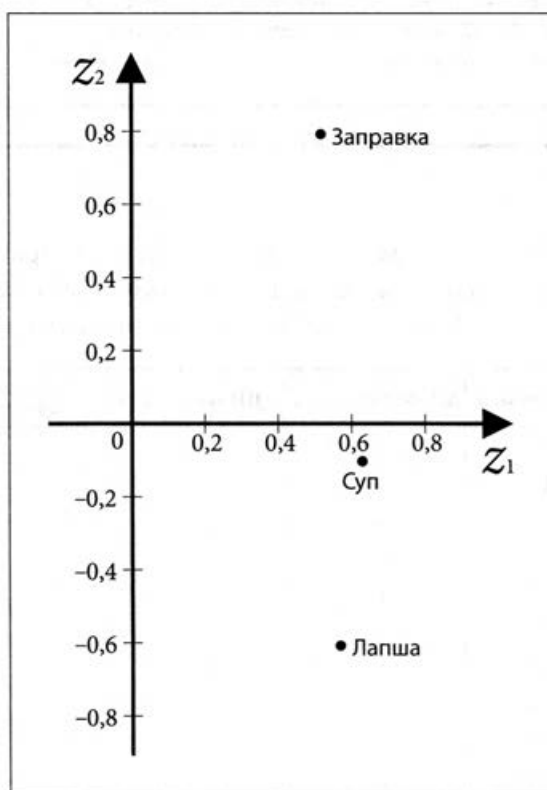
Собственные значения λ	Собственные векторы $\begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix}$
1,6	$\begin{pmatrix} 0,57 \\ 0,52 \\ 0,63 \end{pmatrix}$
0,8	$\begin{pmatrix} -0,60 \\ 0,79 \\ -0,11 \end{pmatrix}$
0,6	$\begin{pmatrix} -0,55 \\ -0,32 \\ 0,77 \end{pmatrix}$

Шаг 4

Строим точечный график на основе:

- собственного вектора, соответствующего максимальному собственному значению;
- собственного вектора, соответствующего второму по величине собственному значению, найденных на прошлом шаге:

	Координаты
Лапша	(0,57 ; -0,60)
Заправка	(0,52 ; 0,79)
Суп	(0,63 ; -0,11)



Шаг 5

Составляем уравнения для первой и второй главных компонент на основе вычислений Шага 4:

$$\begin{aligned} z_1 &= 0,57u_1 + 0,52u_2 + 0,63u_3 \\ z_2 &= -0,60u_1 + 0,79u_2 - 0,11u_3 \end{aligned}$$

↑
Нормированная
переменная
«Лапша»

↑
Нормированная
переменная
«Заправка»

↑
Нормированная
переменная
«Суп»

КОЭФФИЦИЕНТЫ ПЕРВОЙ ГЛАВНОЙ КОМПОНЕНТЫ ОБРАЗУЮТ СОБСТВЕННЫЙ ВЕКТОР, СООТВЕТСТВУЮЩИЙ МАКСИМАЛЬНОМУ СОБСТВЕННОМУ ЗНАЧЕНИЮ. АНАЛОГИЧНО, КОЭФФИЦИЕНТЫ p -ОЙ ГЛАВНОЙ КОМПОНЕНТЫ ОБРАЗУЮТ СОБСТВЕННЫЙ ВЕКТОР, СООТВЕТСТВУЮЩИЙ p -МУ ПО ВЕЛИЧИНЕ СОБСТВЕННОМУ ЗНАЧЕНИЮ.



Шаг 6

Находим координаты каждого экземпляра относительно первой и второй главных компонент. Другими словами, находим для каждого экземпляра значение первой главной компоненты и значение второй главной компоненты.

Название ресторана	НП «лапша», u_1	НП «заправка», u_2	НП «суп», u_3	1-я главная компонента, z_1	2-я главная компонента, z_2
Нираку	-0,7	0,3	1,4	0,7	0,5
Юмэдая	-1,3	1,2	-1,3	-1,0	1,9
Тикай	1,3	-0,5	0,8	1,0	-1,3
нанохана	-0,7	-1,4	0,1	-1,1	-0,7
Ханабуси	0,0	1,2	1,4	1,5	0,8
Сёсинкэн	0,7	-0,5	-0,6	-0,3	-0,7
Марудзо Рамэн	0,7	0,3	0,1	0,6	-0,1
Кайракутэй	-1,3	-1,4	-1,3	-2,3	-0,1
Нарумия	0,0	-0,5	-0,6	-0,7	-0,3
Согэцу	1,3	1,2	0,1	1,4	0,1
СА	0	0	0	0	0
СКО	1	1	1	$\sqrt{1,6}$	$\sqrt{0,8}$

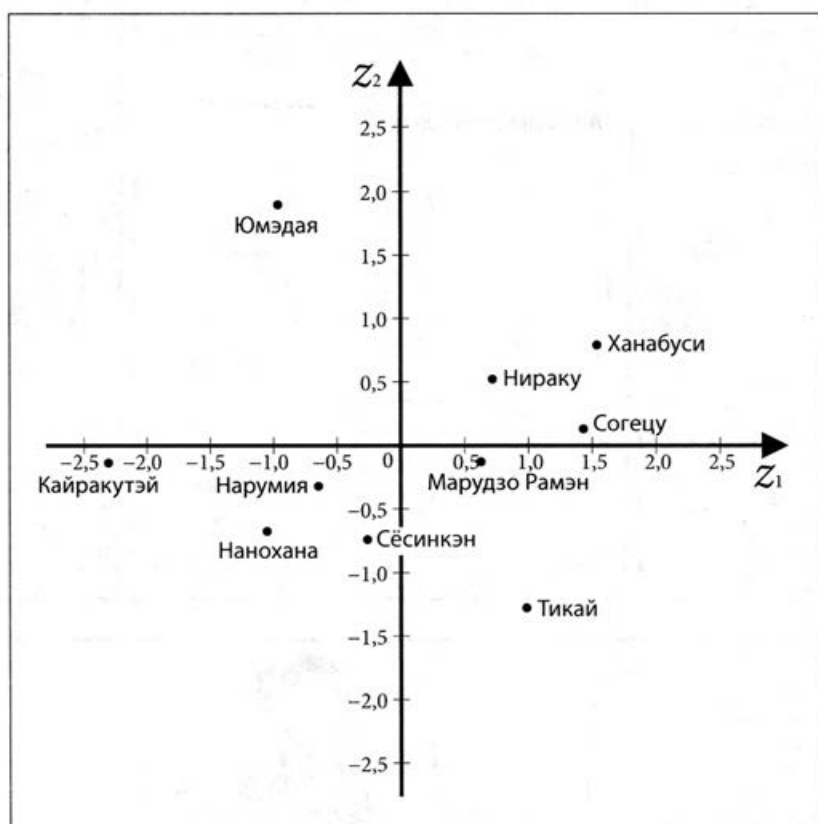
$$0,57 \times 1,3 + 0,52 \times 1,2 + 0,63 \times 0,1 = 1,4$$

↑
√Собственное значение

Шаг 7

Строим точечный график на основе значений первой главной компоненты и значений второй главной компоненты, найденных на Шаге 6.

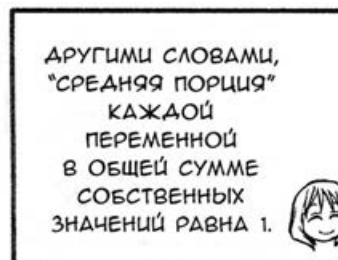
Название ресторана	Координаты
Нираку	(0,7 ; 0,5)
Юмэдая	(-1,0 ; 1,9)
Тикай	(1,0 ; -1,3)
Нанохана	(-1,1 ; -0,7)
Ханабуси	(1,5 ; 0,8)
Сёсинкэн	(-0,3 ; -0,7)
Марудзо Рамэн	(0,6 ; -0,1)
Кайракутэй	(-2,3 ; -0,1)
Нарумия	(-0,7 ; -0,3)
Согецу	(1,4 ; 0,1)



ГОТОВО!



2 ПРОВЕРЯЕМ ТОЧНОСТЬ РЕЗУЛЬТАТОВ АНАЛИЗА



"УДЕЛЬНЫЙ ВКЛАД
i-й ГЛАВНОЙ КОМПОНЕНТЫ"
ВЫЧИСЛЯЕТСЯ ПО СЛЕДУЮЩЕЙ
ФОРМУЛЕ.

ВОТ ЭТО
ДА!

$$\text{Удельный вклад } i\text{-й главной компоненты} = \frac{\lambda_i}{\text{Размер выборки}} \times 100$$

ДАЛЕЕ, СУММАРНЫЙ ВКЛАД
НАХОДИТСЯ ПОСЛЕДОВАТЕЛЬНЫМ
СУММИРОВАНИЕМ УДЕЛЬНЫХ
ВКЛАДОВ ГЛАВНЫХ КОМПОНЕНТ,
НАЧИНАЯ С ПЕРВОЙ.

КРУТО!

	Соб. знач. λ	Удельные вклады	Суммарные вклады
1-я главная компонента	1.6	$\frac{1.6}{3} \times 100 = 52.4 (\%)$ 	$\frac{1.6}{3} \times 100 = 52.4 (\%)$ 
2-я главная компонента	0.8	$\frac{0.8}{3} \times 100 = 27.1 (\%)$ 	$\frac{1.6}{3} \times 100 + \frac{0.8}{3} \times 100 = 79.6 (\%)$ 
3-я главная компонента	0.6	$\frac{0.6}{3} \times 100 = 20.4 (\%)$ 	$\frac{1.6}{3} \times 100 + \frac{0.8}{3} \times 100 + \frac{0.6}{3} \times 100 = 100 (\%)$ 

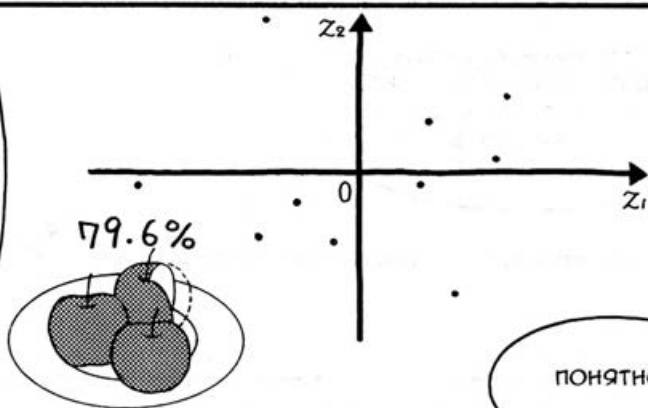
УДЕЛЬНЫЙ ВКЛАД
i-й ГЛАВНОЙ КОМПОНЕНТЫ —
ЭТО КРИТЕРИЙ ТОГО, КАК
МНОГО ИНФОРМАЦИИ
АНАЛИЗИРУЕМЫХ ДАННЫХ
СОСРЕДОТОЧЕНО
В i-й ГЛАВНОЙ
КОМПОНЕНТЕ.

ЧЕМ БОЛЬШЕ
УДЕЛЬНЫЙ ВКЛАД,
ТЕМ БОЛЬШЕ
ИНФОРМАЦИИ
СОСРЕДОТОЧЕНО,
ДА?

ПОЭТОМУ,...

...ТАК КАК В МЕТОДЕ
ГЛАВНЫХ КОМПОНЕНТ
НАХОДЯТ ГЛАВНЫЕ
КОМПОНЕНТЫ ДО ВТОРОЙ
И СТРОЯТ ДЛЯ НИХ
ТОЧЕЧНЫЙ ГРАФИК,

ТО ЧЕМ БОЛЬШЕ
СУММАРНЫЙ ВКЛАД
АВУХ ПЕРВЫХ ГЛАВНЫХ
КОМПОНЕНТ, ТЕМ ЛУЧШЕ.



ПОНЯТНО.

ТОГДА КАКОЕ
ЗНАЧЕНИЕ
МОЖНО
СЧИТАТЬ
КРИТЕРИЕМ
УСПЕШНОГО
АНАЛИЗА?

К СОЖАЛЕНИЮ,
ДОСТОВЕРНОГО
КРИТЕРИЯ
НЕ СУЩЕСТВУЕТ.

КАК?!

НО БУДЕТ НЕПЛОХО, НАВЕРНОЕ,
ЕСЛИ СУММАРНЫЙ ВКЛАД
ПЕРВЫХ ДВУХ КОМПОНЕНТ
СОСТАВИТ ТАК ПРОЦЕНТОВ 50.

В НАШЕМ ПРИМЕРЕ
ОН РАВЕН 79,6%,
ЗНАЧИТ, АНАЛИЗ
МОЖНО СЧИТАТЬ
УСПЕШНЫМ?!

Ура!

3 ПРОВОДИМ РАССМОТРЕНИЕ РЕЗУЛЬТАТОВ АНАЛИЗА.



В МЕТОДЕ ГЛАВНЫХ КОМПОНЕНТ РЕЗУЛЬТАТЫ РАССМАТРИВАЮТ НА ОСНОВАНИИ ЭТИХ ДВУХ ГРАФИКОВ.



График шага 7
(График экземпляров)

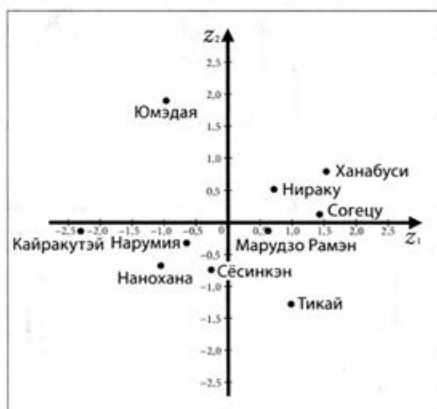
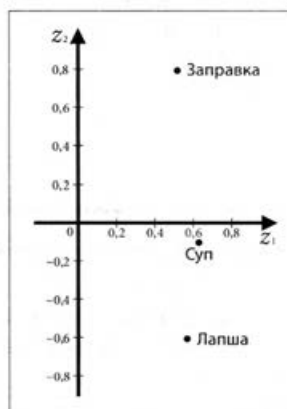
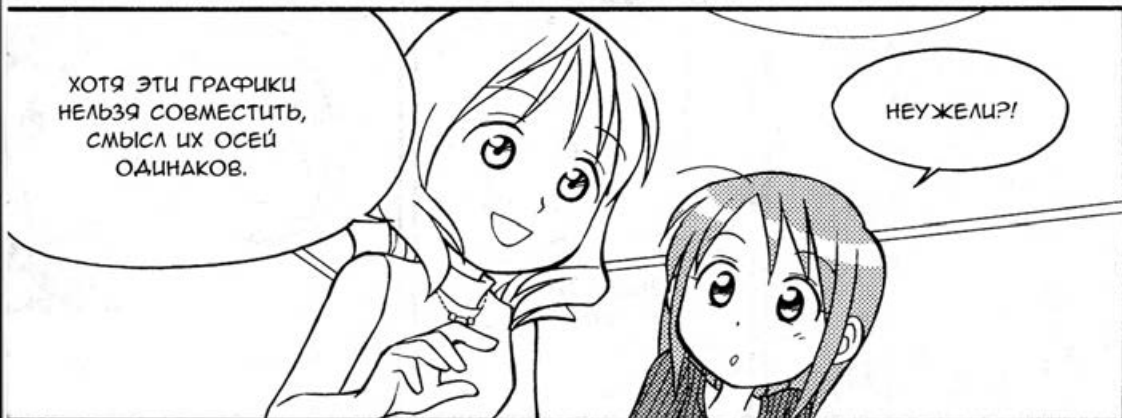
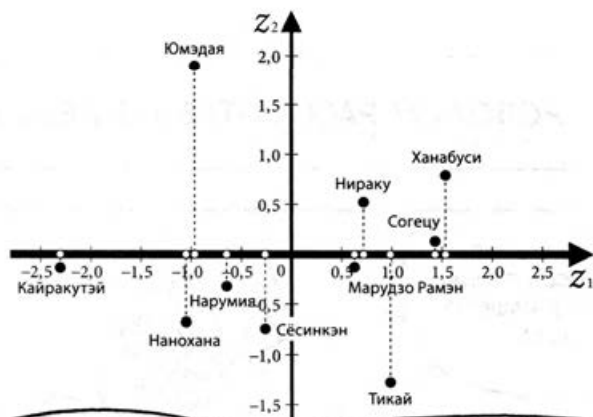


График шага 4
(График переменных)



ХОТЯ ЭТИ ГРАФИКИ НЕЛЬЗЯ СОВМЕСТИТЬ, СМЫСЛ ИХ ОСЕЙ ОДИНАКОВ.



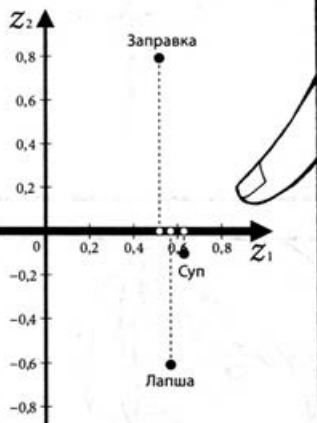


ДРУГИМИ СЛОВАМИ,
ЭТО — ОБЩАЯ
ОЦЕНКА ЛАПШИ
"РАМЭН".

ЗНАЧИТ,
ПЕРВОЕ МЕСТО
ЗАНЯЛ РЕСТОРАН
"ХАНАБУСИ",
А ВТОРОЕ —
"СОГЕЦУ", АА?!

ВЕРНО!

ТЕПЕРЬ РАССМОТРИМ
ГОРИЗОНТАЛЬНУЮ ОСЬ
ГРАФИКА ПЕРЕМЕННЫХ.



КАК ВИДИШЬ,
ЗНАЧЕНИЕ "СУП"
ЗДЕСЬ ЯВЛЯЕТСЯ
МАКСИМАЛЬ-
НЫМ.

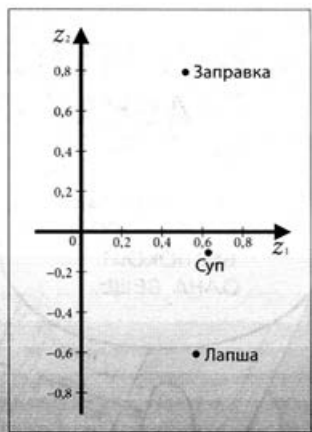
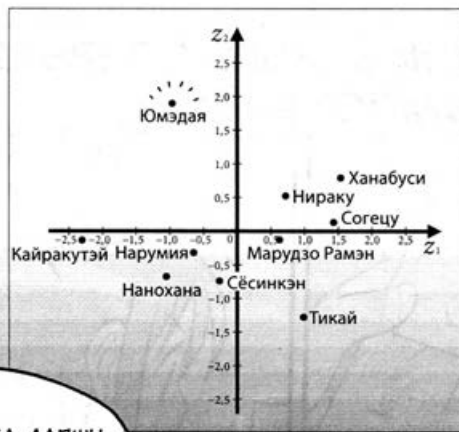
АА.

ЭТО ОЗНАЧАЕТ, ЧТО
МАКСИМАЛЬНОЕ ВЛИЯНИЕ
НА ОБЩУЮ ОЦЕНКУ
ЛАПШИ "РАМЭН"
ОКАЗЫВАЕТ ПЕРЕМЕННАЯ
"СУП".

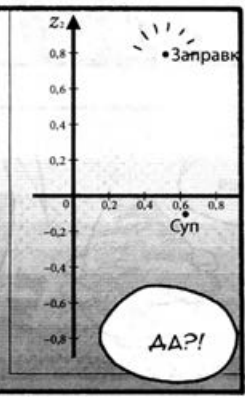
ЗАОРОВО!

В ЗАКЛЮЧЕНИЕ
ПОПРОБУЕМ СРАВНИТЬ
ЭТИ ДВА ГРАФИКА.

ВОТ ТУТ ВВЕРХУ
НАХОДИТСЯ ТОЧКА
ДЛЯ РЕСТОРАНА
"ЮМЭДАЯ", ВИДИШЬ?



А, ВИЖУ. НО ОБЩАЯ ОЦЕНКА ЛАПШИ
"РАМЭН" У НЕГО НЕ ОЧЕНЬ ХОРОШАЯ.



4. ВЫБОР ПЕРЕМЕННЫХ И ПЕРВАЯ ГЛАВНАЯ КОМПОНЕНТА

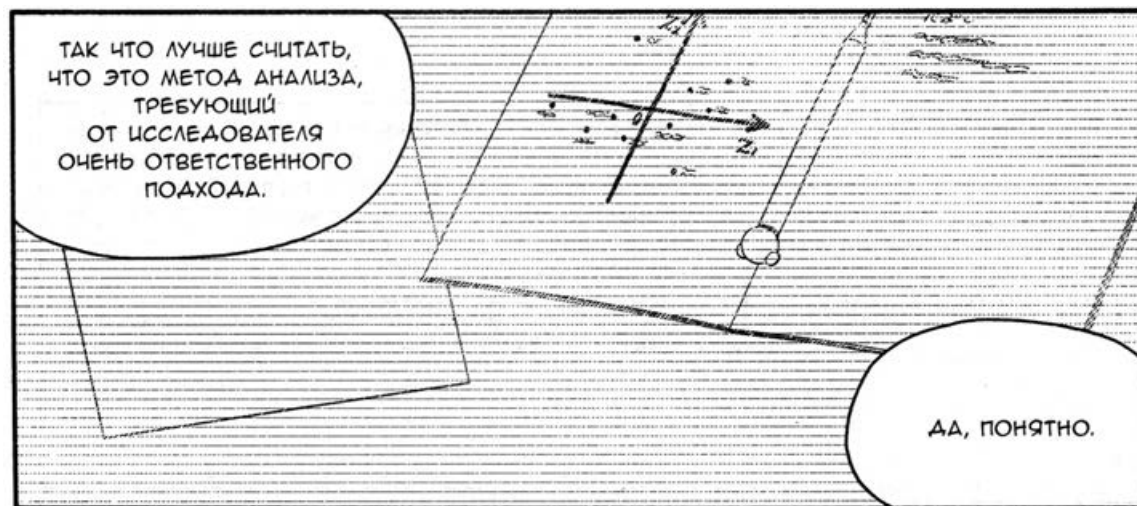
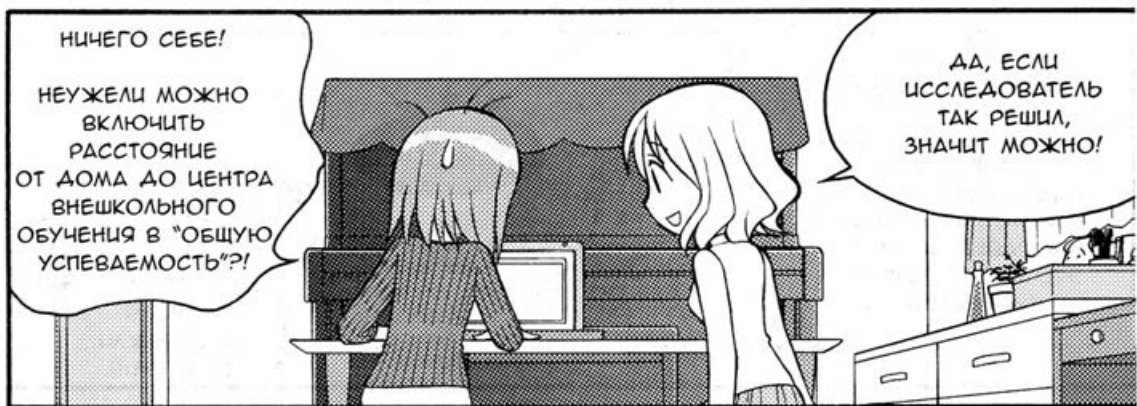


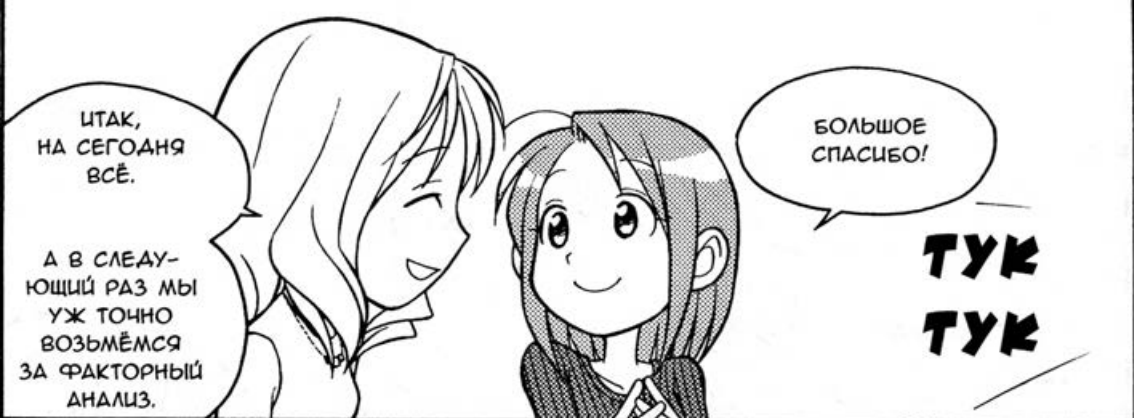


...РАССМОТРИМ СЛЕДУЮЩЕЕ УТВЕРЖДЕНИЕ.

Школьники, живущие недалеко от центров внешкольного обучения, лучше учатся, так как меньше времени тратят на дорогу и у них больше времени на выполнение домашних заданий. Поэтому анализ методом главных компонент должен включать не только предметы «родной язык», «обществоведение», «естествознание», «иностранный» и «математику», но также «число, обратное расстоянию от дома до центра внешкольного обучения», так как это повысит убедительность определения первой главной компоненты как «общей успеваемости».

	Родной язык	Обществоведение	Естествознание	Иностранный язык	Математика	1 / расстоянию от дома до центра внешкольного обучения
ученик Таро	42	62	26	4	20	1/1200
ученица Ханакo	12	28	42	8	84	1/580
⋮	⋮	⋮	⋮	⋮	⋮	⋮





ИТАК,
НА СЕГОДНЯ
ВСЁ.

А В СЛЕДУ-
ЮЩИЙ РАЗ МЫ
УЖ ТОЧНО
ВОЗЬМЁМСЯ
ЗА ФАКТОРНЫЙ
АНАЛИЗ.

БОЛЬШОЕ
СПАСИБО!

ТУК
ТУК



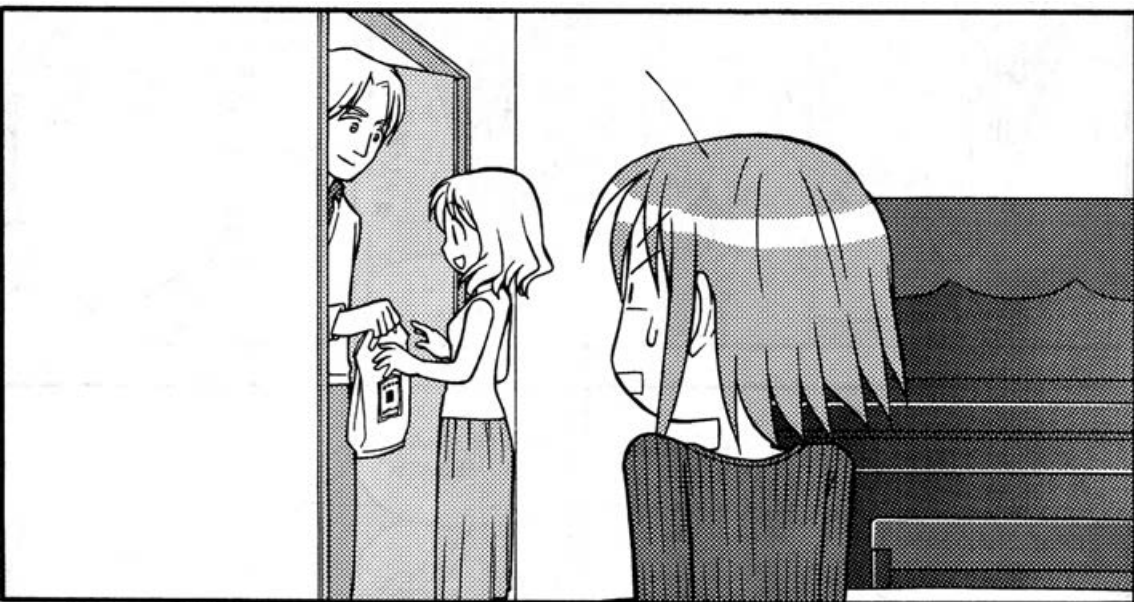
ЭЙ, МУУ...
Я ПРИНЁС
ТЕБЕ
КОЕ-ЧТО.

Ах.

ЭТО МОЙ
БРАТ.

СЕЙЧАС!

ШЁЛК



И...

ИГА...



ИГАРАСУ-САН!

АА?

Что?



ТЕБЯ ЗОВУТ...
ТАК..

Вы что,
знакомы?!



МЕНЯ ЗОВУТ ТА...
ТАКАЦУ РУЦ.



А!
ДОЧЬ ТАКАЦУ,
НАЧАЛЬНИКА
ОТДЕЛА!

ПОМНИШЬ,
Я КАК-ТО
ПРИХОДИЛ
К ВАМ
ДОМОЙ?

АА...



ВОТ ЭТО АА!
А Я И НЕ ЗНАЛА,
ЧТО РУЦ И МОЙ
БРАТ ЗНАКОМЫ!

А Я НЕ ЗНАЛА,
ЧТО МЫ И
ИГАРАСУ-САН —
БРАТ И СЕСТРА!



НУ, НЕ БУДУ
ВАМ МЕШАТЬ.

ПОСТОЙТЕ!



5. ПЕРВАЯ ГЛАВНАЯ КОМПОНЕНТА И ОБЩИЙ ПОКАЗАТЕЛЬ

Из вышесказанного могло сложиться мнение, что первая главная компонента всегда означает общий показатель, но на самом деле это не так.

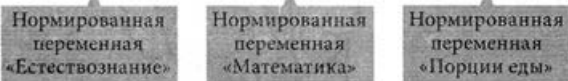
Рассмотрим такой пример. В табл. 4.1 приведены данные тестирования в одной средней школе.

◆ Таблица 4.1. Данные тестирования учеников

Имена учеников	Естествознание (баллов)	Математика (баллов)	Число съеденных порций еды (порций/день)
A	77	82	3
B	68	66	1
C	93	81	2
D	100	92	5
E	75	70	0

При анализе вышеприведённых данных методом главных компонент получается следующее уравнение для первой главной компоненты:

$$z_1 = 0,56u_1 + 0,60u_2 + 0,57u_3.$$



Как вы думаете, какую общую силу выражает эта первая главная компонента? С точки зрения здравого смысла, она не выражает никакого общего показателя.

Приведу ещё один пример. Пусть вы собираетесь выявить «общий показатель спортивных способностей», действуя по следующему плану:

1) сбор различных переменных наобум, не важно каких: «силы захвата» или «любимой телепередачи»;

2) анализ собранных переменных методом главных компонент.

Но будет ли найденная таким образом первая главная компонента выражать «общий показатель спортивных способностей»?

Не кажется ли вам, что только следующий порядок анализа:

1) сбор переменных, которые могут иметь отношение к «общему показателю спортивных способностей»;

2) анализ собранных переменных методом главных компонент; позволит первой главной компоненте быть выражением «общего показателя спортивных способностей».

Подведём итоги. Первая главная компонента, полученная методом главных компонент, не является естественным выражением общего показателя. Первая главная компонента приобретает смысл «общего показателя» только в том случае, когда аналитик, намеренный найти какой-либо «общий показатель», собирает подходящие переменные и применяет в отношении них метод главных компонент. Это можно уподобить овощному рагу. Повар вначале решает приготовить овощное рагу, а затем собирает соответствующие продукты, из которых потом готовит рагу. Если бы он просто брал всё, что попадётся под руку, и варил это, рагу бы не получилось.

Представьте, что вы приготовили рагу из приправы для соуса карри, корейского салата кимчи и крабовых палочек. Однако можно ли будет сказать, что это «рагу»? Вы, конечно, можете убеждать всех: «Это — настоящее рагу!» Однако окружающие не будут воспринимать это как рагу, более того, они станут упрекать вас во лжи. Тогда какие же ингредиенты надо сварить, чтобы получилось настоящее рагу? Хотя всякому ясно, что крабовые палочки тут совсем неуместны, однозначного ответа никто дать не сможет. Наверное, тут всё зависит от повара, его совести и знаний. Эту идею я хотел донести до вас в предыдущем, четвёртом, параграфе с названием «Выбор переменных и первая главная компонента».

6. КРИТЕРИЙ СУММАРНОГО ВКЛАДА

Как я уже заметил, результаты анализа методом главных компонент принято представлять в виде двумерного точечного графика. Следовательно, чем больше будет значение суммарного вклада двух первых главных компонент, тем больше у вас оснований утверждать: «Этот точечный график содержит довольно много информации проанализированных данных». Другими словами, вы сможете считать, что анализ был успешным.

К сожалению, не существует статистического критерия, позволяющего судить об успешности анализа на основе точного значения суммарного вклада двух первых главных компонент.

Критерий «50%», о котором упоминалось на стр. 114 — это всего лишь моё личное мнение, основанное на таком рассуждении: «Трудно сформировать правильное мнение на основе точечного графика, не содержащего и половины информации проанализированных данных».

Следующий рассказ, наверное, повергнет читателя в смятение. Сначала возьмите две переменные, какие угодно, и примените к ним метод главных компонент. Думаю, вы уже поняли, что суммарный вклад двух первых главных компонент в этом случае обязательно окажется равным 100%. Затем примените метод главных компонент к двумстам переменным. Если это не какие-нибудь специфические данные, то суммарный вклад двух первых главных компонент, скорее всего, окажется меньше 50%. Да, в действительности использование критерия «50%» сильно зависит от количества анализируемых данных.

Таким образом, критерий суммарного вклада не является слишком достоверным. Однако это не развязывает вам руки окончательно. Ведь вряд ли кто-нибудь поверит в результаты вашего анализа, если суммарный вклад первых двух главных компонент будет, например, меньше 14%, да и сами вы, наверное, постесняетесь знакомить окружающих с такими результатами.

Я рекомендую вам, учтя вышеизложенное, прилагать усилия, направленные на выработку вашего собственного критерия.

7. ГЛАВНЫЕ КОМПОНЕНТЫ, НАЧИНАЯ СО ВТОРОЙ

Читатели, которым не интересна математика, могут пропустить этот параграф.

Выше говорилось, что главные компоненты, начиная со второй, «находятся автоматически (или математически), помимо воли аналитика». На самом деле это не совсем так.

Например, вторая главная компонента — это ось, которую аналитик должен построить так, чтобы она удовлетворяла следующим условиям:

- была ортогональна первой главной компоненте;
- проходила через область, дисперсия данных в которой вторая по величине после первой главной компоненты¹.

Аналогично, третья главная компонента — это ось, которую аналитик должен построить так, чтобы она удовлетворяла следующим условиям:

- была ортогональна первой и второй главным компонентам;
- проходила через область, дисперсия данных в которой третья по величине после первой и второй главных компонент².

¹ Другими словами вторая главная компонента — это ось, ортогональная оси первой главной компоненты и проходящая через область со второй по величине дисперсией данных.

² Другими словами третья главная компонента — это ось, ортогональная осям первой и второй главных компонент и проходящая через область с третьей по величине дисперсией данных. Аналогично, i -я главная компонента — это ось, ортогональная осям предыдущих главных компонент и проходящая через область со i -й по величине дисперсией данных.

Таким образом, главные компоненты, начиная со второй, не «находятся автоматически (математически), помимо воли аналитика», а «находятся только в результате удовлетворения аналитиком определённых условий».

Некоторые из читателей, возможно, решат, что аналитик, даже не прилагая усилий для удовлетворения условий, может найти главные компоненты, начиная со второй, при помощи одной из программ анализа данных. Это действительно так. Просто программисты заранее позаботились о том, чтобы облегчить нам труд, заложив вышеупомянутые условия в свои программы.

8. ДИСПЕРСИЯ И СОБСТВЕННЫЕ ЗНАЧЕНИЯ

Читатели, которым не интересна математика, могут пропустить этот параграф.

Выше говорилось, что первая главная компонента — это ось, проходящая через область, в которой дисперсия данных максимальна. Помня об этом, посмотрите ещё раз на процесс вычисления главных компонент на стр. 107-111. Среди этих вычислений нет ни одного, связанного с дисперсией данных. Вместо этого почему-то упоминаются собственные векторы и собственные значения.

Опустив подробное математическое доказательство, я замечу, что «нахождение оси, проходящей через область максимальной дисперсии данных» и «нахождение максимального собственного значения и соответствующего ему собственного вектора для матрицы корреляции» — это одно и то же.

Аналогично, «нахождение оси, проходящей через область с i -й по величине дисперсией данных» и «нахождение i -го по величине собственного значения и соответствующего ему собственного вектора для матрицы корреляции» — это одно и то же.

5

ФАКТОРНЫЙ АНАЛИЗ

1. ЧТО ТАКОЕ ФАКТОРНЫЙ АНАЛИЗ
2. ВАЖНЫЕ МОМЕНТЫ ФАКТОРНОГО АНАЛИЗА
3. КОНКРЕТНЫЙ ПРИМЕР ФАКТОРНОГО АНАЛИЗА
4. ВЫБОРКА В ПРИМЕРЕ ДАННОЙ ГЛАВЫ
5. ДОПОЛНЕНИЯ К ВАЖНЫМ МОМЕНТАМ
6. ПЕРЕМЕННЫЕ С МАЛЫМИ ЗНАЧЕНИЯМИ ФАКТОРНЫХ НАГРУЗОК
7. МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ
8. ПОЧЕМУ ВРАЩЕНИЕ ПОДРАЗУМЕВАЕТ МЕТОД ВАРИМАКС
9. МАТРИЦА ФАКТОРНЫХ НАГРУЗОК И МАТРИЦА ФАКТОРНОЙ СТРУКТУРЫ
10. МЕТОД ПРОМАКС
11. ВЕРХНИЙ ПРЕДЕЛ ЗАДАВАЕМОГО ЧИСЛА ОБЩИХ ФАКТОРОВ
12. СТОИТ ЛИ СЧИТАТЬ МЕТОД ГЛАВНЫХ ФАКТОРОВ
И МЕТОД ВАРИМАКС УСТАРЕВШИМИ?
13. ТЕРМИНОЛОГИЯ ФАКТОРНОГО АНАЛИЗА

ТЩЕК ТЩЕК

ЯМАМОТО-САН...

Лектор

ЗАДОРОВ ЛИ ТЫ?

НАВЕРНОЕ,
ОЧЕНЬ ЗАНЯТ?

ПИТАЕШЬСЯ ЛИ
НОРМАЛЬНО?

Я БЕСПОКОЮСЬ...

Ах!



НА ЗАВТРА

СЕГОДНЯ МЫ НАКОНЕЦ
ПРИСТУПИМ К ТВОЕМУ
ДОЛГОЖДАННОМУ
ФАКТОРНОМУ АНАЛИЗУ!

ПОЖАЛУЙСТА!

Поклон

СЕГОДНЯШНИЙ УРОК
БУДЕТ ПОСВЯЩЁН
ФАКТОРНОМУ АНАЛИЗУ,

НО ПО СЛУЧАЮ
Я ДУМАЮ ТАКЖЕ
ИСПРАВИТЬ АНКЕТУ
КАФЕ NORNS, КОТОРУЮ
ТЫ, РУЦ, ПОКАЗЫВАЛА
НАМ ТОГАА.

ЗАТЕМ МЫ С ЕЁ
ПОМОЩЬЮ ПРОВЕДЁМ
АНКЕТИРОВАНИЕ И
ПОПРОБУЕМ ПРИМЕНИТЬ
ФАКТОРНЫЙ АНАЛИЗ
К ПОЛУЧЕННЫМ
ДАННЫМ.

ХОРОШО!

ЭТО —
ПОСЛЕДНЯЯ
ТЕМА, ПОЭТОМУ
ДАВАЙ
КАК СЛЕДУЕТ
ПОСТАРАЕМСЯ.

ДАВАЙ!

ДЛЯ НАЧАЛА,
ПОСМОТРИ ВОТ
ЭТОТ ЖУРНАЛ.

ШУРХ

Кан-найти
Работу

1. ЧТО ТАКОЕ ФАКТОРНЫЙ АНАЛИЗ

ТУТ ЕСТЬ ДАННЫЕ,
КОТОРЫЕ ОЧЕНЬ
ДЛЯ НАС ПОДХОДЯТ.

Как найти работу

Информационный
журнал по тридо-
устройству?

Респонденты	Известная компания	Хорошая система обучения новых сотрудников	Активно поручает работу молодым сотрудникам	Возможность получения специальных знаний и навыков	Имеет богатые традиции	Разрешает сотрудникам продолжать образование, в том числе за границей	Перспективная компания
А.	2	5	1	5	1	5	2
Б.	3	4	2	5	3	4	1
В.	4	1	1	2	5	3	2
Г.	4	1	3	2	5	3	4
Д.	1	2	5	1	2	1	4
Е.	5	1	1	1	4	2	2
Ж.	4	1	1	2	3	2	2
З.	3	3	3	4	4	5	4
И.	3	2	4	3	5	3	5
К.	3	1	2	2	4	3	3
Л.	2	2	3	2	3	1	1
М.	4	3	2	3	5	3	3
Н.	2	1	1	2	3	3	1
О.	3	1	1	1	4	2	2
П.	3	3	2	4	4	5	3

ВЗГЛЯНИ
НА ЭТО.

ЭТА ТАБЛИЦА СОДЕРЖИТ РЕЗУЛЬТАТЫ ОПРОСА СТУДЕНТОВ ТРЕТЬИХ КУРСОВ С ВОПРОСОМ: "КАКИЕ КАЧЕСТВА КОМПАНИЙ ВАЖНЫ ДЛЯ ВАС ПРИ ВЫБОРЕ МЕСТА РАБОТЫ?" И ПЯТЬЮ ГРАДАЦИЯМИ ОТВЕТОВ НА НЕГО.

Чем больше значение, тем важнее для респондента данный пункт.

ВЗГЛЯНИ НА ДАННЫЕ
РЕСПОНДЕНТА А.

ДЛЯ НЕГО ВАЖНЫ ПУНКТЫ,
КАСАЮЩИЕСЯ
ВОЗМОЖНОСТИ РАЗВИВАТЬ
ЛИЧНЫЕ СПОСОБНОСТИ.

ДА, МОЖНО
СКАЗАТЬ ТАК.

Это относится также
к респонденту Б.

Возможность
получения
специальных
знаний и
навыков.

Хорошая
система
подготовки
новых
сотруд-
ников.

Разрешает сотрудни-
кам продолжать об-
разование, в том
числе за границей.

ТЕПЕРЬ, ЧТО ТЫ
ДУМАЕШЬ ПО ПОВОДУ
РЕСПОНДЕНТА В.?

Имеет
богатые
традиции

Известная
компания

ТАК...
КАЖЕТСЯ, ЭТОТ РЕСПОНДЕНТ
ОТДАЁТ ПРЕДПОЧТЕНИЕ
КАЧЕСТВАМ, ХАРАКТЕРНЫМ
ДЛЯ БОЛЬШИХ И СТАБИЛЬНЫХ
КОМПАНИЙ.

То же самое можно сказать
и про респондента Г.

ТОГДА ЧТО
НАСЧЁТ РЕСПОН-
ДЕНТКИ Д.?

ПОХОЖЕ, ЕЙ НРАВЯТСЯ
КОМПАНИИ, В КОТОРЫХ
МОЖНО БЫСТРО
ДОСТИЧЬ УСПЕХА.

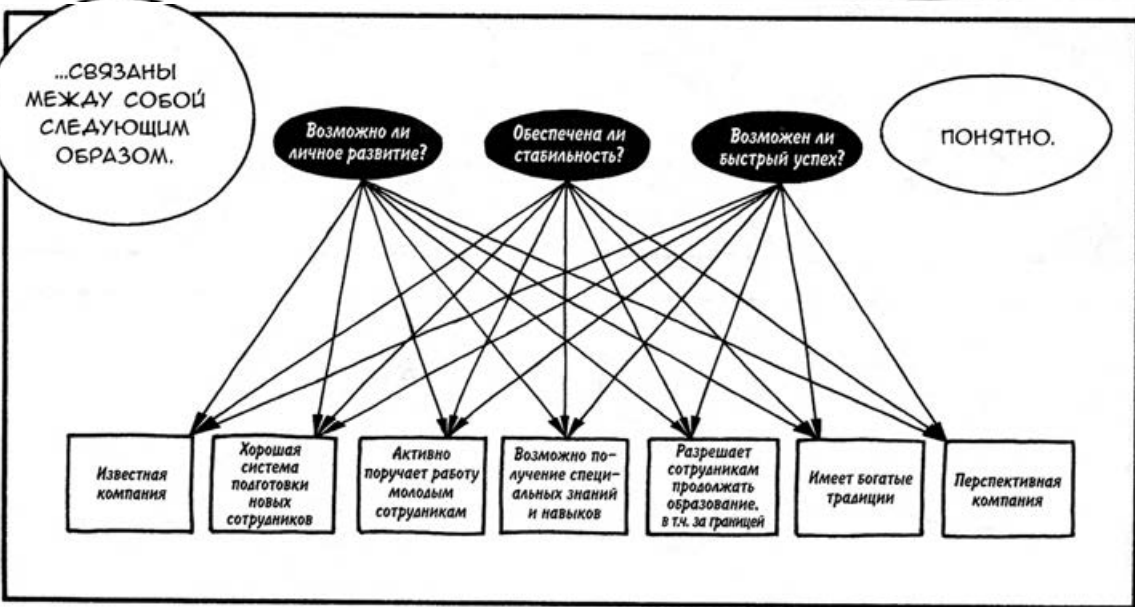
Активно поручает
работу молодым
сотрудникам.

Перспек-
тивная
компания

МОЛОДЕЦ! ТЫ
УМЕЕШЬ ПРАВИЛЬНО
ОЦЕНИВАТЬ ДАННЫЕ!

Наверное,
и Ямамото-сяна
научилась?

ДА НЕТ, НИЧЕГО
ОСОБЕННОГО...
ХЕ-ХЕ...



ЭТО — РЕЗУЛЬТАТЫ ТЕСТИРОВАНИЯ
УЧЕНИКОВ ТРЕТЬЕГО КЛАССА СРЕДНЕЙ
ШКОЛЫ, ЗАНИМАЮЩИХСЯ В ОДНОМ
ИЗ ЦЕНТРОВ ВНЕШКОЛЬНОГО
ОБРАЗОВАНИЯ.

ЯСНО.

ПРИВЕДУ ЕЩЁ
ОДИН ПРИМЕР.

Ученики	Родной язык	Обществове- дение	Естествозна- ние	Иностранный язык	Математика
A	93	100	89	84	77
B	100	98	89	95	86
C	84	84	99	85	100
D	70	73	92	66	77
E	70	72	89	66	75
F	66	68	95	57	82
G	74	70	96	93	88
H	74	75	95	70	79
I	76	77	92	78	83
J	79	88	100	86	100

ПОПРОБУЮ ИЗУЧИТЬ
ЭТИ ДАННЫЕ,
КАК ТЫ ДЕЛАЛА
ТОЛЬКО ЧТО.

ТАК...
С ЧЕГО БЫ
ЛУЧШЕ НАЧАТЬ?

Я ВИЖУ, ЧТО
УЧЕНИЦА А. И УЧЕНИК В.
ИМЕЮТ ВЫСОКИЕ БАЛЛЫ
ПО ГУМАНИТАРНЫМ
ДИСЦИПЛИНАМ.

РАССУЖДАЯ, КАК
В ПРЕДЫДУЩЕМ
ПРИМЕРЕ, МОЖНО
СКАЗАТЬ, ЧТО..

... ЗАДЕСЬ ОКАЗЫВАЮТ
ВЛИЯНИЕ ДВА ФАКТОРА:
- УРОВЕНЬ ГУМАНИТАР-
НЫХ СПОСОБНОСТЕЙ
- УРОВЕНЬ ЕСТЕСТВЕННО-
НАУЧНЫХ СПОСОБНОСТЕЙ,
ПРИЧЁМ СТЕПЕНЬ
ВЛИЯНИЯ КАЖДОГО ИЗ
ФАКТОРОВ РАЗЛИЧНА
ДЛЯ РАЗНЫХ УЧЕНИКОВ.

А ЕЩЁ?

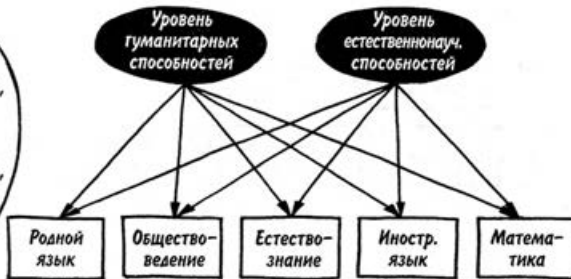
УЧЕНИК С. И
УЧЕНИЦА J. ИМЕЮТ
ВЫСОКИЕ БАЛЛЫ
ПО ЕСТЕСТВЕННЫМ
ДИСЦИПЛИНАМ.



TH — гумани-
тарные
науки
TN — точные
науки

ПОНЯТНО.

ДРУГИМИ СЛОВАМИ,
МОЖНО СКАЗАТЬ,
ЧТО СУЩЕСТВУЮТ
СЛЕДУЮЩИЕ СВЯЗИ,
НЕ ТАК ЛИ?



НО МЕНЯ
БЕСПОКОИТ
ОДИН МОМЕНТ.

А ЧТО
ИМЕННО?

АГА.

ТО, ЧТО ВСЕ ЭТИ МЫСЛИ:
"УРОВЕНЬ ГУМАНИТАРНЫХ
СПОСОБНОСТЕЙ", "УРОВЕНЬ
ЕСТЕСТВЕННОНАУЧНЫХ
СПОСОБНОСТЕЙ",
О КОТОРЫХ ТЫ
ГОВОРИШЬ,...

...НА САМОМ ДЕЛЕ
СУЩЕСТВУЮТ
ТОЛЬКО В НАШЕМ
ВООБРАЖЕНИИ.

ДА, ВСЕ ОНИ —
ПЛОДЫ НАШЕГО
ВООБРАЖЕНИЯ.

НО МОЖНО ЛИ ТОГДА
ИСПОЛЬЗОВАТЬ ИХ
ДЛЯ ИНТЕРПРЕТАЦИИ
ДАННЫХ?

КОНЕЧНО.
ВЕДЬ ЕСЛИ
ПРЕДПОЛОЖИТЬ,
ЧТО ЭТО — СКРЫТЫЕ
ПРИЧИНЫ, СТОЯЩИЕ
ЗА ДАННЫМИ,
ТО МОЖНО КРАСИВО
ОБЪЯСНИТЬ ДАННЫЕ.

Общественные	Естественнонаучные	
100	80	84
98	89	95
84	99	85
73	92	66
64	80	66
64	95	57
50	70	51

Аксиомы по-разному решают проблему, формулировки	Возможные причины спонтанных действий и поведения	Сложные процессы трансформации	Разрешает структурные проблемы, проблемы, проблемы
1	5	1	5
2	5	2	5
3	2	1	5
4	2	2	5
5	1	3	5
6	1	3	5
7	1	3	5
8	1	3	5
9	1	3	5
10	1	3	5

ЭТО, КОНЕЧНО,
ПОНЯТНО...

я хочу,
чтобы ты, руц,
посмотрела на это
под другим углом.

?

попробуй представить,
что эти мысли:
"уровень гуманитарных
способностей" и "уровень
естественнонаучных
способностей" на самом
деле не плоды нашего
воображения,...

...а реально существующие
в этом мире объясняющие
переменные, которые
скрываются за данными.

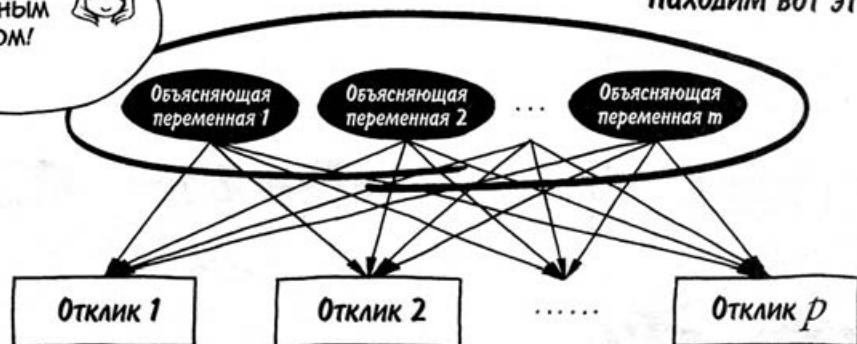
ведь метод анализа,
предназначенный
для нахождения этих
объясняющих переменных,
скрывающихся
за данными — ...

шурх-шурх

...называется...

...факторным
анализом!

Находим вот это!



Классно!!

1. ЧТО ТАКОЕ ФАКТОРНЫЙ АНАЛИЗ

КСТАТИ, i -Ю
ОБЪЯСНЯЮЩУЮ
ПЕРЕМЕННУЮ НАЗЫВАЮТ
 i -М ОБЩИМ ФАКТОРОМ,
 i -М ФАКТОРОМ ИЛИ
ФАКТОРОМ i .

1-й общий фактор

2-й общий фактор

Уровень гуманитарных способностей

Уровень естественнонаучн. способностей

КОНКРЕТНЫЕ ЗНАЧЕНИЯ
ОБЩИХ ФАКТОРОВ
ДЛЯ КАЖДОГО
ИЗ ЭКЗЕМПЛЯРОВ НАЗЫВАЮТ
ФАКТОРНЫМИ ЗНАЧЕНИЯМИ.

ПОНЯТНО.

Имена учеников	Уровень гуманитарных способностей	Уровень естественнонаучных способностей
A	XX	XX
B	XX	XX
C	XX	XX

СПРОШУ НА ВСЯКИЙ СЛУЧАЙ:
ПОНЯТНО ЛИ ТЕБЕ,
ЧТО ФАКТОРНЫЙ АНАЛИЗ
И МЕТОД ГЛАВНЫХ
КОМПОНЕНТ —
РАЗНЫЕ МЕТОДЫ?

АА!

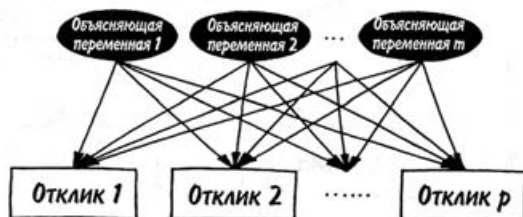
ВЕДЬ ЦАПИ И СХЕМЫ
ЭТИХ МЕТОДОВ
СОВЕРШЕННО РАЗНЫЕ!



Метод главных компонент



Факторный анализ



ВАЖНЫЕ МОМЕНТЫ ФАКТОРНОГО АНАЛИЗА



Первая главная компонента	Общий показатель
Другие главные компоненты	Находятся автоматически (математически), помимо воли аналитика



Факторный анализ
Важный момент **2**

2-й момент.

В МЕТОДЕ ГЛАВНЫХ КОМПОНЕНТ СУЩЕСТВОВАЛА ЦЕРАРХИЯ: СНАЧАЛА — ПЕРВАЯ ГЛАВНАЯ КОМПОНЕНТА, ЗА НЕЙ — ОСТАЛЬНЫЕ.

1-я главная компонента

2-я главная компонента

ОДНАКО В ФАКТОРНОМ АНАЛИЗЕ ТАКОЙ ЦЕРАРХИИ НЕ СУЩЕСТВУЕТ: ВСЕ ОБЩИЕ ФАКТОРЫ РАВНОЗНАЧНЫ.

ЭТО Я ТОЖЕ ПОНЯЛА ИЗ ПРЕДЫДУЩЕГО ПРИМЕРА!

Факторный анализ
Важный момент **3**

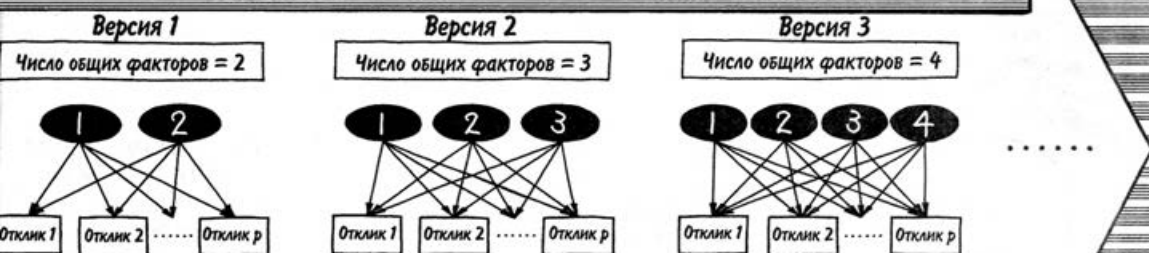
3-й момент.

К СОЖАЛЕНИЮ, ДЛЯ ВЫЧИСЛЕНИЙ НЕОБХОДИМО, ЧТОБЫ АНАЛИТИК ПЕРЕД ПРОВЕДЕНИЕМ ФАКТОРНОГО АНАЛИЗА ВЫДВИНУЛ ВЕРСИЮ О ЧИСЛЕ ОБЩИХ ФАКТОРОВ.

КАК?!

НО КАК УЗНАТЬ ЧИСЛО ОБЩИХ ФАКТОРОВ ПЕРЕД ПРОВЕДЕНИЕМ АНАЛИЗА? ВЕДЬ ОНИ ЖЕ СКРЫТЫЕ!

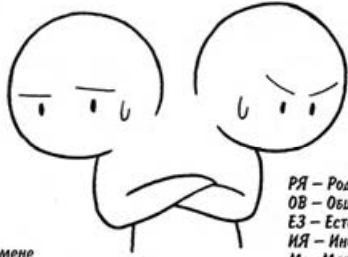
ЧТО ЖЕ ДЕЛАТЬ?!



4-й момент.
ОН ИМЕЕТ ОТНОШЕНИЕ
К ПРЕДЫДУЩЕМУ
МОМЕНТУ.

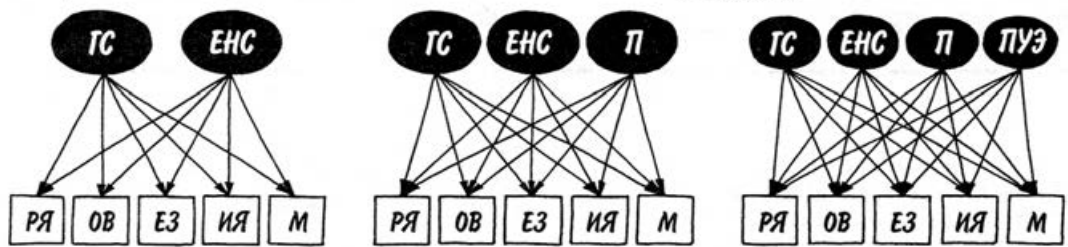
Для всех версий результаты
получаются неплохие...

ХОТЯ АНАЛИЗИРУЮТСЯ
ОДНИ И ТЕ ЖЕ ДАННЫЕ,
В НЕКОТОРЫХ СЛУЧАЯХ
РЕЗУЛЬТАТЫ АНАЛИЗА
ОКАЗЫВАЮТСЯ
УДОВЛЕТВОРИТЕЛЬНЫМИ
И ДЛЯ ДВУХ, И ДЛЯ ТРЁХ,
И ДЛЯ ДРУГОГО ЧИСЛА
ОБЩИХ ФАКТОРОВ.

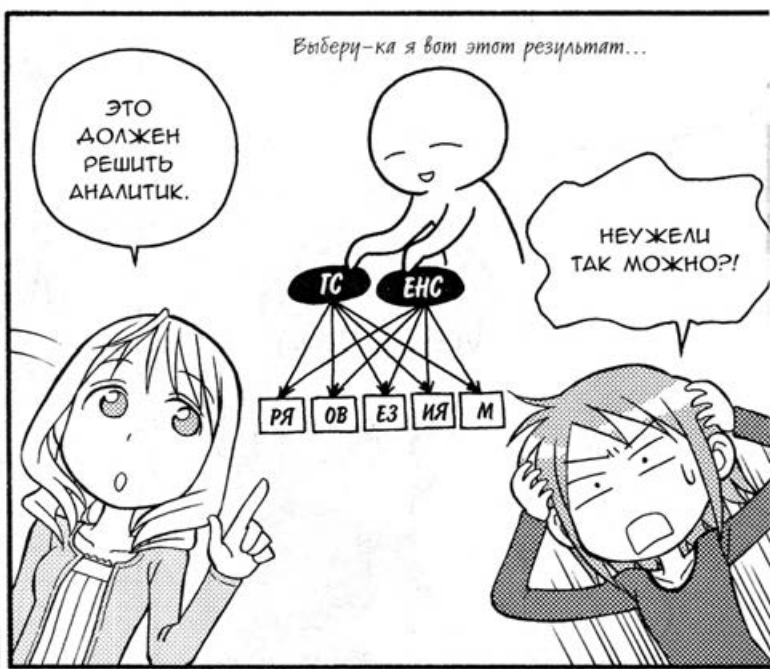


- ТС – Уровень гуманитарных способностей
- ЕНС – Уровень естественнонаучных способностей
- П – Память
- ПУЭ – Уровень психологической устойчивости на экзамене

- РЯ – Родной язык
- ОВ – Обществоведение
- ЕЗ – Естествознание
- ИЯ – Иностраннный язык
- М – Математика



ПОНЯТНО.
НО ЧТО ТОГДА
СЧИТАТЬ
ОКОНЧАТЕЛЬНЫМ
РЕЗУЛЬТАТОМ?



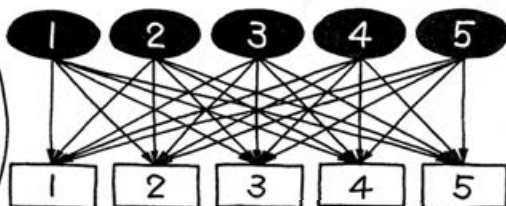
ЭТО
ДОЛЖЕН
РЕШИТЬ
АНАЛИТИК.

Выберу-ка я вот этот результат...

НЕУЖЕЛИ
ТАК МОЖНО?!

5-й МОМЕНТ.

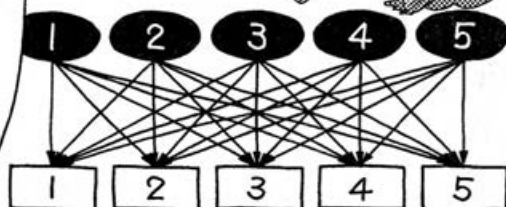
ПРИ ИСПОЛЬЗОВАНИИ НЕКОТОРЫХ МЕТОДОВ ВЫЧИСЛЕНИЙ ЧИСЛО НАЙДЕННЫХ ОБЩИХ ФАКТОРОВ НЕ МОЖЕТ ПРЕВЫСИТЬ ЧИСЛО ОТКЛИКОВ...



...ДАЖЕ ЕСЛИ НА САМОМ ДЕЛЕ В ДАННЫХ СКРЫТО ГОРАЗДО БОЛЬШЕ ОБЩИХ ФАКТОРОВ.

ВОТ ОНО ЧТО?!

ПОЭТОМУ АНАЛИТИК ДОЛЖЕН ВЫБРАТЬ "НАСТОЯЩИЕ ОБЩИЕ ФАКТОРЫ" И ОТМЕСТИ "ВСЁ ОСТАЛЬНОЕ".



НИЧЕГО СЕБЕ!

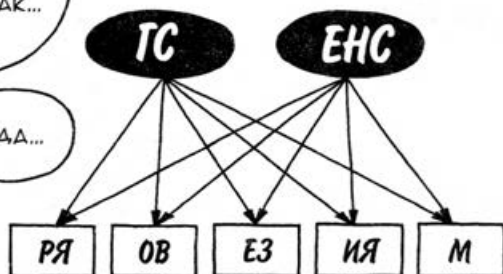
6-й МОМЕНТ.

ОН СВЯЗАН С ПРЕДЫДУЩИМ МОМЕНТОМ.

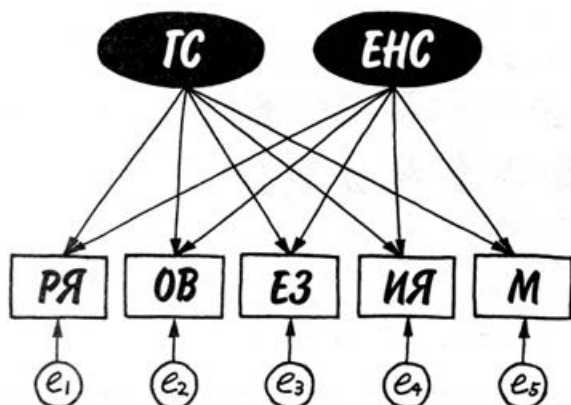
Всё остальное

РАНЬШЕ Я РИСОВАЛА СХЕМУ ФАКТОРНОГО АНАЛИЗА ВОТ ТАК...

ДА...



Всё остальное



ОДНАКО НА САМОМ ДЕЛЕ ПРАВИЛЬНО РИСОВАТЬ ЕЁ ВОТ ТАК, УЧИТЫВАЯ "ВСЁ ОСТАЛЬНОЕ"



ОГО!

ШУРХ

e_i — ЭТО И ЕСТЬ "ВСЁ ОСТАЛЬНОЕ"?



ДА!
 e_i НАЗЫВАЮТ СПЕЦИФИЧЕСКИМИ ФАКТОРАМИ.



Факторный анализ
Важный момент **7**

7-й МОМЕНТ.

В ФАКТОРНОМ АНАЛИЗЕ ПЕРЕД ВЫЧИСЛЕНИЯМИ ОБЫЧНО ПРОВОДИТСЯ НОРМИРОВКА АНАЛИЗИРУЕМЫХ ДАННЫХ ПО КАЖДОЙ ИЗ ПЕРЕМЕННЫХ.



ПОНЯТНО.



ПОЭТОМУ В ДАЛЬНЕЙШЕМ Я БУДУ ИСХОДИТЬ ИЗ ЭТОГО.

8-й МОМЕНТ.



ВЫРАЖАЯ СТРУКТУРУ ФАКТОРНОГО АНАЛИЗА С ПОМОЩЬЮ ФОРМУЛЫ И СХЕМЫ,...



...МЫ ПОЛУЧИМ ВОТ ЧТО.

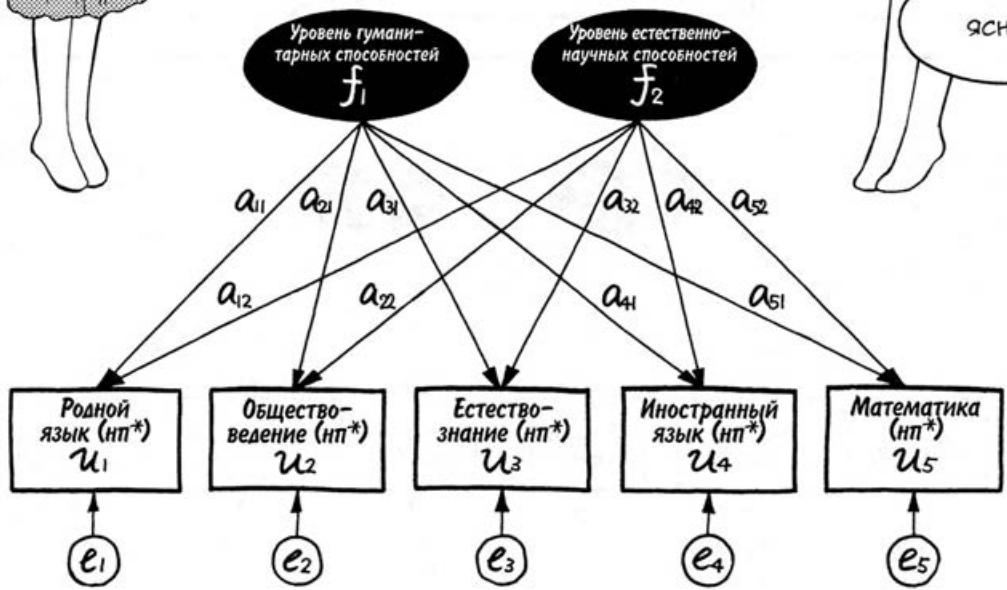
$$\begin{cases}
 u_1 = a_{11} f_1 + a_{12} f_2 + e_1 \\
 u_2 = a_{21} f_1 + a_{22} f_2 + e_2 \\
 u_3 = a_{31} f_1 + a_{32} f_2 + e_3 \\
 u_4 = a_{41} f_1 + a_{42} f_2 + e_4 \\
 u_5 = a_{51} f_1 + a_{52} f_2 + e_5
 \end{cases}$$

↑ 1-й общий фактор ↑ 2-й общий фактор ↑ Специфические факторы

Факторные нагрузки



ЯСНО!



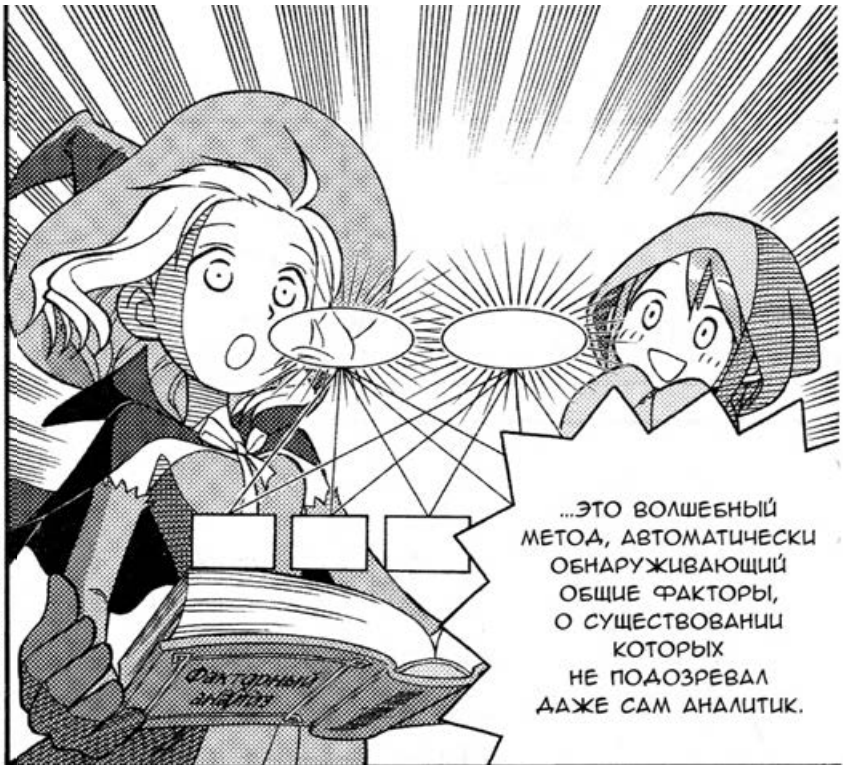
*нп – нормированная переменная

Факторный анализ
Важный момент 9

ШУРХ

Факторный анализ

9-й момент.
ФАКТОРНЫЙ АНАЛИЗ —...



...ЭТО ВОЛШЕБНЫЙ МЕТОД, АВТОМАТИЧЕСКИ ОБНАРУЖИВАЮЩИЙ ОБЩИЕ ФАКТОРЫ, О СУЩЕСТВОВАНИИ КОТОРЫХ НЕ ПОДОЗРЕВАЛ ДАЖЕ САМ АНАЛИТИК.

хлоп

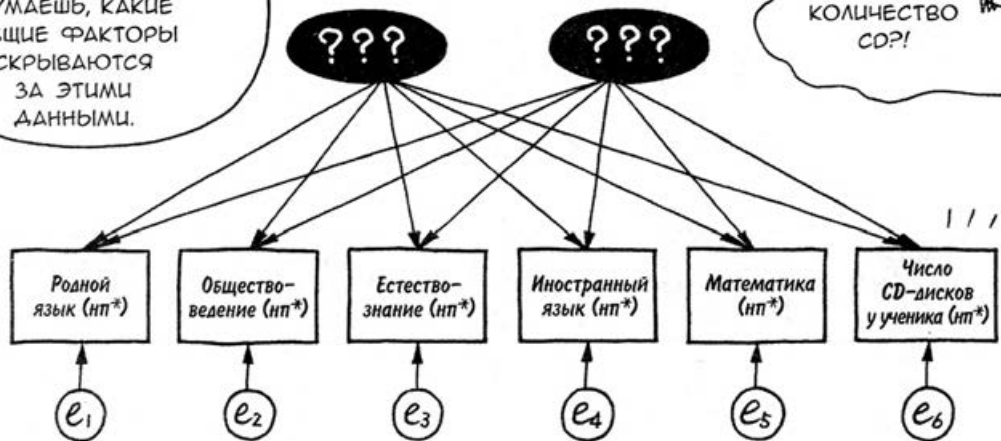
ДУМАТЬ ТАК — ОШИБКА.

КАК, ОШИБКА?!

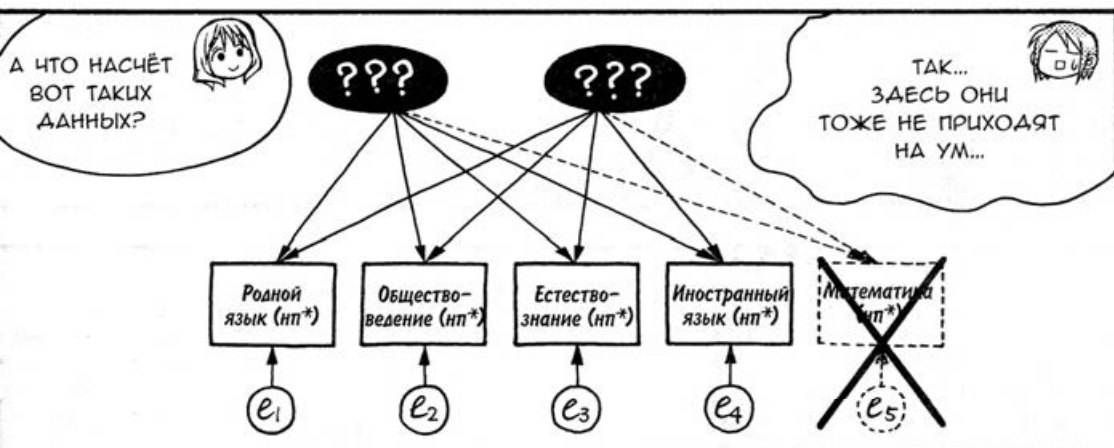
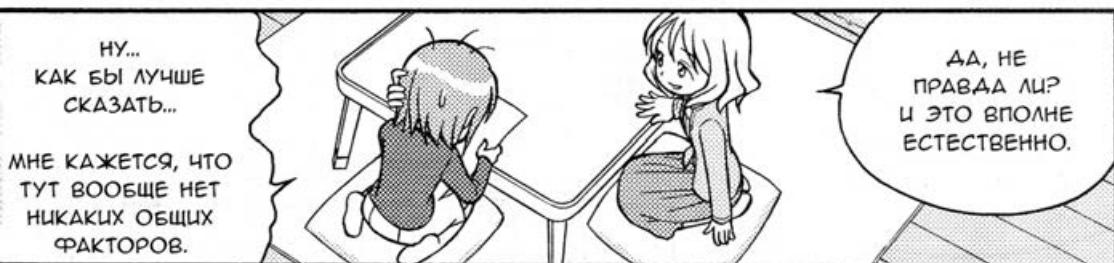
хлоп

НАПРИМЕР, КАК ТЫ ДУМАЕШЬ, КАКИЕ ОБЩИЕ ФАКТОРЫ СКРЫВАЮТСЯ ЗА ЭТИМИ ДАННЫМИ.

ЧТО?! КОЛИЧЕСТВО CD?!



*нп — нормированная переменная



*нп – нормированная переменная



ДРУГИМИ СЛОВАМИ,
ДЛЯ УСПЕШНОГО
ФАКТОРНОГО АНАЛИЗА
НЕОБХОДИМО СНАЧАЛА
В НЕКОТОРОЙ МЕРЕ
ПОСТРОИТЬ ГИПОТЕЗУ...



Известная компания	Хорошая система подготовки новых сотрудников	Активно поручает работу молодым сотрудникам	Возможно получение специальных знаний и навыков	Разрешает сотрудникам продолжать образование	Имеет богатые традиции	Перспективная компания
--------------------	--	---	---	--	------------------------	------------------------

...ЗА ЭТИМИ ОТКЛИКАМИ,
ВОЗМОЖНО, СКРЫВАЮТСЯ ВОТ
ТАКЕ ОБЩИЕ ФАКТОРЫ."

ДА...
НО ВЕДЬ ФАКТОРНЫЙ
АНАЛИЗ ПОТОМУ
И ПРОВОДЯТ, ЧТО
НЕ ЗНАЮТ, КАКИЕ ОБЩИЕ
ФАКТОРЫ СКРЫВАЮТСЯ
ЗА ДАННЫМИ?

НЕужели анализ
НЕ БУДЕТ УСПЕШНЫМ
БЕЗ ПРЕДВАРИТЕЛЬНОГО
ПРЕДСТАВЛЕНИЯ
О ЕГО РЕЗУЛЬТАТАХ?

ЛИЧНО МНЕ ЭТО
КАЖЕТСЯ ДОВОЛЬНО
СТРАННЫМ...

Факторный анализ
Важный момент 10

ИТАК, ПОСЛЕДНИЙ
ВАЖНЫЙ МОМЕНТ!

НА САМОМ ДЕЛЕ
ФАКТОРНЫЙ АНАЛИЗ —
ВО ВСЕ НЕ МЕТОД
ДЛЯ ОБНАРУЖЕНИЯ
СКРЫТЫХ ОБЩИХ
ФАКТОРОВ.

ЧТО?!

ЧТО?!

ТОГДА ЧТО ЖЕ ЭТО ТАКОЕ?!

$$\begin{cases} u_1 = a_{11}f_1 + a_{12}f_2 + e_1 \\ u_2 = a_{21}f_1 + a_{22}f_2 + e_2 \\ u_3 = a_{31}f_1 + a_{32}f_2 + e_3 \\ u_4 = a_{41}f_1 + a_{42}f_2 + e_4 \\ u_5 = a_{51}f_1 + a_{52}f_2 + e_5 \end{cases}$$

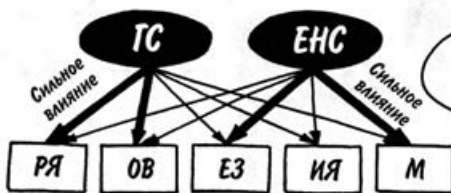
ЭТО МЕТОД АНАЛИЗА ДЛЯ ПРОВЕРКИ ЗНАЧЕНИЙ ФАКТОРНЫХ НАГРУЗОК, КОТОРЫЕ БЫЛИ В УРАВНЕНИЯХ И НА СХЕМЕ 8-ГО ВАЖНОГО МОМЕНТА.



ТАК ВОТ ОНО ЧТО?

ОДНАКО ТЕПЕРЬ Я, КАЖЕТСЯ, ПОНИМАЮ... ВЕДЬ ПРИСТУПАЯ К ФАКТОРНОМУ АНАЛИЗУ, АНАЛИТИК УЖЕ ИМЕЕТ НЕКОТОРОЕ ПРЕДСТАВЛЕНИЕ О ТОМ, КАКИЕ ОБЩИЕ ФАКТОРЫ СКРЫВАЮТСЯ ЗА ДАННЫМИ.

СЧИТАЕТСЯ, ЧТО ЧЕМ БОЛЬШЕ АБСОЛЮТНОЕ ЗНАЧЕНИЕ ФАКТОРНОЙ НАГРУЗКИ, ТЕМ БОЛЬШЕ ВЛИЯНИЕ ЭТОТ ОБЩИЙ ФАКТОР ОКАЗЫВАЕТ НА ЭТОТ ОТКЛИК.



ЯСНО.

ВОТ И ВСЁ О ВАЖНЫХ МОМЕНТАХ!

ТЕПЕРЬ ДАВАЙ ВМЕСТЕ СОСТАВИМ АНКЕТУ ДЛЯ ФАКТОРНОГО АНАЛИЗА С УЧЁТОМ ВСЕГО ИЗУЧЕННОГО!

ДАВАЙ!

ЧТО НАСЧЁТ ЭТОГО ПУНКТА?

ИСПРАВЬ-КА ВОТ ЭТО...

УРА!

3. КОНКРЕТНЫЙ ПРИМЕР ФАКТОРНОГО АНАЛИЗА



АНКЕТА ДЛЯ ПОСЕТИТЕЛЕЙ

С помощью этого анкетирования мы хотим сделать кафе NORNS лучше. Просим Вашего содействия. Каждому заполнившему анкету на кассе будет выдан бесплатный купон!

■ Пожалуйста, оцените Ваше впечатление от кафе NORNS.

Q1. Как бы Вы оценили следующее? (в каждой строке — только 1 кружок)

	Очень плохо	Плохо	Неоднозначно	Хорошо	Очень хорошо
а. Внешний вид здания	1	2	3	4	5
б. Атмосфера в кафе.	1	2	3	4	5
с. Работа официанток	1	2	3	4	5
д. Вкус чёрного чая	1	2	3	4	5
е. Цена чёрного чая	1	2	3	4	5
ф. Дизайн чашек	1	2	3	4	5

■ Пожалуйста, расскажите о себе.

Q2. Ваш пол. (только 1 кружок)

1. муж.

2. жен.

Ваш возраст.

21 лет

Ваш род занятий. (только 1 кружок)

1. Служащий

4. Домохозяйка

2. Предприниматель

ЭТО АНКЕТЫ,
ЗАПОЛНЕННЫЕ
ПОСЕТИТЕЛЯМИ!

ПОРЯДОК В КАФЕ
МЫ УЖЕ НАВЕЛИ...
ДАВАЙ
ПРИСТУПИМ
К АНАЛИЗУ!

УЖЕ МОЖНО?
БОЛЬШОЕ
СПАСИБО!

В ОПРОСЕ
УЧАСТВОВАЛИ
ПОСЕТИТЕЛИ,
ПРИШЕДШИЕ В НАШЕ
КАФЕ СЕГОДНЯ
С 15:00 ДО 16:00...

■ Пожалуйста, оцените Ваше впечатление от кафе NORNIS.

Q1. Как бы Вы оценили следующее? (в каждой строке — только 1 кружок)

	Очень плохо	Плохо	Неоднозначно	Хорошо	Очень хорошо
a. Внешний вид здания	1	2	3	4	5
b. Атмосфера в кафе	1	2	3	4	5
c. Работа официанток	1	2	3	4	5
d. Вкус чёрного чая	1	2	3	4	5
e. Цена чёрного чая	1	2	3	4	5
f. Дизайн чашек	1	2	3	4	5

ДАВАЙ ИЗ ОТВЕТОВ
НА ШЕСТЬ ВОПРОСОВ
ПУНКТА "ВПЕЧАТЛЕНИЕ
ОТ КАФЕ NORNIS"
ВЫЯСНИМ СЛЕДУЮЩЕЕ:

"МЫСЛИ", СТОЯЩИЕ ЗА ОТВЕТАМИ,
ТО ЕСТЬ ОБЩИЕ ФАКТОРЫ
"ЧТО ВАМ ПРАВИТСЯ В КАФЕ NORNIS"
И ЗНАЧЕНИЯ ФАКТОРНЫХ НАГРУЗОК.
ИТАК, ПРИСТУПАЕМ!

ДАВАЙ!

Респонденты	Q1a Внешний вид здания	Q1b Атмосфера в кафе	Q1c Работа официанток	Q1d Вкус чёрного чая	Q1e Цена чёрного чая	Q1f Дизайн чашек
А.	5	5	5	4	4	2
Б.	5	4	5	2	2	2
В.	4	4	4	4	4	4
Г.	2	3	4	3	3	3
Д.	3	3	3	3	4	1
Е.	5	4	5	3	2	3
Ж.	5	5	5	4	5	5
З.	3	1	2	5	4	4
И.	4	1	3	3	2	3
К.	1	2	2	2	2	2
Л.	3	2	3	1	1	1
М.	4	3	4	4	3	4
Н.	3	2	3	4	5	5
О.	4	3	4	5	4	5
П.	2	2	3	5	5	4

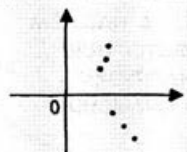
ВВОД
ДАННЫХ
ЗАКОНЧЕН!

МЫ ПРОВЕДАЕМ ДЛЯ ЭТИХ ДАННЫХ
ФАКТОРНЫЙ АНАЛИЗ, ИСХОДЯ
ИЗ ПРЕДОПОЛОЖЕНИЯ, ЧТО ЧИСЛО
ОБЩИХ ФАКТОРОВ РАВНО 2.

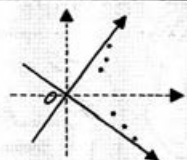


ВОТ ПОРЯДОК
ПРОВЕДЕНИЯ
ФАКТОРНОГО
АНАЛИЗА.

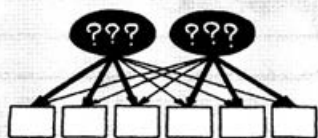
1 РАССЧИТЫВАЕМ ФАКТОРНЫЕ НАГРУЗКИ
ДО ВРАЩЕНИЯ



2 РАССЧИТЫВАЕМ ФАКТОРНЫЕ НАГРУЗКИ
ПОСЛЕ ВРАЩЕНИЯ



3 ИНТЕРПРЕТИРУЕМ КАЖДЫЙ
ИЗ ОБЩИХ ФАКТОРОВ



4 ПРОВЕРЯЕМ ТОЧНОСТЬ
РЕЗУЛЬТАТОВ АНАЛИЗА



5 НАХОДИМ ФАКТОРНЫЕ ЗНАЧЕНИЯ
И ВЫЯВЛЯЕМ ОСОБЕННОСТИ
КАЖДОГО ЭКЗЕМПЛЯРА

	1-й общий фактор	2-й общий фактор
А.	Х Х	Х Х
Б.	Х Х	Х Х
В.	Х Х	Х Х
⋮	⋮	⋮

АГА, В ПУНКТЕ **2**
ПРОВОДИТСЯ
"ВРАЩЕНИЕ", КОТОРОЕ
МЫ ДЕЛАЛИ РАНЬШЕ,
АА?



АА.
Подробности
потом.



1 РАСЧИТЫВАЕМ ФАКТОРНЫЕ НАГРУЗКИ ДО ВРАЩЕНИЯ

СНАЧАЛА НАИДЕМ ФАКТОРНЫЕ НАГРУЗКИ ДО ВРАЩЕНИЯ.

ТО ЕСТЬ УЗНАЕМ СТЕПЕНИ ВЛИЯНИЯ ОБЩИХ ФАКТОРОВ, АА?

ВОТ ЭТИ!

1-й общий фактор f_1

2-й общий фактор f_2

Внешний вид здания

Атмосфера в кафе

Работа официанток

Вкус чёрного чая

Цена чёрного чая

Дизайн чашек

Вот эта!

СУЩЕСТВУЕТ МНОГО СПОСОБОВ РАСЧЁТА ФАКТОРНЫХ НАГРУЗОК, НАПРИМЕР, МЕТОД ГЛАВНЫХ ФАКТОРОВ, МЕТОД МАКСИМАЛЬНОГО ПРАВОПОДОБИЯ И Т.Д.

СЕГОДНЯ Я ОБЪЯСНЮ МЕТОД ГЛАВНЫХ ФАКТОРОВ.

Метод главных факторов

ХОРОШО.

РАСЧЁТ ПРОИЗВОДИТСЯ В 16 ШАГОВ.

ЧТО?! В ШЕСТНАДЦАТЬ?!

ИТАК, НАЧНЁМ!

ОХ...

Шаг 1

Проводим нормировку по каждой из переменных.

(ип – нормированная переменная, СА – среднее арифметическое, СКО – среднее квадратичное отклонение)

Респонденты	Q1a Внешний вид здания	...	Q1f Дизайн чашек
А.	5	...	2
Б.	5	...	2
В.	4	...	4
Г.	2	...	3
Д.	3	...	1
Е.	5	...	3
Ж.	5	...	5
З.	3	...	4
И.	4	...	3
К.	1	...	2
Л.	3	...	1
М.	4	...	4
Н.	3	...	5
О.	4	...	5
П.	2	...	4
СА	3,5	...	3,2
СКО	1,2	...	1,4



Респонденты	Q1a (ип), u_1	...	Q1f (ип), u_6
А.	1,2	...	-0,9
Б.	1,2	...	-0,9
В.	0,4	...	0,6
Г.	-1,2	...	-0,1
Д.	-0,4	...	-1,6
Е.	1,2	...	-0,1
Ж.	1,2	...	1,3
З.	-0,4	...	0,6
И.	0,4	...	-0,1
К.	-2,0	...	-0,9
Л.	-0,4	...	-1,6
М.	0,4	...	0,6
Н.	-0,4	...	1,3
О.	0,4	...	1,3
П.	-1,2	...	0,6
СА	0	...	0
СКО	1	...	1

$$\sqrt{\frac{(5-3,5)^2 + \dots + (2-3,5)^2}{15-1}} = 1,2$$

$$\frac{2-3,5}{1,2} = -1,2$$

$$\sqrt{\frac{(-0,9-0)^2 + \dots + (0,6-0)^2}{15-1}} = 1$$

ЗНАМЕНАТЕЛЬ СРЕДНЕКВАДРАТИЧНОГО ОТКЛОНЕНИЯ, ИСПОЛЬЗУЕМОГО ПРИ НОРМИРОВКЕ В ФАКТОРНОМ АНАЛИЗЕ, ПРИНЯТО ВЫЧИСЛЯТЬ КАК "РАЗМЕР ВЫБОРКИ МИНУС 1".



Шаг 2

Предполагаем, что нормированные данные можно выразить следующим образом: (нп – нормированная переменная, СА – среднее арифметическое, СКО – среднее квадратичное отклонение)

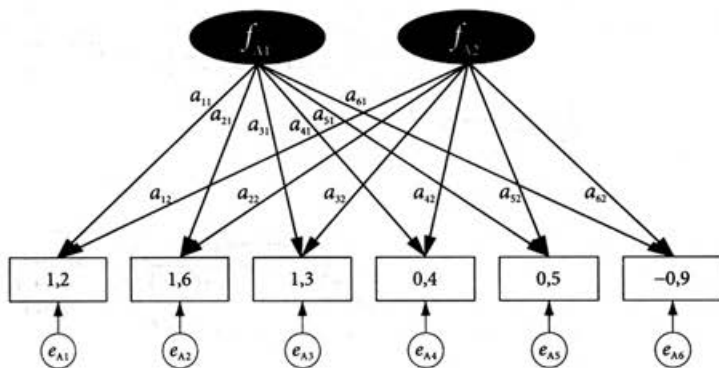
	Q1a Внешний вид здания u_1	...	Q1f Дизайн чашек u_6	=		Q1a (нп), u_1	...	Q1f (нп), u_6
А.	1,2	...	-0,9		А.	$a_{11}f_{A1} + a_{12}f_{A2} + e_{A1}$...	$a_{61}f_{A1} + a_{62}f_{A2} + e_{A6}$
:	:	:	:		:	:	:	:
П.	-1,2	...	0,6		П.	$a_{11}f_{П1} + a_{12}f_{П2} + e_{П1}$...	$a_{61}f_{П1} + a_{62}f_{П2} + e_{П6}$
СА	0	...	0		СА	0	...	0
СКО	1	...	1		СКО	1	...	1

Предполагаем, что в каждой из рамок среднее арифметическое равно 0, а дисперсия равна 1

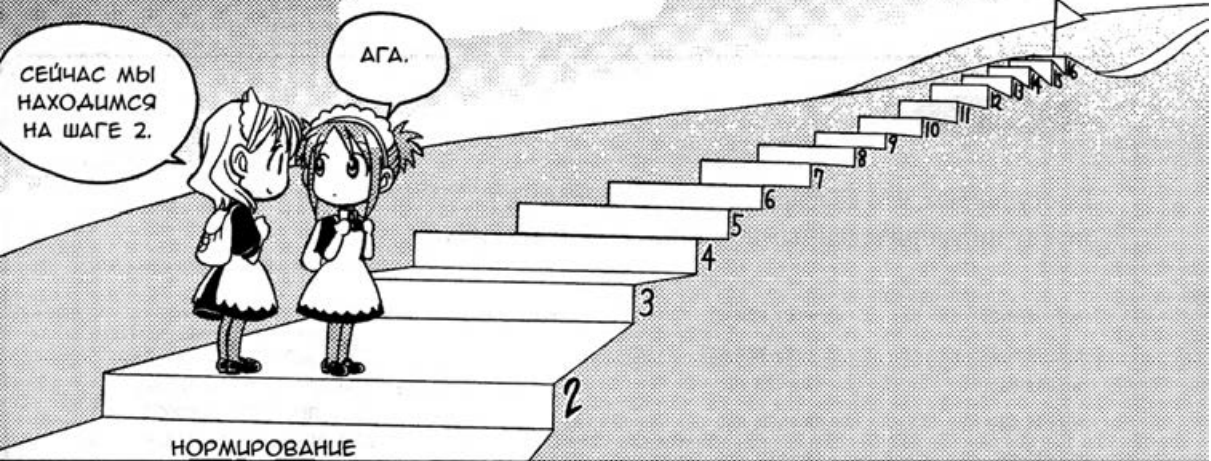
Предполагаем, что среднее арифметическое равно 0, а дисперсия равна d_1^2

Предполагаем, что среднее арифметическое равно 0, а дисперсия равна d_6^2

ДАННЫЕ, НАПРИМЕР, РЕСПОНДЕНТА А., МОЖНО ПРЕДСТАВИТЬ В ВИДЕ СЛЕДУЮЩЕЙ СХЕМЫ:



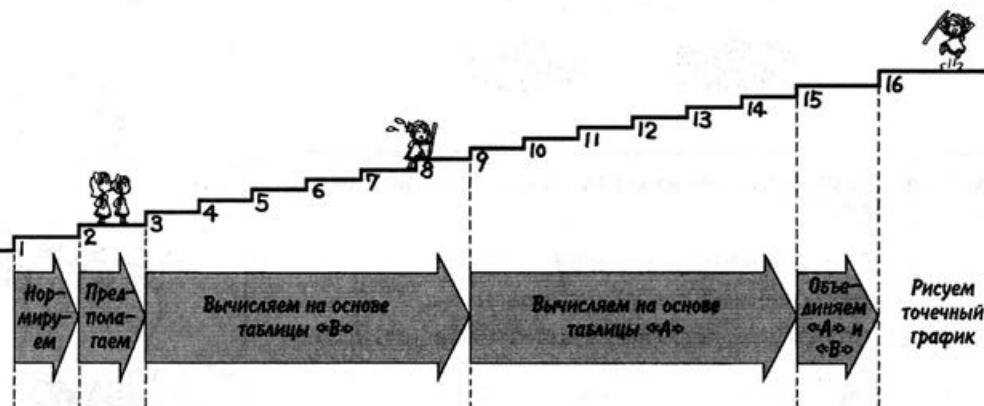
ЭТО ЖЕ
ФОРМУЛА И СХЕМА ИЗ
"ВАЖНОГО МОМЕНТА
№8", КОТОРЫЙ МЫ
НЕДАВНО ИЗУЧИЛИ!



Респонденты	Q1a (нп), u_1	...	Q1f (нп), u_6
А.	1,2	...	-0,9
:	:	:	:
П.	-1,2	...	0,6
СА	0	...	0
СКО	1	...	1

=

Респонденты	Q1a (нп), u_1	...	Q1f (нп) u_6
А.	$a_{11}f_1^A + a_{12}f_2^A + e_1^A$...	$a_{61}f_1^A + a_{62}f_2^A + e_6^A$
:	:	:	:
П.	$a_{11}f_1^П + a_{12}f_2^П + e_1^П$...	$a_{61}f_1^П + a_{62}f_2^П + e_6^П$
СА	0	...	0
СКО	1	...	1



Шаг 3

Постулируем, что коэффициенты корреляции между нижеприведёнными величинами равны нулю:

- f_1 и e_1 ;
- f_1 и e_2 ;
- f_1 и e_3 ;
- f_1 и e_4 ;
- f_1 и e_5 ;
- f_1 и e_6 ;
- f_2 и e_1 ;
- f_2 и e_2 ;
- f_2 и e_3 ;
- f_2 и e_4 ;
- f_2 и e_5 ;
- f_2 и e_6 ;
- e_1 и e_2 ;
- e_1 и e_3 ;
- e_1 и e_4 ;
- e_1 и e_5 ;
- e_1 и e_6 ;
- e_2 и e_3 ;
- e_2 и e_4 ;
- e_2 и e_5 ;
- e_2 и e_6 ;
- e_3 и e_4 ;
- e_3 и e_5 ;
- e_3 и e_6 ;
- e_4 и e_5 ;
- e_4 и e_6 ;
- e_5 и e_6 .

ДРУГИМИ СЛОВАМИ, ПОЛАГАЕМ, ЧТО СПЕЦИФИЧЕСКОЕ ФАКТОРЫ НЕ СВЯЗАНЫ НИ ДРУГ С ДРУГОМ, НИ С ОБЩИМИ ФАКТОРАМИ, ДА?



НАПРИМЕР, РАВЕНСТВО НУЛЮ КОЭФФИЦИЕНТА КОРРЕЛЯЦИИ f_2 и e_6 , ТО ЕСТЬ

$$\frac{\sum (\text{отклонения } f_2 \times \text{отклонения } e_6)}{\sqrt{\sum (\text{отклонения } f_2)^2} \times \sqrt{\sum (\text{отклонения } e_6)^2}} = 0$$

ОЗНАЧАЕТ, ЧТО СУММА ПРОИЗВЕДЕНИЙ ОТКЛОНЕНИЙ f_2 и e_6

$$= (f_2^A - \bar{f}_2)(e_6^A - \bar{e}_6) + \dots + (f_2^n - \bar{f}_2)(e_6^n - \bar{e}_6) =$$

$$= (f_2^A - 0)(e_6^A - 0) + \dots + (f_2^n - 0)(e_6^n - 0) =$$

$$= f_2^A e_6^A + \dots + f_2^n e_6^n = 0.$$

Исходя из предположений Шага 2



Шаг 4

Предполагаем, что коэффициент корреляции между f_1 и f_2 равен нулю, то есть:

$$\frac{\sum (\text{отклонения } f_1 \times \text{отклонения } f_2)}{\sqrt{\sum (\text{отклонения } f_1)^2} \times \sqrt{\sum (\text{отклонения } f_2)^2}} = 0.$$

В этом случае для суммы произведений отклонений f_1 и f_2 получаем следующее:

$$\begin{aligned} & (f_1^A - \bar{f}_1)(f_2^A - \bar{f}_2) + \dots + (f_1^n - \bar{f}_1)(f_2^n - \bar{f}_2) = \\ & = (f_1^A - 0)(f_2^A - 0) + \dots + (f_1^n - 0)(f_2^n - 0) = \\ & = f_1^A f_2^A + \dots + f_1^n f_2^n = 0. \end{aligned}$$

Исходя из предположений Шага 2

ТО ЕСТЬ МЫ ПРЕДПОЛАГАЕМ, ОБЩИЕ ФАКТОРЫ ТОЖЕ НЕ СВЯЗАНЫ ДРУГ С ДРУГОМ, ДА?

МОДЕЛЬ, В КОТОРОЙ КОЭФФИЦИЕНТЫ КОРРЕЛЯЦИИ МЕЖДУ ЛЮБЫМИ РАЗЛИЧНЫМИ ОБЩИМИ ФАКТОРАМИ ПРЕДПОЛАГАЮТСЯ РАВНЫМИ НУЛЮ, НАЗЫВАЕТСЯ **МОДЕЛЬЮ ОРТОГОНАЛЬНЫХ ФАКТОРОВ**. МОДЕЛЬ, НЕ ПРЕДПОЛАГАЮЩАЯ ЭТОГО, НАЗЫВАЕТСЯ **МОДЕЛЬЮ НЕОРТОГОНАЛЬНЫХ ФАКТОРОВ**.

РАНЬШЕ В ОСНОВНОМ ИСПОЛЬЗОВАЛИ МОДЕЛЬ ОРТОГОНАЛЬНЫХ ФАКТОРОВ, ТАК КАК "ВЫЧИСЛЕНИЯ В НЕЙ СРАВНИТЕЛЬНО ПРОЩЕ". ОДНАКО В НАСТОЯЩЕЕ ВРЕМЯ МОДЕЛЬ НЕОРТОГОНАЛЬНЫХ ФАКТОРОВ ТОЖЕ, ПОХОЖЕ, НАЧИНАЕТ ПРИМЕНЯТЬСЯ ПО СЛЕДУЮЩИМ ПРИЧИНАМ:

- УВЕЛИЧЕНИЕ ПРОИЗВОДИТЕЛЬНОСТИ КОМПЬЮТЕРОВ;
- УЩЕРБНОСТЬ ПОСТУЛАТА МОДЕЛИ ОРТОГОНАЛЬНЫХ ФАКТОРОВ "О РАВЕНСТВЕ НУЛЮ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ МЕЖДУ ЛЮБЫМИ РАЗЛИЧНЫМИ ОБЩИМИ ФАКТОРАМИ"

В ДАННОМ ПРИМЕРЕ МЫ ИСПОЛЬЗУЕМ МОДЕЛЬ ОРТОГОНАЛЬНЫХ ФАКТОРОВ.

Шаг 5

Убеждаемся, что выражение коэффициента корреляции, например, между переменной u_2 , которая получена нормировкой Q1b, и переменной u_6 , которая получена нормировкой Q1f, можно преобразовать следующим образом:

$$\frac{\sum (\text{отклонения } u_2 \times \text{отклонения } u_6)}{\sqrt{\sum (\text{отклонения } u_2)^2} \times \sqrt{\sum (\text{отклонения } u_6)^2}}$$

$$= \frac{\sum (\text{отклонения } u_2 \times \text{отклонения } u_6)}{15 - 1} \div \frac{\sqrt{\sum (\text{отклонения } u_2)^2} \times \sqrt{\sum (\text{отклонения } u_6)^2}}{15 - 1}$$

Делим числитель и знаменатель на размер выборки минус 1

$$= \frac{\sum (\text{отклонения } u_2 \times \text{отклонения } u_6)}{15 - 1} \div \left(\sqrt{\frac{\sum (\text{отклонения } u_2)^2}{15 - 1}} \times \sqrt{\frac{\sum (\text{отклонения } u_6)^2}{15 - 1}} \right)$$

Представляем знаменатель в виде произведения среднеквадратичных отклонений u_2 и u_6

$$= \frac{\sum (\text{отклонения } u_2 \times \text{отклонения } u_6)}{15 - 1} \div 1 \times 1$$

Согласно шагу 1

$$= \frac{\sum (\text{отклонения } u_2 \times \text{отклонения } u_6)}{15 - 1}$$

$$= \frac{(a_{21}f_1^{\wedge} + a_{22}f_2^{\wedge} + e_2^{\wedge})(a_{61}f_1^{\wedge} + a_{62}f_2^{\wedge} + e_6^{\wedge}) + \dots + (a_{21}f_1^{\wedge} + a_{22}f_2^{\wedge} + e_2^{\wedge})(a_{61}f_1^{\wedge} + a_{62}f_2^{\wedge} + e_6^{\wedge})}{15 - 1}$$

Упорядочиваем числитель:

$$\begin{aligned} & (a_{21}f_1^{\wedge} + a_{22}f_2^{\wedge} + e_2^{\wedge})(a_{61}f_1^{\wedge} + a_{62}f_2^{\wedge} + e_6^{\wedge}) + \dots + (a_{21}f_1^{\wedge} + a_{22}f_2^{\wedge} + e_2^{\wedge})(a_{61}f_1^{\wedge} + a_{62}f_2^{\wedge} + e_6^{\wedge}) = \\ & = \boxed{a_{21}f_1^{\wedge}a_{61}f_1^{\wedge}} + \boxed{a_{21}f_1^{\wedge}a_{62}f_2^{\wedge}} + \boxed{a_{21}f_1^{\wedge}e_6^{\wedge}} + \boxed{a_{22}f_2^{\wedge}a_{61}f_1^{\wedge}} + \boxed{a_{22}f_2^{\wedge}a_{62}f_2^{\wedge}} + \boxed{a_{22}f_2^{\wedge}e_6^{\wedge}} + \boxed{e_2^{\wedge}a_{61}f_1^{\wedge}} + \boxed{e_2^{\wedge}a_{62}f_2^{\wedge}} + \boxed{e_2^{\wedge}e_6^{\wedge}} + \\ & + \boxed{a_{21}f_1^{\wedge}a_{61}f_1^{\wedge}} + \boxed{a_{21}f_1^{\wedge}a_{62}f_2^{\wedge}} + \boxed{a_{21}f_1^{\wedge}e_6^{\wedge}} + \boxed{a_{22}f_2^{\wedge}a_{61}f_1^{\wedge}} + \boxed{a_{22}f_2^{\wedge}a_{62}f_2^{\wedge}} + \boxed{a_{22}f_2^{\wedge}e_6^{\wedge}} + \boxed{e_2^{\wedge}a_{61}f_1^{\wedge}} + \boxed{e_2^{\wedge}a_{62}f_2^{\wedge}} + \boxed{e_2^{\wedge}e_6^{\wedge}} \\ & = a_{21}a_{61}(f_1^{\wedge 2} + \dots + f_1^{\wedge 2}) + a_{21}a_{62}(f_1^{\wedge}f_2^{\wedge} + \dots + f_1^{\wedge}f_2^{\wedge}) + a_{21}(f_1^{\wedge}e_6^{\wedge} + \dots + f_1^{\wedge}e_6^{\wedge}) + \\ & + a_{22}a_{61}(f_2^{\wedge}f_1^{\wedge} + \dots + f_2^{\wedge}f_1^{\wedge}) + a_{22}a_{62}(f_2^{\wedge 2} + \dots + f_2^{\wedge 2}) + a_{22}(f_2^{\wedge}e_6^{\wedge} + \dots + f_2^{\wedge}e_6^{\wedge}) + \\ & + a_{61}(f_1^{\wedge}e_2^{\wedge} + \dots + f_1^{\wedge}e_2^{\wedge}) + a_{62}(f_2^{\wedge}e_2^{\wedge} + \dots + f_2^{\wedge}e_2^{\wedge}) + (e_2^{\wedge}e_6^{\wedge} + \dots + e_2^{\wedge}e_6^{\wedge}) = \end{aligned}$$

$$= a_{21}a_{61} \Sigma[(\text{отклон. } f_1)^2] + a_{21}a_{62} \Sigma[\text{отклон. } f_1 \times \text{отклон. } f_2] + a_{21} \Sigma[\text{отклон. } f_1 \times \text{отклон. } e_6] +$$

$$+ a_{22}a_{61} \Sigma[\text{отклон. } f_1 \times \text{отклон. } f_2] + a_{22}a_{62} \Sigma[(\text{отклон. } f_2)^2] + a_{22} \Sigma[\text{отклон. } f_2 \times \text{отклон. } e_6] +$$

$$+ a_{61} \Sigma[\text{отклон. } f_1 \times \text{отклон. } e_2] + a_{62} \Sigma[\text{отклон. } f_2 \times \text{отклон. } e_2] + \Sigma[\text{отклон. } e_2 \times \text{отклон. } e_6]$$

$$= a_{21}a_{61} \Sigma[(\text{отклон. } f_1)^2] + 0 + 0$$

$$+ 0 + a_{22}a_{62} \Sigma[(\text{отклон. } f_2)^2] + 0$$

$$+ 0 + 0 + 0$$

Согласно
шагам 3 и 4

Подставляя в числитель упрощённое выражение, получаем:

$$= \frac{a_{21}a_{61} \Sigma[(\text{отклонения } f_1)^2] + a_{22}a_{62} \Sigma[(\text{отклонения } f_2)^2]}{15 - 1} =$$

$$= a_{21}a_{61} \times \frac{\Sigma[(\text{отклонения } f_1)^2]}{15 - 1} + a_{22}a_{62} \times \frac{\Sigma[(\text{отклонения } f_2)^2]}{15 - 1} =$$

$$= a_{21}a_{61} \times \text{Дисперсия } f_1 + a_{22}a_{62} \times \text{Дисперсия } f_2 =$$

Согласно шагу 2

$$= a_{21}a_{61} + a_{22}a_{62}$$

ТО ЕСТЬ МЫ ЗАДЕСЬ
ИССЛЕДУЕМ
ВЗАИМОСВЯЗИ
МЕЖДУ ОТКЛОНЕНИЯМИ,
ДА?



	Внешний вид здания (пт)	Атмосфера в кафе (пт)	Работа официанток (пт)	Вкус чёрного чая (пт)	Цена чёрного чая (пт)	Дизайн чашек (пт)
Внешний вид здания (пт)						
Атмосфера в кафе (пт)						
Работа официанток (пт)						
Вкус чёрного чая (пт)						
Цена чёрного чая (пт)						
Дизайн чашек (пт)						

Шаг 6

Убеждаемся, что выражение коэффициента корреляции, например, переменных u_1 и u_2 можно преобразовать следующим образом:

$$\begin{aligned}
 & \frac{\Sigma(\text{отклонения } u_2 \times \text{отклонения } u_1)}{\sqrt{\Sigma[(\text{отклонения } u_2)^2] \times \Sigma[(\text{отклонения } u_1)^2]}} \\
 &= \frac{\Sigma(\text{отклонения } u_2 \times \text{отклонения } u_1)}{\frac{15-1}{\sqrt{\Sigma[(\text{отклонения } u_2)^2] \times \Sigma[(\text{отклонения } u_1)^2]}}} \\
 &= \frac{\Sigma(\text{отклонения } u_2 \times \text{отклонения } u_1)}{15-1} \times \frac{1}{\sqrt{\frac{\Sigma[(\text{отклонения } u_2)^2]}{15-1} \times \frac{\Sigma[(\text{отклонения } u_1)^2]}{15-1}}} \\
 &= \frac{\Sigma(\text{отклонения } u_2 \times \text{отклонения } u_1)}{15-1} \times \frac{1}{1 \times 1} \\
 &= \frac{\Sigma[(\text{отклонения } u_2 \times \text{отклонения } u_1)]}{15-1} \\
 &= \frac{(a_{21}f_1^{\hat{A}} + a_{22}f_2^{\hat{A}} + e_2^{\hat{A}})^2 + \dots + (a_{21}f_1^{\hat{N}} + a_{22}f_2^{\hat{N}} + e_2^{\hat{N}})^2}{15-1}
 \end{aligned}$$

Делим числитель и знаменатель на размер выборки минус 1

Представляем знаменатель в виде произведения среднеквадратичных отклонений u_2 и u_1

Согласно шагу 1

Упорядочиваем числитель:

$$\begin{aligned}
 & (a_{21}f_1^{\hat{A}} + a_{22}f_2^{\hat{A}} + e_2^{\hat{A}})^2 + \dots + (a_{21}f_1^{\hat{N}} + a_{22}f_2^{\hat{N}} + e_2^{\hat{N}})^2 = \\
 &= \begin{matrix} \boxed{(a_{21}f_1^{\hat{A}})^2} + \boxed{(a_{22}f_2^{\hat{A}})^2} + \boxed{(e_2^{\hat{A}})^2} + \boxed{2(a_{21}f_1^{\hat{A}})(a_{22}f_2^{\hat{A}})} + \boxed{2(a_{22}f_2^{\hat{A}})(e_2^{\hat{A}})} + \boxed{2(a_{21}f_1^{\hat{A}})(e_2^{\hat{A}})} \\ + \dots + \\ + \boxed{(a_{21}f_1^{\hat{N}})^2} + \boxed{(a_{22}f_2^{\hat{N}})^2} + \boxed{(e_2^{\hat{N}})^2} + \boxed{2(a_{21}f_1^{\hat{N}})(a_{22}f_2^{\hat{N}})} + \boxed{2(a_{22}f_2^{\hat{N}})(e_2^{\hat{N}})} + \boxed{2(a_{21}f_1^{\hat{N}})(e_2^{\hat{N}})} \end{matrix} = \\
 &= a_{21}^2(f_1^{\hat{A}2} + \dots + f_1^{\hat{N}2}) + a_{22}^2(f_2^{\hat{A}2} + \dots + f_2^{\hat{N}2}) + (e_2^{\hat{A}2} + \dots + e_2^{\hat{N}2}) + \\
 &+ 2a_{21}a_{22}(f_1^{\hat{A}}f_2^{\hat{A}} + \dots + f_1^{\hat{N}}f_2^{\hat{N}}) + 2a_{22}(f_2^{\hat{A}}e_2^{\hat{A}} + \dots + f_2^{\hat{N}}e_2^{\hat{N}}) + 2a_{21}(f_1^{\hat{A}}e_2^{\hat{A}} + \dots + f_1^{\hat{N}}e_2^{\hat{N}}) = \\
 &= a_{21}^2\Sigma[(\text{отклонения } f_1)^2] + a_{22}^2\Sigma[(\text{отклонения } f_2)^2] + \Sigma[(\text{отклонения } e_2)^2] + \\
 &+ 2a_{21}a_{22}\Sigma[(\text{отклонения } f_1) \times (\text{отклонения } f_2)] + 2a_{22}\Sigma[(\text{отклонения } f_2) \times (\text{отклонения } e_2)] + \\
 &+ 2a_{21}\Sigma[(\text{отклонения } f_1) \times (\text{отклонения } e_2)] =
 \end{aligned}$$

$$= a_{21}^2 \Sigma[(\text{отклонения } f_1)^2] + a_{22}^2 \Sigma[(\text{отклонения } f_2)^2] + \Sigma[(\text{отклонения } e_2)^2] +$$

$$+ \quad 0 \quad + \quad 0 \quad + \quad 0$$

Согласно
шагам 3 и 4

Подставляя в числитель упрощённое выражение, получаем:

$$= \frac{a_{21}^2 \Sigma[(\text{отклонения } f_1)^2] + a_{22}^2 \Sigma[(\text{отклонения } f_2)^2] + \Sigma[(\text{отклонения } e_2)^2]}{15 - 1} =$$

$$= a_{21}^2 \frac{\Sigma[(\text{отклонения } f_1)^2]}{15 - 1} + a_{22}^2 \frac{\Sigma[(\text{отклонения } f_2)^2]}{15 - 1} + \frac{\Sigma[(\text{отклонения } e_2)^2]}{15 - 1} =$$

$$= a_{21}^2 (\text{дисперсия } f_1) + a_{22}^2 (\text{дисперсия } f_2) + (\text{дисперсия } e_2) =$$

$$= a_{21}^2 + a_{22}^2 + d_2^2 \quad \text{Согласно шагу 2}$$

ТО ЕСТЬ МЫ
ЗДЕСЬ ИССЛЕДУЕМ
ВЗАИМОСВЯЗИ
ОТКЛОНЕНИЙ САМИХ
С СОБОЙ, ДА?



	Внешний вид здания (нп)	Атмосфера в кафе (нп)	Работа официанток (нп)	Вкус чёрного чая (нп)	Цена чёрного чая (нп)	Дизайн чашек (нп)
Внешний вид здания (нп)						
Атмосфера в кафе (нп)						
Работа официанток (нп)						
Вкус чёрного чая (нп)						
Цена чёрного чая (нп)						
Дизайн чашек (нп)						

Шаг 7

Убедимся, что матрицу корреляции можно, на основе Шагов 5 и 6, переписать следующим образом:

$$\begin{pmatrix} r_{11} & r_{12} & \dots & r_{16} \\ r_{21} & r_{22} & \dots & r_{26} \\ \vdots & \vdots & \ddots & \vdots \\ r_{61} & r_{62} & \dots & r_{66} \end{pmatrix} = \begin{pmatrix} a_{11}^2 + a_{12}^2 + d_1^2 & a_{11}a_{21} + a_{12}a_{22} & \dots & a_{11}a_{61} + a_{12}a_{62} \\ a_{21}a_{11} + a_{22}a_{12} & a_{21}^2 + a_{22}^2 + d_2^2 & \dots & a_{21}a_{61} + a_{22}a_{62} \\ \vdots & \vdots & \ddots & \vdots \\ a_{61}a_{11} + a_{62}a_{12} & a_{61}a_{21} + a_{62}a_{22} & \dots & a_{61}^2 + a_{62}^2 + d_6^2 \end{pmatrix} =$$

Понятно.



$$= \begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} & \dots & a_{11}a_{61} + a_{12}a_{62} \\ a_{21}a_{11} + a_{22}a_{12} & a_{21}^2 + a_{22}^2 & \dots & a_{21}a_{61} + a_{22}a_{62} \\ \vdots & \vdots & \ddots & \vdots \\ a_{61}a_{11} + a_{62}a_{12} & a_{61}a_{21} + a_{62}a_{22} & \dots & a_{61}^2 + a_{62}^2 \end{pmatrix} + \begin{pmatrix} d_1^2 & 0 & \dots & 0 \\ 0 & d_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_6^2 \end{pmatrix}$$

Шаг 8

Упорядочим выражение Шага 7. Элементы $1 - d_i^2$ главной диагонали матрицы в левой части уравнения, появляющиеся в процессе вычислений, называются *общностями* и иногда обозначаются как h_i^2 .

$$\begin{pmatrix} r_{11} & r_{12} & \dots & r_{16} \\ r_{21} & r_{22} & \dots & r_{26} \\ \vdots & \vdots & \ddots & \vdots \\ r_{61} & r_{62} & \dots & r_{66} \end{pmatrix} - \begin{pmatrix} d_1^2 & 0 & \dots & 0 \\ 0 & d_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_6^2 \end{pmatrix} = \begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} & \dots & a_{11}a_{61} + a_{12}a_{62} \\ a_{21}a_{11} + a_{22}a_{12} & a_{21}^2 + a_{22}^2 & \dots & a_{21}a_{61} + a_{22}a_{62} \\ \vdots & \vdots & \ddots & \vdots \\ a_{61}a_{11} + a_{62}a_{12} & a_{61}a_{21} + a_{62}a_{22} & \dots & a_{61}^2 + a_{62}^2 \end{pmatrix}$$

Переносим последнее слагаемое (см. Шаг 7) справа налево

$$\begin{pmatrix} 1 & r_{12} & \dots & r_{16} \\ r_{21} & 1 & \dots & r_{26} \\ \vdots & \vdots & \ddots & \vdots \\ r_{61} & r_{62} & \dots & 1 \end{pmatrix} - \begin{pmatrix} d_1^2 & 0 & \dots & 0 \\ 0 & d_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & d_6^2 \end{pmatrix} = \begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} & \dots & a_{11}a_{61} + a_{12}a_{62} \\ a_{21}a_{11} + a_{22}a_{12} & a_{21}^2 + a_{22}^2 & \dots & a_{21}a_{61} + a_{22}a_{62} \\ \vdots & \vdots & \ddots & \vdots \\ a_{61}a_{11} + a_{62}a_{12} & a_{61}a_{21} + a_{62}a_{22} & \dots & a_{61}^2 + a_{62}^2 \end{pmatrix}$$

Заменяем r_{ii} на единицу.

$$\begin{pmatrix} 1 - d_1^2 & r_{12} & \dots & r_{16} \\ r_{21} & 1 - d_2^2 & \dots & r_{26} \\ \vdots & \vdots & \ddots & \vdots \\ r_{61} & r_{62} & \dots & 1 - d_6^2 \end{pmatrix} = \begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} & \dots & a_{11}a_{61} + a_{12}a_{62} \\ a_{21}a_{11} + a_{22}a_{12} & a_{21}^2 + a_{22}^2 & \dots & a_{21}a_{61} + a_{22}a_{62} \\ \vdots & \vdots & \ddots & \vdots \\ a_{61}a_{11} + a_{62}a_{12} & a_{61}a_{21} + a_{62}a_{22} & \dots & a_{61}^2 + a_{62}^2 \end{pmatrix}$$

Упорядочиваем правую часть.

$$\begin{pmatrix} a_{11}^2 + a_{12}^2 & a_{11}a_{21} + a_{12}a_{22} & \dots & a_{11}a_{61} + a_{12}a_{62} \\ a_{21}a_{11} + a_{22}a_{12} & a_{21}^2 + a_{22}^2 & \dots & a_{21}a_{61} + a_{22}a_{62} \\ \vdots & \vdots & \ddots & \vdots \\ a_{61}a_{11} + a_{62}a_{12} & a_{61}a_{21} + a_{62}a_{22} & \dots & a_{61}^2 + a_{62}^2 \end{pmatrix} =$$

$$\begin{pmatrix} a_{11}a_{11} + a_{12}a_{12} & a_{11}a_{21} + a_{12}a_{22} & \dots & a_{11}a_{61} + a_{12}a_{62} \\ a_{21}a_{11} + a_{22}a_{12} & a_{21}a_{21} + a_{22}a_{22} & \dots & a_{21}a_{61} + a_{22}a_{62} \\ \vdots & \vdots & \ddots & \vdots \\ a_{61}a_{11} + a_{62}a_{12} & a_{61}a_{21} + a_{62}a_{22} & \dots & a_{61}a_{61} + a_{62}a_{62} \end{pmatrix} =$$

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \dots & a_{61} \\ a_{12} & a_{22} & \dots & a_{62} \end{pmatrix}$$

Подставляем

$$\begin{pmatrix} 1 - d_1^2 & r_{12} & \dots & r_{16} \\ r_{21} & 1 - d_2^2 & \dots & r_{26} \\ \vdots & \vdots & \ddots & \vdots \\ r_{61} & r_{62} & \dots & 1 - d_6^2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \dots & a_{61} \\ a_{12} & a_{22} & \dots & a_{62} \end{pmatrix}$$



8

Респонденты	Q1a (нп), u_1		Q1f (нп), u_6
A.		A	
:			
П.			
СА	0	...	0
СКО	1	...	1

=

Респонденты	Q1a (нп), u_1		Q1f (нп), u_6
A.	$a_{11}f_1 + a_{12}f_2 + \dots + a_{61}f_6$	СДЕЛАНО	$a_{11}f_1 + a_{12}f_2 + \dots + a_{61}f_6$
:			
П.	$a_{11}f_1 + a_{12}f_2 + \dots + a_{61}f_6$		$a_{11}f_1 + a_{12}f_2 + \dots + a_{61}f_6$
СА	0	...	0
СКО	1	...	1

ИТАК, МЫ ВЫПОЛНИЛИ ВЫЧИСЛЕНИЯ ПО ТАБЛИЦЕ "В". ПЕРЕХОДИМ К ВЫЧИСЛЕНИЯМ ПО ТАБЛИЦЕ "А".

АА.

Вычисляем на основе таблицы "А"

Объединяем вычисления "А" и "В".

Рисуем точечный график

Шаг 9

Находим матрицу корреляции для анализируемых данных и вычитаем из неё ма-

трицу $\begin{pmatrix} d_1^2 & 0 & \dots & 0 \\ 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & d_6^2 \end{pmatrix}$:

$$\begin{pmatrix} 1 & 0,65 & 0,80 & 0,11 & 0,01 & 0,14 \\ 0,65 & 1 & 0,89 & 0,02 & 0,19 & 0,01 \\ 0,80 & 0,89 & 1 & 0,02 & 0,04 & 0,10 \\ 0,11 & 0,02 & 0,02 & 1 & 0,82 & 0,77 \\ 0,01 & 0,19 & 0,04 & 0,82 & 1 & 0,64 \\ 0,14 & 0,01 & 0,10 & 0,77 & 0,64 & 1 \end{pmatrix} - \begin{pmatrix} d_1^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & d_2^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & d_3^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & d_4^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & d_5^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & d_6^2 \end{pmatrix} =$$

$$= \begin{pmatrix} 1-d_1^2 & 0,65 & 0,80 & 0,11 & 0,01 & 0,14 \\ 0,65 & 1-d_2^2 & 0,89 & 0,02 & 0,19 & 0,01 \\ 0,80 & 0,89 & 1-d_3^2 & 0,02 & 0,04 & 0,10 \\ 0,11 & 0,02 & 0,02 & 1-d_4^2 & 0,82 & 0,77 \\ 0,01 & 0,19 & 0,04 & 0,82 & 1-d_5^2 & 0,64 \\ 0,14 & 0,01 & 0,10 & 0,77 & 0,64 & 1-d_6^2 \end{pmatrix}$$

ЭТОТ ШАГ
ПРОСТОЙ!



Шаг 10

Принимаем значения элементов главной диагонали матрицы Шага 9, то есть значения общностей $1-d_i^2$, равными коэффициентам детерминации (КД) уравнений множественной регрессии (УМР) для соответствующих откликов и объясняющих переменных (ОП):

$$1-d_1^2 = R_1^2 \text{ (КД УМР с откликом } u_1 \text{ и ОП } u_2, u_3, u_4, u_5, u_6),$$

$$1-d_2^2 = R_2^2 \text{ (КД УМР с откликом } u_2 \text{ и ОП } u_1, u_3, u_4, u_5, u_6),$$

$$1-d_3^2 = R_3^2 \text{ (КД УМР с откликом } u_3 \text{ и ОП } u_1, u_2, u_4, u_5, u_6),$$

$$1-d_4^2 = R_4^2 \text{ (КД УМР с откликом } u_4 \text{ и ОП } u_1, u_2, u_3, u_5, u_6),$$

$$1-d_5^2 = R_5^2 \text{ (КД УМР с откликом } u_5 \text{ и ОП } u_1, u_2, u_3, u_4, u_6),$$

$$1-d_6^2 = R_6^2 \text{ (КД УМР с откликом } u_6 \text{ и ОП } u_1, u_2, u_3, u_4, u_5),$$

и получаем следующую матрицу:

$$\begin{pmatrix} 1-d_1^2 & 0,65 & 0,80 & 0,11 & 0,01 & 0,14 \\ 0,65 & 1-d_2^2 & 0,89 & 0,02 & 0,19 & 0,01 \\ 0,80 & 0,89 & 1-d_3^2 & 0,02 & 0,04 & 0,10 \\ 0,11 & 0,02 & 0,02 & 1-d_4^2 & 0,82 & 0,77 \\ 0,01 & 0,19 & 0,04 & 0,82 & 1-d_5^2 & 0,64 \\ 0,14 & 0,01 & 0,10 & 0,77 & 0,64 & 1-d_6^2 \end{pmatrix} = \begin{pmatrix} 0,68 & 0,65 & 0,80 & 0,11 & 0,01 & 0,14 \\ 0,65 & 0,88 & 0,89 & 0,02 & 0,19 & 0,01 \\ 0,80 & 0,89 & 0,91 & 0,02 & 0,04 & 0,10 \\ 0,11 & 0,02 & 0,02 & 0,81 & 0,82 & 0,77 \\ 0,01 & 0,19 & 0,04 & 0,82 & 0,81 & 0,64 \\ 0,14 & 0,01 & 0,10 & 0,77 & 0,64 & 0,66 \end{pmatrix}$$

В соответствии с законами математики, для последующих вычислений необходимо задать значения $1-d_i^2$. Существуют различные способы задания значений $1-d_i^2$, но здесь я расскажу о самом известном из них.



А ЧТО ТАКОЕ УРАВНЕНИЕ МНОЖЕСТВЕННОЙ РЕГРЕССИИ И КОЭФФИЦИЕНТ ДЕТЕРМИНАЦИИ?



ВОТ УРАВНЕНИЕ МНОЖЕСТВЕННОЙ РЕГРЕССИИ:

$$\begin{matrix} \text{Отклик} & \text{Объясняющие переменные} \\ \underline{y} = \alpha_1 \underline{x}_1 + \alpha_2 \underline{x}_2 + \dots + \alpha_p \underline{x}_p + \beta \end{matrix}$$

А ЕГО ТОЧНОСТЬ ВЫРАЖАЕТСЯ КОЭФФИЦИЕНТОМ ДЕТЕРМИНАЦИИ. ПОДРОБНОСТИ В МОЕЙ СТАРОЙ ТЕТРАДКЕ*.

* См. «Манга о статистике. Регрессионный анализ», глава 3

Шаг 11

Находим собственные значения λ и собственные векторы $\{t_1; \dots; t_6\}$, удовлетворяющие следующему уравнению:

$$\begin{pmatrix} 0,68 & 0,65 & 0,80 & 0,11 & 0,01 & 0,14 \\ 0,65 & 0,88 & 0,89 & 0,02 & 0,19 & 0,01 \\ 0,80 & 0,89 & 0,91 & 0,02 & 0,04 & 0,10 \\ 0,11 & 0,02 & 0,02 & 0,81 & 0,82 & 0,77 \\ 0,01 & 0,19 & 0,04 & 0,82 & 0,81 & 0,64 \\ 0,14 & 0,01 & 0,10 & 0,77 & 0,64 & 0,66 \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{pmatrix} = \lambda \begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{pmatrix}.$$

Собственные векторы находим так, чтобы они удовлетворяли соотношению $t_1^2 + t_2^2 + t_3^2 + t_4^2 + t_5^2 + t_6^2 = 1$.

С помощью программы для анализа данных должны быть получены следующие результаты:

Собственные значения λ	Собственные векторы $\begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{pmatrix}$
2,55	$\begin{pmatrix} 0,43 \\ 0,48 \\ 0,50 \\ 0,34 \\ 0,34 \\ 0,32 \end{pmatrix}$
2,11	$\begin{pmatrix} -0,28 \\ -0,34 \\ -0,38 \\ 0,51 \\ 0,47 \\ 0,43 \end{pmatrix}$

ДЛЯ ЭТОГО ПРИМЕРА НА САМОМ ДЕЛЕ МОЖНО НАЙТИ 6 ПАР СОБСТВЕННЫХ ЗНАЧЕНИЙ И СОБСТВЕННЫХ ВЕКТОРОВ. ОДНАКО Я НЕ БУДУ ОСОБО ОСТАНАВЛИВАТЬСЯ НА СОБСТВЕННЫХ ЗНАЧЕНИЯХ С ТРЕТЬЕГО ПО ВЕЛИЧИНЕ И МЕНЬШЕ, ТАК КАК В ДАЛЬНЕЙШЕМ ОНИ НАМ СОВЕРШЕННО НЕ ПОНАДОБЯТСЯ.



Шаг 12

Вспомним, что если принять равными нулю собственные значения начиная с третьего по величине, то будет выполняться следующее приближённое равенство:

$$\begin{pmatrix} 0,68 & 0,65 & 0,80 & 0,11 & 0,01 & 0,14 \\ 0,65 & 0,88 & 0,89 & 0,02 & 0,19 & 0,01 \\ 0,80 & 0,89 & 0,91 & 0,02 & 0,04 & 0,10 \\ 0,11 & 0,02 & 0,02 & 0,81 & 0,82 & 0,77 \\ 0,01 & 0,19 & 0,04 & 0,82 & 0,81 & 0,64 \\ 0,14 & 0,01 & 0,10 & 0,77 & 0,64 & 0,66 \end{pmatrix} \approx \begin{pmatrix} \sqrt{2,55} \times 0,43 & \sqrt{2,11} \times (-0,28) \\ \sqrt{2,55} \times 0,48 & \sqrt{2,11} \times (-0,34) \\ \sqrt{2,55} \times 0,50 & \sqrt{2,11} \times (-0,38) \\ \sqrt{2,55} \times 0,34 & \sqrt{2,11} \times 0,51 \\ \sqrt{2,55} \times 0,34 & \sqrt{2,11} \times 0,47 \\ \sqrt{2,55} \times 0,32 & \sqrt{2,11} \times 0,43 \end{pmatrix} \times$$

См.
стр.
76–78

$$\times \begin{pmatrix} \sqrt{2,55} \times 0,43 & \sqrt{2,55} \times 0,48 & \sqrt{2,55} \times 0,50 & \sqrt{2,55} \times 0,34 & \sqrt{2,55} \times 0,34 & \sqrt{2,55} \times 0,32 \\ \sqrt{2,11} \times (-0,28) & \sqrt{2,11} \times (-0,34) & \sqrt{2,11} \times (-0,38) & \sqrt{2,11} \times 0,51 & \sqrt{2,11} \times 0,47 & \sqrt{2,11} \times 0,43 \end{pmatrix} =$$

$$= \begin{pmatrix} 0,64 & 0,72 & 0,77 & 0,07 & 0,10 & 0,10 \\ 0,72 & 0,83 & 0,88 & 0,05 & 0,08 & 0,08 \\ 0,77 & 0,88 & 0,94 & 0,03 & 0,06 & 0,06 \\ 0,07 & 0,05 & 0,03 & 0,85 & 0,81 & 0,74 \\ 0,10 & 0,08 & 0,06 & 0,81 & 0,77 & 0,70 \\ 0,10 & 0,08 & 0,06 & 0,74 & 0,70 & 0,64 \end{pmatrix}$$

ЧЕТЫРЕ
СТОЛБЦА
И СТРОКИ
ИСЧЕЗАЮТ!



Я помню, мы
изучали это
раньше!

Шаг 13

Заменяем значения на главной диагонали матрицы Шага 11 на значения из главной диагонали матрицы Шага 12.

$$\begin{pmatrix} 0,64 & 0,65 & 0,80 & 0,11 & 0,01 & 0,14 \\ 0,65 & 0,83 & 0,89 & 0,02 & 0,19 & 0,01 \\ 0,80 & 0,89 & 0,94 & 0,02 & 0,04 & 0,10 \\ 0,11 & 0,02 & 0,02 & 0,85 & 0,82 & 0,77 \\ 0,01 & 0,19 & 0,04 & 0,82 & 0,77 & 0,64 \\ 0,14 & 0,01 & 0,10 & 0,77 & 0,64 & 0,66 \end{pmatrix}$$



0,68 0,88 0,91 0,81 0,81 0,66

Шаг 14

Повторяем Шаг 11, но для матрицы с новыми значениями на главной диагонали, то есть находим собственные значения λ и собственные векторы $\{t_1; \dots; t_6\}$, удовлетворяющие следующему уравнению:

$$\begin{pmatrix} 0,68 & 0,65 & 0,80 & 0,11 & 0,01 & 0,14 \\ 0,65 & 0,88 & 0,89 & 0,02 & 0,19 & 0,01 \\ 0,80 & 0,89 & 0,91 & 0,02 & 0,04 & 0,10 \\ 0,11 & 0,02 & 0,02 & 0,81 & 0,82 & 0,77 \\ 0,01 & 0,19 & 0,04 & 0,82 & 0,81 & 0,64 \\ 0,14 & 0,01 & 0,10 & 0,77 & 0,64 & 0,66 \end{pmatrix} \begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{pmatrix} = \lambda \begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{pmatrix}.$$

Собственные векторы находим так, чтобы они удовлетворяли соотношению $t_1^2 + t_2^2 + t_3^2 + t_4^2 + t_5^2 + t_6^2 = 1$.

С помощью программы для анализа данных должны быть получены следующие результаты:

Собственные значения λ	Собственные векторы $\begin{pmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{pmatrix}$
2,54	$\begin{pmatrix} 0,42 \\ 0,47 \\ 0,50 \\ 0,36 \\ 0,35 \\ 0,32 \end{pmatrix}$
2,11	$\begin{pmatrix} -0,28 \\ -0,34 \\ -0,40 \\ 0,52 \\ 0,46 \\ 0,42 \end{pmatrix}$

ЗНАЧИТ,
МЫ ОПЯТЬ ИЩЕМ
СОБСТВЕННЫЕ ЗНАЧЕНИЯ
И СОБСТВЕННЫЕ ВЕКТОРЫ,
НО УЖЕ ДЛЯ МАТРИЦЫ
С ЗАМЕНЁННОЙ ГЛАВНОЙ
ДИАГОНАЛЬЮ, ДА?



Шаг 15

Повторяем раз за разом Шаги 12...14 до тех пор, пока какое-либо значение на главной диагонали матрицы Шага 12, то есть какая-либо общность $1-d_i^2$ не будет готова превзойти 1. Матрицу

$$\begin{pmatrix} \sqrt{\lambda_1} \times t_{11} & \sqrt{\lambda_2} \times t_{12} \\ \sqrt{\lambda_1} \times t_{21} & \sqrt{\lambda_2} \times t_{22} \\ \sqrt{\lambda_1} \times t_{31} & \sqrt{\lambda_2} \times t_{32} \\ \sqrt{\lambda_1} \times t_{41} & \sqrt{\lambda_2} \times t_{42} \\ \sqrt{\lambda_1} \times t_{51} & \sqrt{\lambda_2} \times t_{52} \\ \sqrt{\lambda_1} \times t_{61} & \sqrt{\lambda_2} \times t_{62} \end{pmatrix}, \text{ полученную на последнем цикле, рассматриваем как } \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \\ a_{61} & a_{62} \end{pmatrix}.$$



В случае нашего примера, если один раз повторить Шаги 12–14 для результата, приведённого на предыдущей странице, то значение общности $1-d_3^2$ превзойти 1. Следовательно результат предыдущей страницы, то есть матрицу

$$\begin{pmatrix} \sqrt{2,54} \times 0,42 & \sqrt{2,11} \times (-0,28) \\ \sqrt{2,54} \times 0,47 & \sqrt{2,11} \times (-0,34) \\ \sqrt{2,54} \times 0,50 & \sqrt{2,11} \times (-0,40) \\ \sqrt{2,54} \times 0,36 & \sqrt{2,11} \times 0,52 \\ \sqrt{2,54} \times 0,35 & \sqrt{2,11} \times 0,46 \\ \sqrt{2,54} \times 0,32 & \sqrt{2,11} \times 0,42 \end{pmatrix} \text{ можно интерпретировать как } \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \\ a_{61} & a_{62} \end{pmatrix}.$$

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \\ a_{61} & a_{62} \end{pmatrix} = \begin{pmatrix} \sqrt{\lambda_1} \times t_{11} & \sqrt{\lambda_2} \times t_{12} \\ \sqrt{\lambda_1} \times t_{21} & \sqrt{\lambda_2} \times t_{22} \\ \sqrt{\lambda_1} \times t_{31} & \sqrt{\lambda_2} \times t_{32} \\ \sqrt{\lambda_1} \times t_{41} & \sqrt{\lambda_2} \times t_{42} \\ \sqrt{\lambda_1} \times t_{51} & \sqrt{\lambda_2} \times t_{52} \\ \sqrt{\lambda_1} \times t_{61} & \sqrt{\lambda_2} \times t_{62} \end{pmatrix} = \begin{pmatrix} \sqrt{2,54} \times 0,42 & \sqrt{2,11} \times (-0,28) \\ \sqrt{2,54} \times 0,47 & \sqrt{2,11} \times (-0,34) \\ \sqrt{2,54} \times 0,50 & \sqrt{2,11} \times (-0,40) \\ \sqrt{2,54} \times 0,36 & \sqrt{2,11} \times 0,52 \\ \sqrt{2,54} \times 0,35 & \sqrt{2,11} \times 0,46 \\ \sqrt{2,54} \times 0,32 & \sqrt{2,11} \times 0,42 \end{pmatrix} = \begin{pmatrix} 0,67 & -0,41 \\ 0,74 & -0,49 \\ 0,80 & -0,57 \\ 0,57 & 0,75 \\ 0,55 & 0,66 \\ 0,51 & 0,60 \end{pmatrix}.$$



УРА! МЫ УЖЕ
ПРОШЛИ 15 ШАГОВ!

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \\ a_{61} & a_{62} \end{pmatrix} = \begin{pmatrix} 0,67 & -0,41 \\ 0,74 & -0,49 \\ 0,80 & -0,57 \\ 0,57 & 0,75 \\ 0,55 & 0,66 \\ 0,51 & 0,60 \end{pmatrix}$$

МОЛОДЕЦ!



МАТРИЦУ

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \\ a_{61} & a_{62} \end{pmatrix}$$

НАЗЫВАЮТ МАТРИЦЕЙ
ФАКТОРНЫХ НАГРУЗОК, ИЛИ
МАТРИЦЕЙ ФАКТОРНОГО
ОТБРАЖЕНИЯ.

ЯСНО.

ИТАК, ШАГ 16!

РАССМАТРИВАЯ МАТРИЦУ
ФАКТОРНЫХ НАГРУЗОК,
НАЙДЕННУЮ НА ШАГЕ 15,
КАК ЗНАЧЕНИЯ КООРДИНАТ,
ПОСТРОИМ НА ЕЁ
ОСНОВЕ ТОЧЕЧНЫЙ
ГРАФИК.

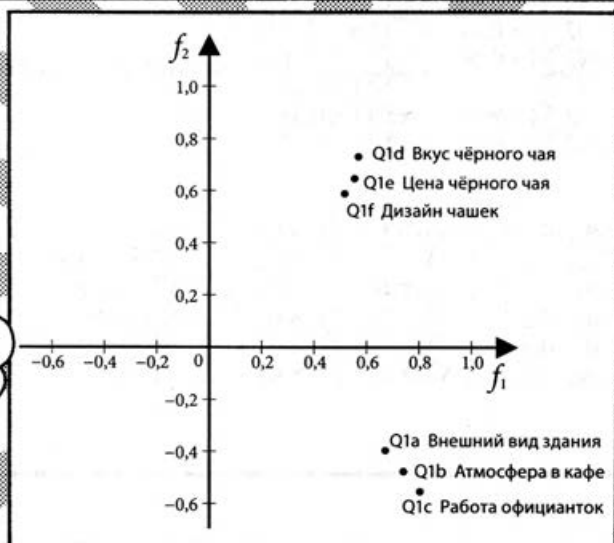
	Координаты
Q1a Внешний вид здания	(0,67 ; -0,41)
Q1b Атмосфера внутри кафе	(0,74 ; -0,49)
Q1c Работа официанток	(0,80 ; -0,57)
Q1d Вкус чёрного чая	(0,57 ; 0,75)
Q1e Цена чёрного чая	(0,55 ; 0,66)
Q1f Дизайн чашек	(0,51 ; 0,60)

ВОТ ОНИ —
ФАКТОРНЫЕ НАГРУЗКИ
ДО ВРАЩЕНИЯ!

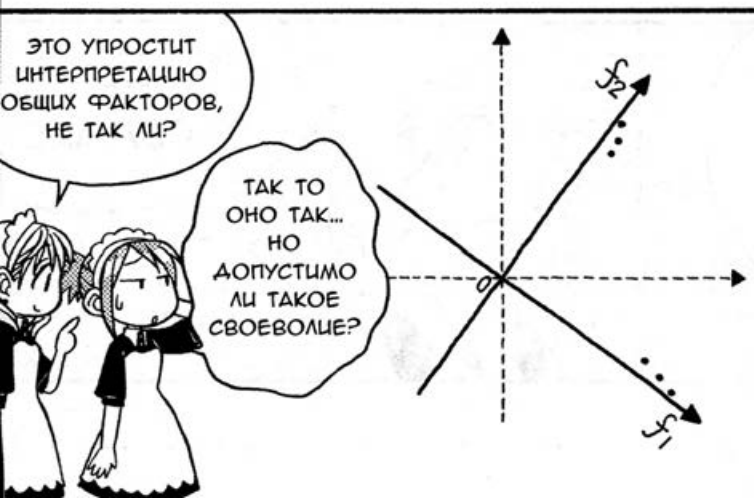
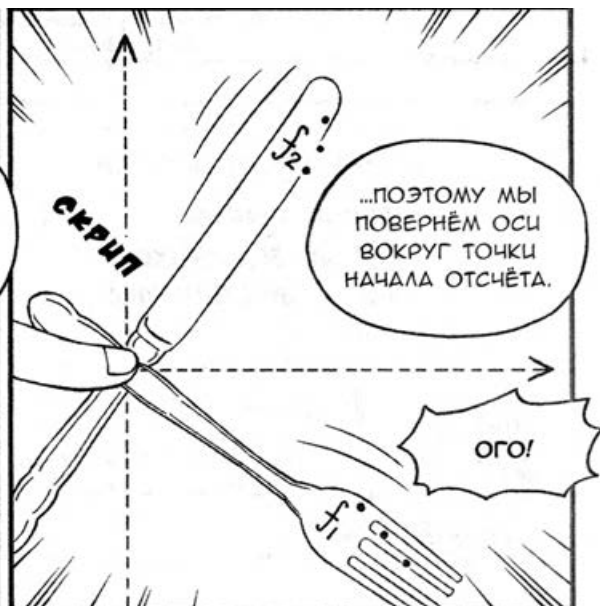
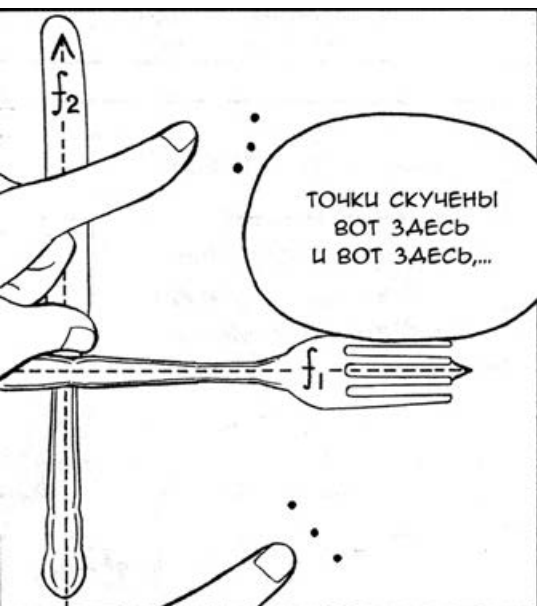


МЫ
СДЕЛАЛИ
ЭТО!

16



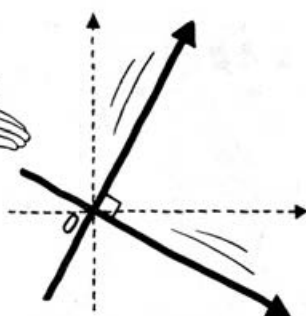
2 РАСЧИТЫВАЕМ ФАКТОРНЫЕ НАГРУЗКИ ПОСЛЕ ВРАЩЕНИЯ





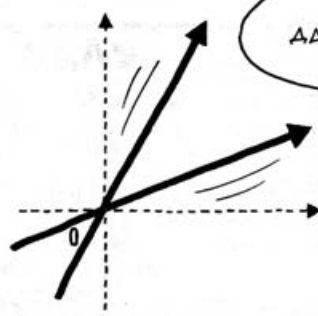
СУЩЕСТВУЮТ
РАЗНЫЕ ВИДЫ
ВРАЩЕНИЯ,...

...НО ВСЕ ОНИ ДЕЛЯТСЯ
НА ОРТОГОНАЛЬНОЕ ВРАЩЕНИЕ
И НАКЛОННОЕ ВРАЩЕНИЕ!



Ортогональное вращение

Оси после поворота пересекаются
под прямым углом



Наклонное вращение

Оси после поворота пересекаются
под непрямым углом

АА?!

Ортогональное вращение

- Метод варимакс
- Метод кватримакс
- Метод бикватримакс

Наклонное вращение

- Метод промакс
- Метод кватримин
- Метод бикватримин
- Метод коваримин



ОНИ, В СВОЮ ОЧЕРЕДЬ,
ДЕЛЯТСЯ НА РАЗНЫЕ
МЕТОДЫ.



ТАК МНОГО?!

САМЫЙ ИЗВЕСТНЫЙ МЕТОД
ОРТОГОНАЛЬНОГО ВРАЩЕНИЯ —
ЭТО МЕТОД ВАРИМАКС,
А НАКЛОННОГО —
МЕТОД ПРОМАКС.



ОРТОГОНАЛЬНОЕ
ВРАЩЕНИЕ
ВАРИМАКС

КАКОЕ
КРАСИВОЕ
НАЗВАНИЕ!

СЕГОДНЯ Я ОБЪЯСНЮ, КАК
ПРОВОДЯТСЯ ВЫЧИСЛЕНИЯ
САМЫМ ИЗВЕСТНЫМ ИЗ ВСЕХ
МЕТОДОВ ВРАЩЕНИЯ —
МЕТОДОМ ВАРИМАКС.

ХОРОШО!

МЫ ТОЛЬКО ЧТО НАШЛИ
МАТРИЦУ ФАКТОРНЫХ
НАГРУЗОК, ПОМНИШЬ?

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \\ a_{61} & a_{62} \end{pmatrix} = \begin{pmatrix} 0.67 & -0.41 \\ 0.74 & -0.49 \\ 0.80 & -0.57 \\ 0.57 & 0.75 \\ 0.55 & 0.66 \\ 0.51 & 0.60 \end{pmatrix}$$

АА.

ВОТ ЭТУ ФОРМУЛУ,
ОСНОВАННУЮ
НА МАТРИЦЕ
ФАКТОРНЫХ НАГРУЗОК,...



$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \dots & a_{61} \\ a_{12} & a_{22} & \dots & a_{62} \end{pmatrix} = \begin{pmatrix} 0.67 & -0.41 \\ 0.74 & -0.49 \\ \vdots & \vdots \\ 0.51 & 0.60 \end{pmatrix} \begin{pmatrix} 0.67 & 0.74 & \dots & 0.51 \\ -0.41 & -0.49 & \dots & 0.60 \end{pmatrix}$$

...МОЖНО
ПРЕОБРАЗОВАТЬ
СЛЕДУЮЩИМ
ОБРАЗОМ.



$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \dots & a_{61} \\ a_{12} & a_{22} & \dots & a_{62} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \dots & a_{61} \\ a_{12} & a_{22} & \dots & a_{62} \end{pmatrix} \quad \leftarrow \text{См. стр. 69}$$

$$= \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{pmatrix} \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \dots & a_{61} \\ a_{12} & a_{22} & \dots & a_{62} \end{pmatrix} \quad \leftarrow \text{См. стр. 72}$$

$$= \left\{ \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{pmatrix} \begin{pmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{pmatrix} \right\} \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \dots & a_{61} \\ a_{12} & a_{22} & \dots & a_{62} \end{pmatrix} \right\} \quad \leftarrow \text{Заклучаем в скобки}$$

$$= \left\{ \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \right\} \left\{ \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \dots & a_{61} \\ a_{12} & a_{22} & \dots & a_{62} \end{pmatrix} \right\} \quad \leftarrow \begin{matrix} \text{Используем} \\ \cos(-\theta) = \cos\theta \\ \sin(-\theta) = -\sin\theta \end{matrix}$$

$$= \begin{pmatrix} a_{11} \cos\theta - a_{12} \sin\theta & a_{11} \sin\theta - a_{12} \cos\theta \\ a_{21} \cos\theta - a_{22} \sin\theta & a_{21} \sin\theta - a_{22} \cos\theta \\ \vdots & \vdots \\ a_{61} \cos\theta - a_{62} \sin\theta & a_{61} \sin\theta - a_{62} \cos\theta \end{pmatrix} \begin{pmatrix} a_{11} \cos\theta - a_{12} \sin\theta & a_{21} \cos\theta - a_{22} \sin\theta & \dots & a_{61} \cos\theta - a_{62} \sin\theta \\ a_{11} \sin\theta - a_{12} \cos\theta & a_{21} \sin\theta - a_{22} \cos\theta & \dots & a_{61} \sin\theta - a_{62} \cos\theta \end{pmatrix}$$

$$= \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{61} & b_{62} \end{pmatrix} \begin{pmatrix} b_{11} & b_{21} & \dots & b_{61} \\ b_{12} & b_{22} & \dots & b_{62} \end{pmatrix}$$

Для облегчения восприятия вводим замену:

$$\begin{pmatrix} a_{11} \cos\theta - a_{12} \sin\theta & a_{11} \sin\theta - a_{12} \cos\theta \\ a_{21} \cos\theta - a_{22} \sin\theta & a_{21} \sin\theta - a_{22} \cos\theta \\ \vdots & \vdots \\ a_{61} \cos\theta - a_{62} \sin\theta & a_{61} \sin\theta - a_{62} \cos\theta \end{pmatrix} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{61} & b_{62} \end{pmatrix} =$$

$$= \begin{pmatrix} 0,67 \cos\theta - (-0,41) \sin\theta & 0,67 \sin\theta - (-0,41) \cos\theta \\ 0,74 \cos\theta - (-0,49) \sin\theta & 0,74 \sin\theta - (-0,49) \cos\theta \\ \vdots & \vdots \\ 0,51 \cos\theta - 0,60 \sin\theta & 0,51 \sin\theta - 0,60 \cos\theta \end{pmatrix}$$



ТО ЕСТЬ, ХОТЯ МАТРИЦА

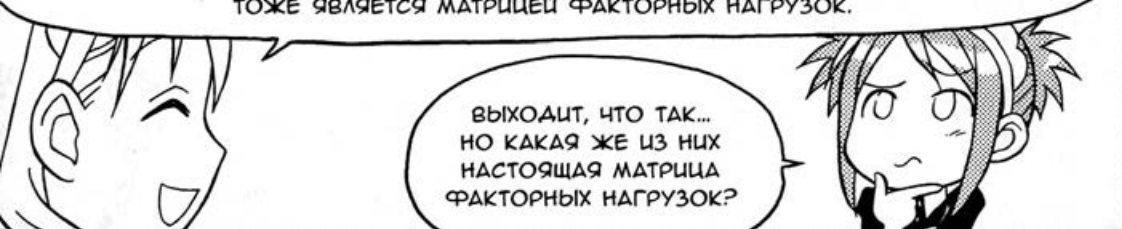
$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{pmatrix} = \begin{pmatrix} 0,67 & -0,41 \\ 0,74 & -0,49 \\ \vdots & \vdots \\ 0,51 & 0,60 \end{pmatrix}$$

НЕСОМНЕННО, ЯВЛЯЕТСЯ
МАТРИЦЕЙ ФАКТОРНЫХ
НАГРУЗОК,...



$$\dots \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{61} & b_{62} \end{pmatrix} = \begin{pmatrix} 0,67 \cos\theta - (-0,41) \sin\theta & 0,67 \sin\theta - (-0,41) \cos\theta \\ 0,74 \cos\theta - (-0,49) \sin\theta & 0,74 \sin\theta - (-0,49) \cos\theta \\ \vdots & \vdots \\ 0,51 \cos\theta - 0,60 \sin\theta & 0,51 \sin\theta - 0,60 \cos\theta \end{pmatrix}$$

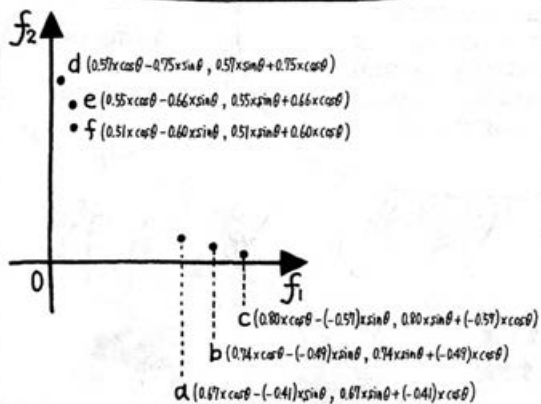
ТОЖЕ ЯВЛЯЕТСЯ МАТРИЦЕЙ ФАКТОРНЫХ НАГРУЗОК.



ВЫХОДИТ, ЧТО ТАК...
НО КАКАЯ ЖЕ ИЗ НИХ
НАСТОЯЩАЯ МАТРИЦА
ФАКТОРНЫХ НАГРУЗОК?

ВОТ ЭТА:

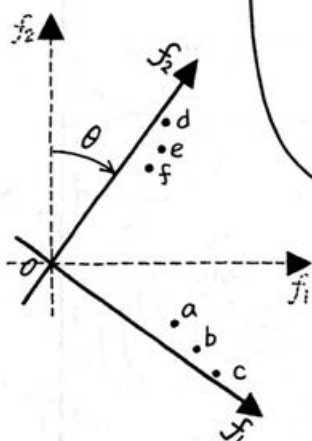
$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{61} & b_{62} \end{pmatrix}$$



МАТРИЦА

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{pmatrix}$$

ЯВЛЯЕТСЯ
ВСЕГО ЛИШЬ
ОСОБЫМ
СЛУЧАЕМ
ДЛЯ УГЛА θ ,
РАВНОГО 0° .



ВОТ КАК?!



В МЕТОДЕ ВАРИМАКС, ЕСЛИ ОБЪЯСНЯТЬ НА ОСНОВЕ ЭТОГО ПРИМЕРА, ПРИБАВЛЯЮТ К СУММЕ КВАДРАТОВ ОТКЛОНЕНИЙ "КВАДРАТА ФАКТОРНОЙ НАГРУЗКИ ПЕРВОГО ОБЩЕГО ФАКТОРА" СУММУ КВАДРАТОВ ОТКЛОНЕНИЙ "КВАДРАТА ФАКТОРНОЙ НАГРУЗКИ ВТОРОГО ОБЩЕГО ФАКТОРА".



Так...
Сумма квадратов отклонений x находится суммированием $(x_i - \bar{x})^2$, значит...



КОНКРЕТНО ГОВОРЯ, В МЕТОДЕ ВАРИМАКС ВРАЩАЮТ ОСИ НА ТАКОЙ УГОЛ θ , ПРИ КОТОРОМ ВОТ ЭТО ВЫРАЖЕНИЕ БУДЕТ МАКСИМАЛЬНЫМ.

ОГО!
ВОТ ОНО ЧТО?!

$$\left(b_{11} - \frac{b_{11}^2 + b_{21}^2 + \dots + b_{61}^2}{6}\right)^2 + \dots + \left(b_{61} - \frac{b_{11}^2 + b_{21}^2 + \dots + b_{61}^2}{6}\right)^2 + \left(b_{12} - \frac{b_{12}^2 + b_{22}^2 + \dots + b_{62}^2}{6}\right)^2 + \dots + \left(b_{62} - \frac{b_{12}^2 + b_{22}^2 + \dots + b_{62}^2}{6}\right)^2$$

ОДНАКО НА САМОМ ДЕЛЕ СУЩЕСТВУЕТ ДВА ВИДА МЕТОДА ВАРИМАКС: ВАРИМАКС БЕЗ НОРМАЛИЗАЦИИ И НОРМАЛИЗОВАННЫЙ ВАРИМАКС.

ТОЛЬКО ЧТО МЫ ИСПОЛЬЗОВАЛИ ВАРИМАКС БЕЗ НОРМАЛИЗАЦИИ.

СЕЙЧАС МЕТОДОМ ВАРИМАКС БОЛЬШЕ ПРИНЯТО НАЗЫВАТЬ НОРМАЛИЗОВАННЫЙ ВАРИМАКС.

А ЧТО ОН ИЗ СЕБЯ ПРЕДСТАВЛЯЕТ?

ЧТО?

Исходные факторные нагрузки

Нормированные факторные нагрузки

СТУК-СТУК

$$\left(\frac{b_{11}^2}{b_{11}^2 + b_{12}^2} - \frac{b_{11}^2}{b_{11}^2 + b_{12}^2} + \dots + \frac{b_{61}^2}{b_{61}^2 + b_{62}^2} \right)^2 + \dots + \left(\frac{b_{61}^2}{b_{61}^2 + b_{62}^2} - \frac{b_{11}^2}{b_{11}^2 + b_{12}^2} + \dots + \frac{b_{61}^2}{b_{61}^2 + b_{62}^2} \right)^2$$

$$+ \left(\frac{b_{12}^2}{b_{11}^2 + b_{12}^2} - \frac{b_{12}^2}{b_{11}^2 + b_{12}^2} + \dots + \frac{b_{62}^2}{b_{61}^2 + b_{62}^2} \right)^2 + \dots + \left(\frac{b_{62}^2}{b_{61}^2 + b_{62}^2} - \frac{b_{12}^2}{b_{11}^2 + b_{12}^2} + \dots + \frac{b_{62}^2}{b_{61}^2 + b_{62}^2} \right)^2$$

Общность $1 - d_1^2 (= h_1^2 = a_{11}^2 + a_{12}^2)$

Общность $1 - d_6^2 (= h_6^2 = a_{61}^2 + a_{62}^2)$

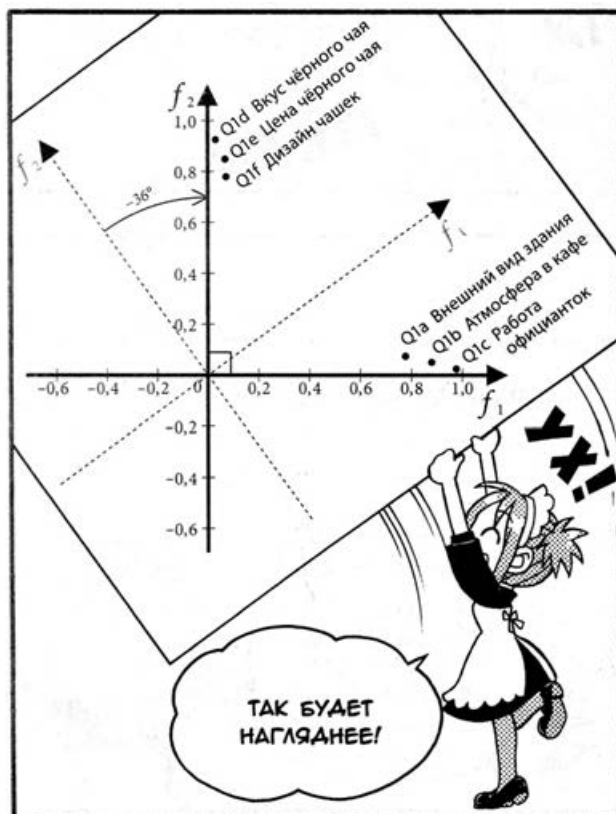
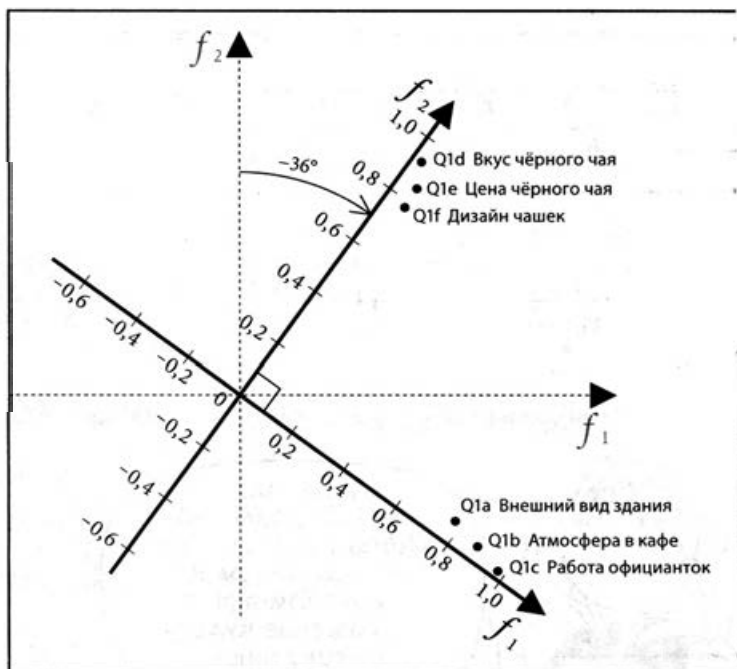


$$\begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{61} & b_{62} \end{pmatrix} = \begin{pmatrix} 0,67 \cos(-36^\circ) - (-0,41) \sin(-36^\circ) & 0,67 \sin(-36^\circ) - (-0,41) \cos(-36^\circ) \\ 0,74 \cos(-36^\circ) - (-0,49) \sin(-36^\circ) & 0,74 \sin(-36^\circ) - (-0,49) \cos(-36^\circ) \\ \vdots & \vdots \\ 0,51 \cos(-36^\circ) - 0,60 \sin(-36^\circ) & 0,51 \sin(-36^\circ) - 0,60 \cos(-36^\circ) \end{pmatrix} = \begin{pmatrix} 0,78 & 0,07 \\ 0,89 & 0,04 \\ 0,99 & 0,01 \\ 0,01 & 0,94 \\ 0,05 & 0,86 \\ 0,06 & 0,79 \end{pmatrix}$$



В ГРАФИЧЕСКОЙ
ФОРМЕ ЭТО
ВЫГЛЯДИТ ТАК.

И ВПРАВАУ!
КАЖЕТСЯ, ЭТО
БУДЕТ ЛЕГЧЕ
ИНТЕРПРЕ-
ТИРОВАТЬ.



ТАК БУДЕТ
НАГЛЯДНЕЕ!

- | | |
|-------|---------------------------------|
| Шаг 1 | Вращаем по общим факторам 1 и 2 |
| Шаг 2 | Вращаем по общим факторам 1 и 3 |
| Шаг 3 | Вращаем по общим факторам 1 и 4 |
| Шаг 4 | Вращаем по общим факторам 2 и 3 |
| Шаг 5 | Вращаем по общим факторам 2 и 4 |
| Шаг 6 | Вращаем по общим факторам 3 и 4 |

КСТАТИ, ЕСЛИ БЫ МЫ
НАЗНАЧИЛИ ЧИСЛО
ОБЩИХ ФАКТОРОВ
РАВНЫМ 4, ТО
ВРАЩАЛИ БЫ ВОТ
ТАК, ОТДЕЛЬНО ПО
КАЖДОЙ ПАРЕ
ОСЕИ.

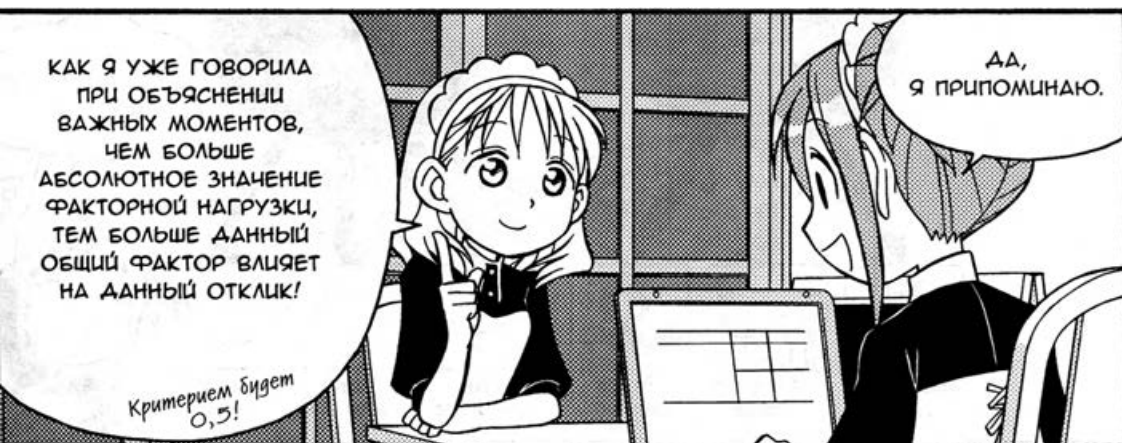
ПОНЯТНО.

3 ИНТЕРПРЕТИРУЕМ КАЖДЫЙ ИЗ ОБЩИХ ФАКТОРОВ

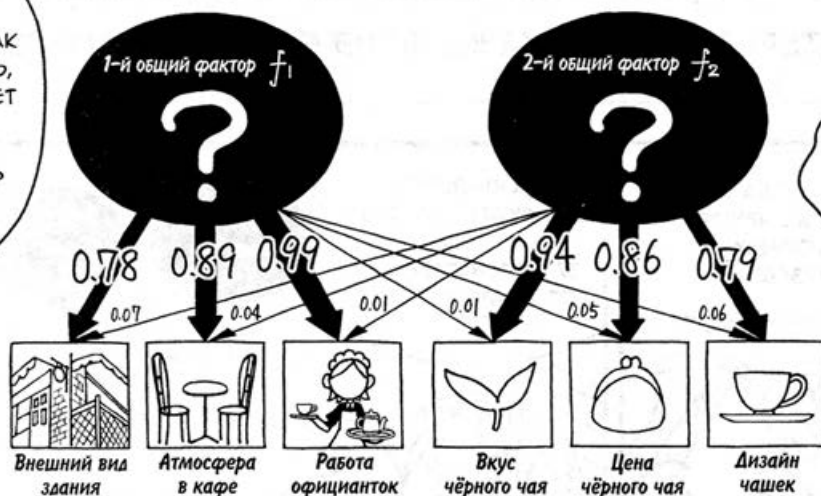


Я СВЕЛА В ОДНУ ТАБЛИЦУ НАЙДЕННЫЕ НАМИ ФАКТОРНЫЕ НАГРУЗКИ И ВЫДЕЛИЛА ЦВЕТОМ ТЕ ИЗ НИХ, КОТОРЫЕ ПО АБСОЛЮТНОМУ ЗНАЧЕНИЮ НЕ МЕНЬШЕ 0,5.

	1-й общий фактор	2-й общий фактор
Q1a Внешний вид здания	0,78	0,07
Q1b Атмосфера в кафе	0,89	0,04
Q1c Работа официанток	0,99	0,01
Q1d Вкус чёрного чая	0,01	0,94
Q1e Цена чёрного чая	0,05	0,86
Q1f Дизайн чашек	0,06	0,79



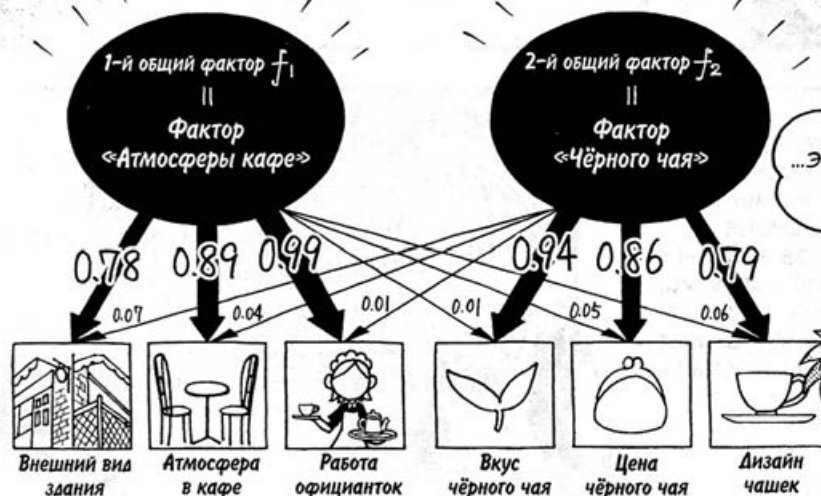
ИТАК, РУКИ, КАК ТЫ ДУМАЕШЬ, ЧТО ОЗНАЧАЕТ ИЗ ОБЩИХ ФАКТОРОВ?



ХММ...

ОДИН ИЗ НИХ СИЛЬНО ВЛИЯЕТ НА ОТКЛИКИ "ВНЕШНИЙ ВИД ЗДАНИЯ", "АТМОСФЕРА В КАФЕ", "РАБОТА ОФИЦИАНТОК"...

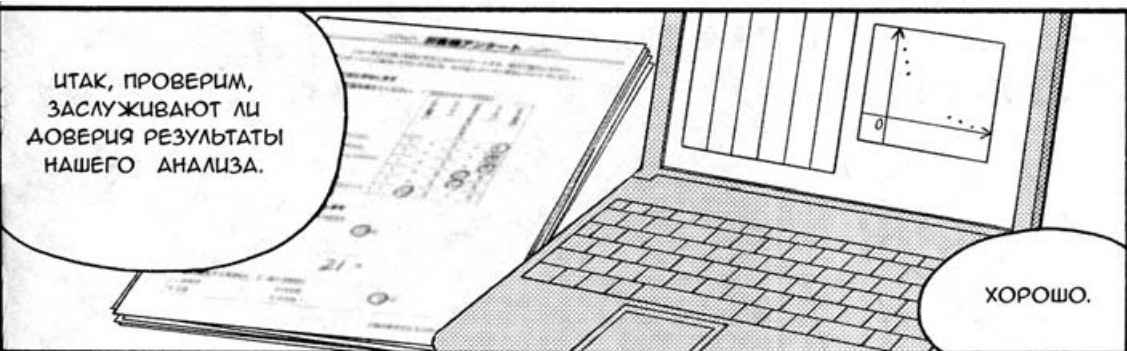
...А ДРУГОЙ — НА ОТКЛИКИ "ВКУС ЧЁРНОГО ЧАЯ", "ЦЕНА ЧЁРНОГО ЧАЯ", "ДИЗАЙН ЧАШЕК", ЗНАЧИТ...



...ЭТО ВЫГЛЯДИТ ВОТ ТАК!

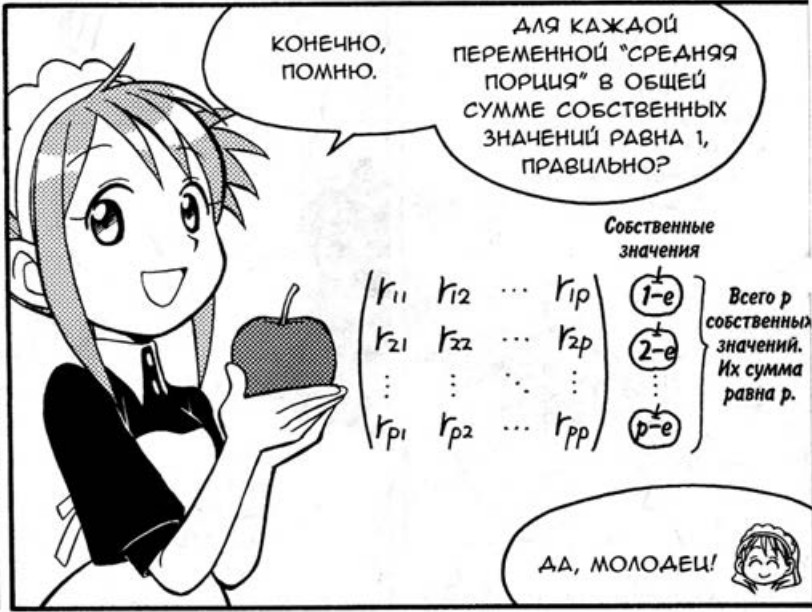


4 ПРОВЕРЯЕМ ТОЧНОСТЬ РЕЗУЛЬТАТОВ АНАЛИЗА





ПОМНИШЬ ЛИ ТЫ,
ЧТО СУММА ВСЕХ
СОБСТВЕННЫХ
ЗНАЧЕНИЙ МАТРИЦЫ
КОРРЕЛЯЦИИ
ДЛЯ p ПЕРЕМЕННЫХ
РАВНА p ?



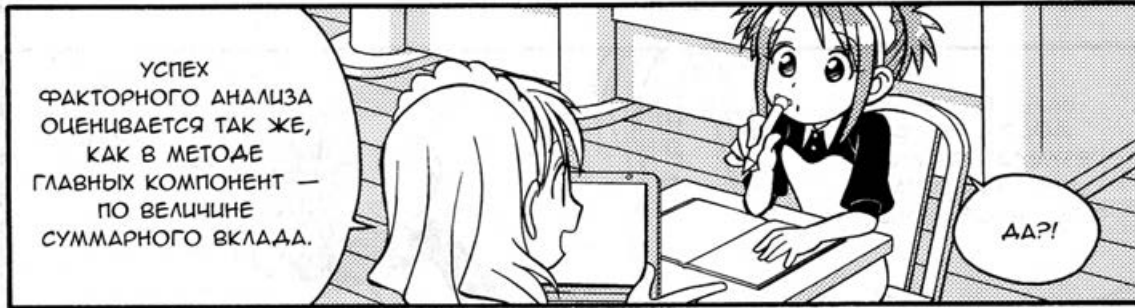
КОНЕЧНО,
ПОМНЮ.

ДЛЯ КАЖДОЙ
ПЕРЕМЕННОЙ "СРЕДНЯЯ
ПОРЦИЯ" В ОБЩЕЙ
СУММЕ СОБСТВЕННЫХ
ЗНАЧЕНИЙ РАВНА 1,
ПРАВИЛЬНО?

Собственные значения

r_{11}	r_{12}	...	r_{1p}	$1-\epsilon$	}	Всего p собственных значений. Их сумма равна p .
r_{21}	r_{22}	...	r_{2p}	$2-\epsilon$		
\vdots	\vdots	\ddots	\vdots	\vdots		
r_{p1}	r_{p2}	...	r_{pp}	$p-\epsilon$		

АА, МОЛОДЕЦ!



УСПЕХ
ФАКТОРНОГО АНАЛИЗА
ОЦЕНИВАЕТСЯ ТАК ЖЕ,
КАК В МЕТОДЕ
ГЛАВНЫХ КОМПОНЕНТ —
ПО ВЕЛИЧИНЕ
СУММАРНОГО ВКЛАДА.

АА?!



СТУК-СТУК



АА?!

ДЛЯ НАШЕГО ПРИМЕРА,
"УДЕЛЬНЫЙ ВКЛАД i -ГО ОБЩЕГО
ФАКТОРА" ВЫГЛЯДИТ ВОТ ТАК.

ЗНАЧИТ, ЗАДЕСЬ
ИСПОЛЬЗУЮТСЯ
ФАКТОРНЫЕ
НАГРУЗКИ?

Удельный вклад i -го общего фактора = $\frac{b_{1i}^2 + b_{2i}^2 + b_{3i}^2 + b_{4i}^2 + b_{5i}^2 + b_{6i}^2}{6} \times 100$

ТОГДА СУММАРНЫЙ ВКЛАД — ЭТО СУММА УДЕЛЬНЫХ ВКЛАДОВ ОБЩИХ ФАКТОРОВ, НАЧИНАЯ С ПЕРВОГО?

$$\begin{aligned} & \text{Удельный вклад} \\ & \text{1-го общего фактора} \\ & = \frac{b_{11}^2 + b_{21}^2 + b_{31}^2 + b_{41}^2 + b_{51}^2 + b_{61}^2}{6} \times 100 \\ & = \frac{(0.78)^2 + (0.89)^2 + (0.99)^2 + (0.01)^2 + (0.05)^2 + (0.06)^2}{6} \times 100 \\ & = 39 \text{ — } \times 100 \end{aligned}$$

ДА, ЭТО ТАК.

КОНКРЕТНЫЕ ВЫЧИСЛЕНИЯ ВЫГЛЯДЯТ ВОТ ТАК.



ДА, ПОНЯТНО.

	$b_{11}^2 + b_{21}^2 + \dots + b_{61}^2$	Удельные вклады	Суммарные вклады
1-й общий фактор	2.39	$\frac{2.39}{6} \times 100 = 39.8 (\%)$	$\frac{2.39}{6} \times 100 = 39.8 (\%)$
2-й общий фактор	2.26	$\frac{2.26}{6} \times 100 = 37.7 (\%)$	$\frac{2.39}{6} \times 100 + \frac{2.26}{6} \times 100 = 77.4 (\%)$

ЧЕМ БОЛЬШЕ СУММАРНЫЙ ВКЛАД, ТЕМ ЛУЧШЕ?



ДА.

ОДНАКО СЧИТАЕТСЯ НЕПЛОХО, ЕСЛИ СУММАРНЫЙ ВКЛАД ДЛЯ ЗАДАННОГО АНАЛИТИКОМ ЧИСЛА ОБЩИХ ФАКТОРОВ СОСТАВИТ 50%.

...ЧТО В НАШЕМ СЛУЧАЕ АНАЛИЗ БЫЛ УСПЕШНЫМ, ТАК КАК СУММАРНЫЙ ВКЛАД РАВЕН 77,4%!

ОДНАКО ТОЧНОГО СТАТИСТИЧЕСКОГО КРИТЕРИЯ, ПОЗВОЛЯЮЩЕГО СЧИТАТЬ АНАЛИЗ УСПЕШНЫМ, НЕ СУЩЕСТВУЕТ.



ТАК ЖЕ, КАК И В МЕТОДЕ ГЛАВНЫХ КОМПОНЕНТ, ДА?



ПОНЯТНО. И ЭТО ОЗНАЧАЕТ,...



ДА, ИМЕННО ТАК!

5 НАХОДИМ ФАКТОРНЫЕ ЗНАЧЕНИЯ И ВЫЯВЛЯЕМ ОСОБЕННОСТИ КАЖДОГО ЭКЗЕМПЛЯРА



ПОМНИШЬ ЛИ ТЫ, ЧТО МЫ ПРОВОДИЛИ НОРМИРОВКУ ПО КАЖДОЙ ИЗ ПЕРЕМЕННЫХ ВОТ ТАКИМ ОБРАЗОМ?



ДА. МЫ СДЕЛАЛИ ЭТО В НАЧАЛЕ ФАКТОРНОГО АНАЛИЗА.



ВЫЧИСЛЕНИЕ ФАКТОРНЫХ ЗНАЧЕНИЙ РЕГРЕССИОННЫМ МЕТОДОМ, ЕСЛИ ОБЪЯСНЯТЬ ВКРАТЦЕ, ВЫГЛЯДИТ ВОТ ТАК.

ДА...?

Респонденты	Q1a (нп), u_1	...	Q1f (нп), u_6
А.	1,2	...	-0,9
⋮	⋮	⋮	⋮
П.	-1,2	...	0,6

Факторное значение 1-го общего фактора для респондента А.	$f_{A1} = w_{11} \times 1,2 + \dots + w_{61} \times (-0,9)$
Факторное значение 2-го общего фактора для респондента П.	$f_{P2} = w_{12} \times (-1,2) + \dots + w_{62} \times 0,6$

Данные, нормированные по каждой из переменных (см. стр. 155)

Матрица, обратная к матрице корреляции (см. стр. 166)

Матрица факторных нагрузок после вращения (см. стр.179)

$$\begin{pmatrix} f_{A1} & f_{A2} \\ f_{B1} & f_{B2} \\ f_{C1} & f_{C2} \\ f_{D1} & f_{D2} \\ f_{E1} & f_{E2} \\ f_{Ж1} & f_{Ж2} \\ f_{З1} & f_{З2} \\ f_{И1} & f_{И2} \\ f_{К1} & f_{К2} \\ f_{Л1} & f_{Л2} \\ f_{М1} & f_{М2} \\ f_{Н1} & f_{Н2} \\ f_{О1} & f_{О2} \\ f_{П1} & f_{П2} \end{pmatrix} = \begin{pmatrix} 1,2 & 1,6 & 1,3 & 0,4 & 0,5 & -0,9 \\ 1,2 & 0,8 & 1,3 & -1,2 & -1,0 & -0,9 \\ 0,4 & 0,8 & 0,3 & 0,4 & 0,5 & 0,6 \\ -1,2 & 0,1 & 0,3 & -0,4 & -0,3 & -0,1 \\ -0,4 & 0,1 & -0,6 & -0,4 & 0,5 & -1,6 \\ 1,2 & 0,8 & 1,3 & -0,4 & -1,0 & -0,1 \\ 1,2 & 1,6 & 1,3 & 0,4 & 1,3 & 1,3 \\ -0,4 & -1,5 & -1,6 & 1,3 & 0,5 & 0,6 \\ 0,4 & -1,5 & -0,6 & -0,4 & -1,0 & -0,1 \\ -2,0 & -0,7 & -1,6 & -1,2 & -1,0 & -0,9 \\ -0,4 & -0,7 & -0,6 & -2,1 & -1,8 & -1,6 \\ 0,4 & 0,1 & 0,3 & 0,4 & -0,3 & 0,6 \\ -0,4 & -0,7 & -0,6 & 0,4 & 1,3 & 1,3 \\ 0,4 & 0,1 & 0,3 & 1,3 & 0,5 & 1,3 \\ -1,2 & -0,7 & -0,6 & 1,3 & 1,3 & 0,6 \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0,65 & 0,80 & 0,11 & 0,01 & 0,14 \\ 0,65 & 1 & 0,89 & 0,02 & 0,19 & 0,01 \\ 0,80 & 0,89 & 1 & 0,02 & 0,04 & 0,10 \\ 0,11 & 0,02 & 0,02 & 1 & 0,82 & 0,77 \\ 0,01 & 0,19 & 0,04 & 0,82 & 1 & 0,64 \\ 0,14 & 0,01 & 0,10 & 0,77 & 0,64 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0,78 & 0,07 \\ 0,89 & 0,04 \\ 0,99 & 0,01 \\ 0,01 & 0,94 \\ 0,05 & 0,86 \\ 0,06 & 0,79 \end{pmatrix}$$

Это и есть

$$\begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \\ w_{31} & w_{32} \\ w_{41} & w_{42} \\ w_{51} & w_{52} \\ w_{61} & w_{62} \end{pmatrix} !$$

ОНКРЕТНЫЕ РАСЧЁТЫ ВЫГЛЯДЯТ ВОТ ТАК!

$$\begin{pmatrix} 1,2 & 1,6 & 1,3 & 0,4 & 0,5 & -0,9 \\ 1,2 & 0,8 & 1,3 & -1,2 & -1,0 & -0,9 \\ 0,4 & 0,8 & 0,3 & 0,4 & 0,5 & 0,6 \\ -1,2 & 0,1 & 0,3 & -0,4 & -0,3 & -0,1 \\ -0,4 & 0,1 & -0,6 & -0,4 & 0,5 & -1,6 \\ 1,2 & 0,8 & 1,3 & -0,4 & -1,0 & -0,1 \\ 1,2 & 1,6 & 1,3 & 0,4 & 1,3 & 1,3 \\ -0,4 & -1,5 & -1,6 & 1,3 & 0,5 & 0,6 \\ 0,4 & -1,5 & -0,6 & -0,4 & -1,0 & -0,1 \\ -2,0 & -0,7 & -1,6 & -1,2 & -1,0 & -0,9 \\ -0,4 & -0,7 & -0,6 & -2,1 & -1,8 & -1,6 \\ 0,4 & 0,1 & 0,3 & 0,4 & -0,3 & 0,6 \\ -0,4 & -0,7 & -0,6 & 0,4 & 1,3 & 1,3 \\ 0,4 & 0,1 & 0,3 & 1,3 & 0,5 & 1,3 \\ -1,2 & -0,7 & -0,6 & 1,3 & 1,3 & 0,6 \end{pmatrix}$$

$$\begin{pmatrix} -0,014 & -0,024 \\ -0,047 & -0,005 \\ 1,048 & 0,001 \\ -0,001 & 0,611 \\ 0,092 & 0,259 \\ -0,104 & 0,155 \end{pmatrix}$$

ВОТ ЭТО ДА! НЕСЛАБАЯ МАТРИЦА!

Я свела результаты
вычислений в одну
таблицу.



Респонденты	1-й общий фактор f_1	2-й общий фактор f_2
А.	1,38	0,24
Б.	1,27	-1,19
В.	0,28	0,49
Г.	0,34	-0,30
Д.	-0,45	-0,35
Е.	1,20	-0,56
Ж.	1,22	0,78
З.	-1,60	1,03
И.	-0,68	-0,53
К.	-1,61	-1,11
Л.	-0,62	-1,97
М.	0,24	0,29
Н.	-0,64	0,83
О.	0,24	1,12
П.	-0,56	1,25
СА	0	0

ОБРАТИ ВНИМАНИЕ
НА ФАКТОРНЫЕ
ЗНАЧЕНИЯ ПЕРВОГО
ОБЩЕГО ФАКТОРА.

У КОГО ЭТО
ЗНАЧЕНИЕ
МАКСИМАЛЬНО?



ТАК...

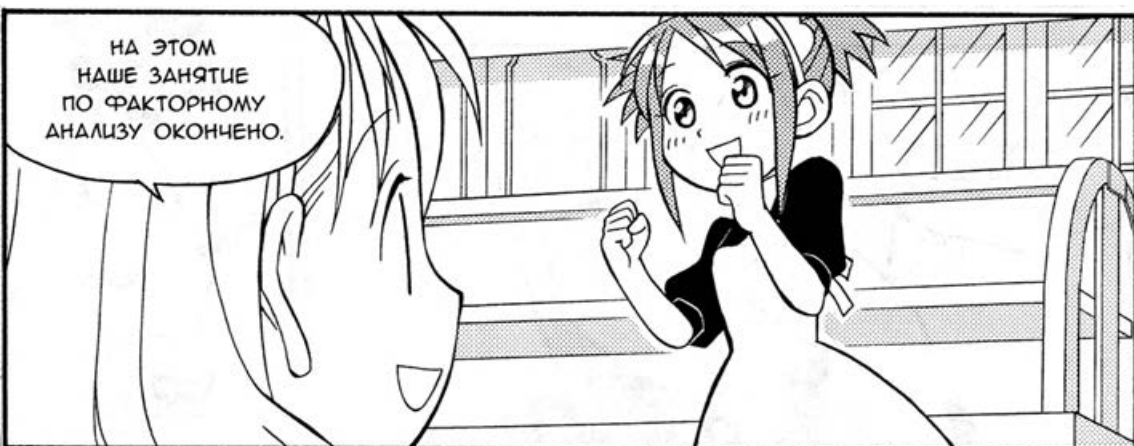
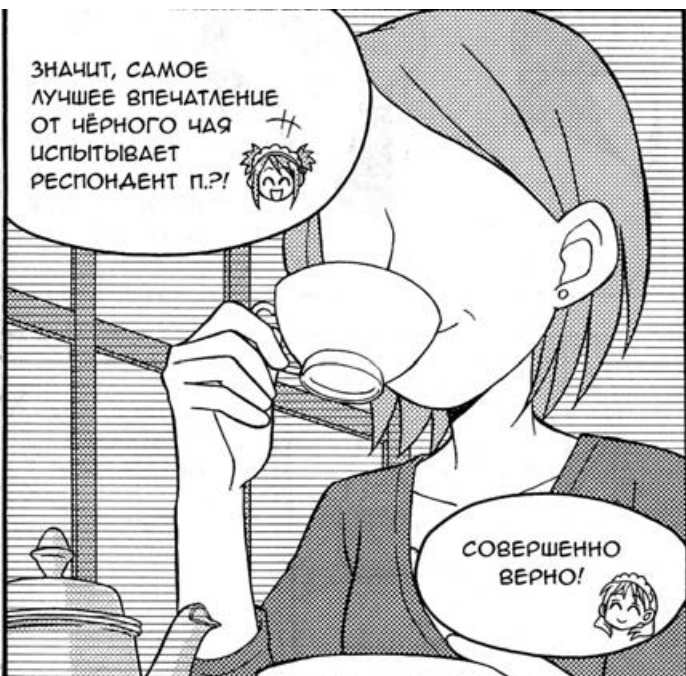
Иденты	1-й общий фактор f_1	2-
А.	1.38	
Б.	1.27	
В.	0.28	
Г.	0.34	
Д.	-0.45	
Е.		
Ж.		
З.		
И.		

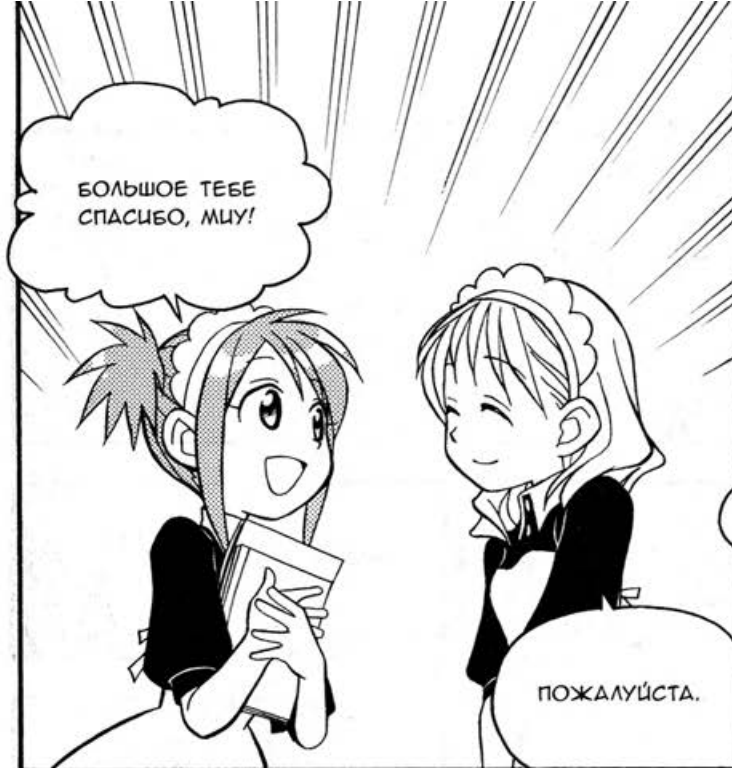
У РЕСПОН-
ДЕНТА А., ДА?

ЭТО ПОЗВОЛЯЕТ
СКАЗАТЬ, ЧТО
САМОЕ ХОРОШЕЕ
ВПЕЧАТЛЕНИЕ ОТ
"АТМОСФЕРЫ КАФЕ"
ИСПЫТЫВАЕТ
РЕСПОНДЕНТ А.



ИНТЕРЕСНО!





Одна маркетинговая компания

НАЧАЛЬНИК
ОТДЕЛА
ТАКАКУ, ВАС
К ТЕЛЕФОНУ.

ДА,
Я СЛУШАЮ...

ЧТО?!
УЖЕ
В АЭРОПОРТУ?!

НО ПОЧЕМУ?
ВЕДЬ ЭТУ РАБОТУ
НЕВОЗМОЖНО
ЗАКОНЧИТЬ ТАК
БЫСТРО...

...

РУЦ,
ВЫТРИ
СЛЁЗЫ.

ХА-ХА-ХА
ИЗВИНИ.

ОЙ, ПРОСТИ,
МОБИЛЬНИК.

АЛЛО...
А, ЭТО ТЫ,
БРАТЕЦ?

ЧТО?!
ПРАВДА?!

РУЦ!
ЯМАМОТО-САН
УЖЕ
В АЭРОПОРТУ!!

КАК?!

4. ВЫБОРКА В ПРИМЕРЕ ДАННОЙ ГЛАВЫ

В примере данной главы генеральная совокупность и выборка определяются следующим образом.

Генеральная совокупность	Все посетители кафе NORNS
Выборка	Посетившие кафе NORNS с 15:00 до 16:00 в тот день, когда Миу и Руи проводили анкетирование.

Очевидно, что эта выборка сформирована не методом простого случайного выбора, а методом преднамеренного выбора, основанного на произвольном решении Миу и Руи.

Наверное, многие читатели обвинят меня в непоследовательности: сказав в главе 1, что нет смысла в выборке, которая не является «точной миниатюрой генеральной совокупности», автор сам же приводит противоречащий этому пример при объяснении основной темы данной книги — факторного анализа. Простите меня, пожалуйста. Однако осмелюсь заметить, что при анализе данных, например, в маркетинге, подобная ситуация неизбежна. Поэтому приходится рассматривать выборку, на самом деле сформированную методом преднамеренного выбора, как сформированную методом простого случайного выбора. Без этого невозможно было бы провести не только факторный анализ, но никакой анализ вообще.

Я считаю, что такой подход, если, конечно, не доводить его до абсурда, вполне приемлем для практического использования, кроме исследований в области науки и техники, при наличии согласия окружающих. Поэтому при формировании выборки всё время размышляйте о том, сможет ли она заслужить доверие окружающих.

5. ДОПОЛНЕНИЯ К ВАЖНЫМ МОМЕНТАМ

В этом разделе перечислены важные моменты факторного анализа, описанные на стр. 139-149, а также даны дополнения к некоторым из них.

Важный момент №1.

Смысл каждого из общих факторов интерпретируется аналитиком субъективно и «задним умом», то есть после того, как он по мере своих возможностей проведёт анализ.

Важный момент №2.

В отличие от главных компонент в методе главных компонент, все общие факторы считаются равноправными.

Важный момент №3.

Расчёты в факторном анализе будут невозможны, если аналитик предварительно не задаст число общих факторов.

Как это ни странно звучит, но для расчётов факторного анализа нужно, чтобы аналитик задал число общих факторов «до» начала анализа.

Существуют математические критерии для задания числа общих факторов. Один из них, описанный на стр. 141 — это число собственных значений матрицы корреляции, которые по величине не меньше 1. Кроме того, есть, например, критерий числа собственных значений непосредственно перед пологой частью *графика каменистой осыпи* (его ещё называют *графиком Кеттела*), на котором собственные значения отображены в порядке убывания.

Вышеизложенное обычно описывается в распространённой литературе, посвящённой факторному анализу. Однако я не уверен, что следует погружаться в какие-либо математические размышления, например о «числе собственных значений», всего лишь ради «критерия». Мне кажется, что более практично будет, не мудрствуя лукаво, просто проанализировать несколько вариантов для разного заданного числа общих факторов, как говорилось на стр. 141.

Важный момент №4.

Бывает так, что оказываются приемлемыми результаты сразу нескольких расчётов для разного заданного числа общих факторов, несмотря на то, что анализируются одни и те же данные. В подобных случаях за «окончательный результат» принимается тот, который больше понравится аналитику.

Важный момент №4 имеет гораздо большее значение, чем предыдущий.

Наверное, некоторые из читателей сомневаются, можно ли считать окончательным результат только на том основании, что он понравился аналитику.

В действительности, существует подход, согласно которому «окончательным» можно считать результат, удовлетворяющий одному из следующих условий:

- заданное число общих факторов равно числу собственных значений матрицы корреляции, величина которых не меньше 1;
- заданное число общих факторов равно числу собственных значений непосредственно перед пологой частью «графика каменистой осыпи»;
- суммарный вклад заданного числа общих факторов превышает определённое значение¹;
- заданное число общих факторов таково, что результат теста согласованности² не достигает уровня значимости;
- заданное число общих факторов таково, что индекс согласованности (Goodness of Fit Index, GFI)³ принимает наилучшее значение.

Однако опыт, по крайней мере мой личный, показывает, что этим условиям нельзя слишком доверять. Другими словами, нельзя рассчитывать на то, что нижеописанная ситуация случится в действительности:

Я провёл расчёты для разного числа общих факторов: 2, 3, 4 и т.д. Наилучшее впечатление на меня произвёл результат для трёх общих факторов. Проверил число собственных значений матрицы корреляции, величина которых не меньше 1 — оно оказалось равно 3. Да, что не говори, основательный подход приводит к заслуживающим доверия результатам.

Конечно, считать «окончательным» тот результат, который больше понравился аналитику — это, наверное, преувеличение автора. Однако, как было отмечено, даже основательный подход не всегда приводит к заслуживающим доверия результатам — решение в конечном счёте принимается аналитиком.

Важный момент №5.

Некоторые методы расчёта применимы только при условии, если заданное число общих факторов не превышает числа откликов, даже если в действительности за данными скрываются гораздо больше общих факторов.

По крайней мере, если в факторном анализе предполагается осуществлять вращение методами варимакс и промакс, то существует верхний предел числа общих факторов. Более подробно об этом я расскажу ниже.

¹ Автор принимает это определённое значение равным 50%.

² Про тест согласованности будет рассказано позже.

³ «Индекс согласованности» в этой книге не объясняется.

Важный момент №6.

См. схему на стр. 143 (в целях экономии бумаги, здесь не приводится).

Важный момент №7.

В факторном анализе перед расчётами обычно производится нормировка по каждой из переменных.

Важный момент №8.

См. формулу и схему на стр. 145 (в целях экономии бумаги, здесь не приводятся).

Как было сказано в начале главы 4, метод главных компонент и факторный анализ — это разные методы. Однако есть много людей, считающих, что это одно и то же. Может быть, в этом виноваты некоторые компьютерные программы, не делающие различия между этими двумя методами. Ещё раз посмотрите на схему стр. 138, и разница между методом главных компонент и факторным анализом станет для вас очевидна.

Важный момент №9.

Факторный анализ не является волшебным методом, автоматически выявляющим общие факторы, о существовании которых не подозревал даже аналитик.

Как было сказано на стр. 148, для успешного факторного анализа необходимо сначала в некоторой мере построить гипотезу: «За этими откликами, возможно, скрываются вот такие общие факторы». Другими словами, если основательно построить гипотезу, то вероятность успеха, то есть вероятность получения желаемого результата, повышается. Да, факторный анализ на самом деле напоминает так называемый «договорной матч»¹.

Для факторного анализа необходима тщательная проработка вопросов анкеты. Но даже это не даёт гарантию успеха — есть вероятность получения таких результатов, что аналитик сможет только покачать головой. Успех факторного анализа абсолютно невозможен, если приступать к нему, думая так: «Вот у меня под рукой есть данные недавнего анкетирования, попробую-ка я их проанализировать». Повторяю, абсолютно невозможен.

¹ Конечно, назвать это «подтасовкой» у меня язык не поворачивается, но как бы там ни было, реальность факторного анализа состоит в том, что аналитик направляет ход вычислений так, чтобы получить желаемый результат.

Важный момент №10.

Факторный анализ является методом, предназначенным для проверки значений факторных нагрузок.

Обычно факторный анализ определяют как:

- метод для описания корреляции между откликами с помощью нескольких общих факторов;
- метод для обнаружения скрытых общих факторов.

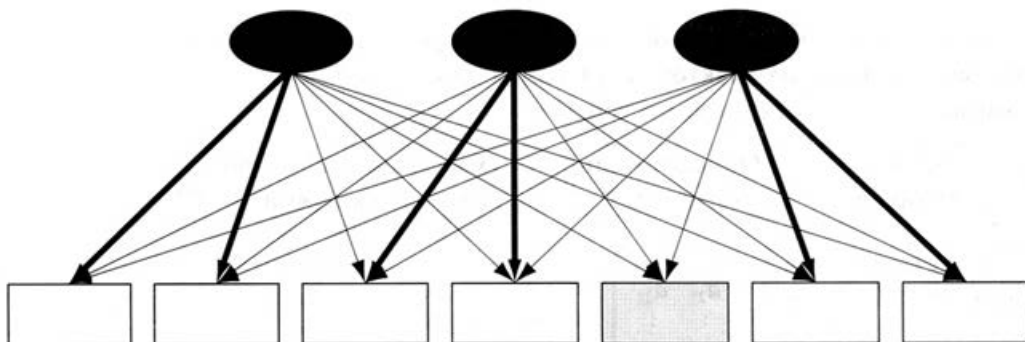
Я не могу согласиться ни с одним из этих определений. И вот почему. Первое из определений, действительно, математически приемлемо, но слишком сложно, чтобы быть понятным при первом взгляде на него, если только человек не обладает «сверхпроницательностью». Что касается второго определения, то как было отмечено в важном моменте №9, факторный анализ является «договорным матчем», поэтому это определение очевидно ошибочно. Однако оно является интуитивно понятным и хорошо подходит в качестве «первого шага» для начинающих изучать факторный анализ.

Сам я определяю факторный анализ как «метод для проверки значений факторных нагрузок». Некоторые из читателей, возможно, покачают головой: «А не является ли это определением *подтверждающего факторного анализа*¹?» Но поймите меня правильно — я ведь не говорю о «тщательной» проверке значений факторных нагрузок.

¹ На самом деле факторный анализ делится на *исследовательский факторный анализ* и *подтверждающий факторный анализ* (проверочный факторный анализ). Гермином «факторный анализ» обычно называют первый из них, и эта книга посвящена только ему.

6. ПЕРЕМЕННЫЕ С МАЛЫМИ ЗНАЧЕНИЯМИ ФАКТОРНЫХ НАГРУЗОК

В факторном анализе часто встречаются отклики, почти не испытывающие влияния ни одного из общих факторов, как показано на рисунке ниже.



■ Рис.5.1. Случай наличия откликов, почти не испытывающих влияния ни одного из общих факторов (факторные нагрузки с абсолютным значением выше 0,5 обозначены жирными стрелками)

В подобных случаях я рекомендую принять одну из следующих мер:

- исключить эти отклики и повторить факторный анализ;
- добиться отсутствия откликов, почти не испытывающих влияния ни одного из общих факторов, путём пошагового снижения минимального абсолютного значения факторной нагрузки, например, вот так: 0,5 -> 0,45->0,4 ->

Минимальное абсолютное значение факторной нагрузки, при котором влияние общего фактора считается «сильным», с точки зрения статистики не имеет под собой особых оснований, но обычно выбирается из диапазона от 0,3 до 0,5.

7. МЕТОД МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

7.1. ОБЩИЕ СВЕДЕНИЯ О МЕТОДЕ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Одним из известных методов расчёта факторных нагрузок, помимо описанного ранее метода главных факторов, является *метод максимального правдоподобия*¹. В этом методе для функции L , имеющей вид

$$L = \left[\begin{array}{c} \text{Число} \\ \text{откликов} \end{array} \right] + \ln \left(\left[\begin{array}{c} \text{Определитель} \\ \text{матрицы } X \end{array} \right] \right) - \left[\begin{array}{c} \text{Сумма значений главной} \\ \text{диагонали матрицы } X \end{array} \right],$$

$$\text{где } X = \begin{pmatrix} r_{11} & r_{12} & \dots & r_{16} \\ r_{21} & r_{22} & \dots & r_{26} \\ \vdots & \vdots & \vdots & \vdots \\ r_{61} & r_{62} & \dots & r_{66} \end{pmatrix} \left[\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} & \dots & a_{61} \\ a_{12} & a_{22} & \dots & a_{62} \end{pmatrix} - \begin{pmatrix} d_1^2 & 0 & \dots & 0 \\ 0 & d_2^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & d_6^2 \end{pmatrix} \right]^{-1}$$

находят такие значения матрицы $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{pmatrix}$, при которых значение L будет максимально².

Похоже, в последнее время гуляет слух, что «метод максимального правдоподобия лучше метода главных факторов». Меня беспокоит наметившаяся в последнее время тенденция безоглядного использования метода максимального правдоподобия. Предпосылкой использования этого метода является соответствие данных в генеральной совокупности *многомерному нормальному распределению*³. Помните об этом.

7.2. ТЕСТ СОГЛАСОВАННОСТИ

Одной из причин слухов, что «метод максимального правдоподобия лучше метода главных факторов», является возможность проведения «теста согласованности»⁴. Тест согласованности, если описывать его в общих чертах, заключается в проверке следующих гипотез.

¹Кстати, есть ещё такие методы, как *метод наименьших квадратов*, *обобщённый метод наименьших квадратов*.

²Определители матриц в этой книге не объясняются.

³Функцию плотности вероятности многомерного нормального распределения я здесь не привожу, так как выражение громоздко само по себе и содержит много обозначений, которые пришлось бы объяснять отдельно. Просто считайте, что это многомерный случай нормального распределения, как и следует из названия.

⁴Другая причина популярности метода максимального правдоподобия — возможность найти значение «индекса согласованности», который в этой книге не рассматривается.

Нулевая гипотеза	Число общих факторов равно m .
Альтернативная гипотеза	Число общих факторов не равно m .

В тесте согласованности, если значение p меньше уровня значимости α , то считается верной альтернативная гипотеза: «число общих факторов не равно m ». Если же значение p больше уровня значимости α , то делается вывод, что нулевую гипотезу нельзя признать ошибочной, другими словами: «Число общих факторов, может быть, равно m , а может быть, не равно m . Давайте истолкуем это широко и предположим, что оно находится где-то в районе m .»

Привлекательность теста согласованности заключается в возможности приближённо оценить число общих факторов. Однако это не означает автоматического определения их оптимального числа, то есть этот метод вовсе не такой замечательный, как кажется на первый взгляд.

8. ПОЧЕМУ ВРАЩЕНИЕ ПОДРАЗУМЕВАЕТ МЕТОД ВАРИМАКС

Как отмечено на стр. 174, существуют различные виды вращения. Однако в действительности в подавляющем большинстве статей или докладов почему-то используется только метод варимакс.

В чём же причина? Я не знаю точного ответа, так как не изучал «историю факторного анализа», но однажды мне довелось услышать слова одного статистика: «Если для вращения используешь метод варимакс, то все будут довольны, а если другой метод — то подвергнешься критике.» В связи с этим, я могу выдвинуть следующую гипотезу:

1. Давным давно один исследователь (*не статистик) опубликовал результаты своих исследований, основанных на методе варимакс.
2. Другой исследователь, узнавший об этих результатах, подумал: «Да, этот факторный анализ выглядит интересно», и тоже опубликовал результаты исследований на основе метода варимакс.
3. Изложенное в пунктах 1 и 2 повторялось много раз, и накопилось огромное количество результатов исследований. В конце концов укоренился безосновательный шаблон: «Вращение в факторном анализе — это метод варимакс!»
4. Вряд ли в прошлом совсем не было исследователей, которые задавались вопросом: «Почему и в тех, и в этих исследованиях используется метод варимакс?» Однако они, чтобы не создавать «лишних проблем», закрывали на это глаза и в конечном итоге последовали утверждению: «Вращение в факторном анализе — это метод варимакс!»

5. Изложенное в пунктах 1 и 2 продолжало повторяться, и до настоящего времени накопился колоссальный объём вышеописанных исследований.

Прошу понять меня правильно: я вовсе не утверждаю, что «метод варимакс — плохой метод». Я просто задался вопросом, почему когда говорят «вращение», подразумевают только метод варимакс.

По моей оценке, та же самая ситуация в ближайшем будущем случится и с методом промакс, о котором я расскажу позже. И так в действительности уже происходит.

9. МАТРИЦА ФАКТОРНЫХ НАГРУЗОК И МАТРИЦА ФАКТОРНОЙ СТРУКТУРЫ

«Содержание данного параграфа является абстрактным, поэтому не очень разбирающиеся в математике читатели могут его пропустить», — я хотел бы сказать так и здесь, однако без усвоения содержания этого параграфа вы не сможете понять последующие, поэтому сделайте над собой усилие, пожалуйста.

Как было сказано на стр. 172, матрицу
$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \dots & a_{pm} \end{pmatrix}$$
 называют *матрицей*

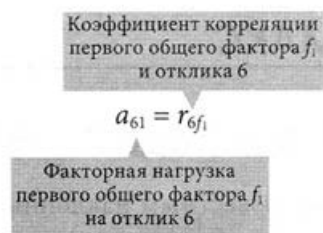
Факторная нагрузка
второго общего фактора f_2
на отклик p

факторных нагрузок ИЛИ матрицей факторного отображения.

Матрицу
$$\begin{pmatrix} r_{1f_1} & r_{1f_2} & \dots & r_{1f_m} \\ r_{2f_1} & r_{2f_2} & \dots & r_{2f_m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{pf_1} & r_{pf_2} & \dots & r_{pf_m} \end{pmatrix}$$
 называют *матрицей факторной структуры*.

Коэффициент корреляции
второго общего фактора f_2
и отклика p .

В случае модели ортогональных факторов, матрицы факторных нагрузок и факторной структуры совпадают друг с другом. Другими словами, выполняются такие равенства:



В случае модели неортогональных факторов подобные равенства не выполняются.

Нужно также обратить внимание на следующее. В таблице ниже приведена часть данных со стр. 152 и факторные значения со стр. 188.

♦ Таблица 5.1. Часть данных со стр. 152 и факторные значения со стр. 188.

Респонденты	Q1a Внешний вид здания	Первый общий фактор f_1
А.	5	1,38
Б.	5	1,27
В.	4	0,28
Г.	2	0,34
Д.	3	-0,45
Е.	5	1,20
Ж.	5	1,22
З.	3	-1,60
И.	4	-0,68
К.	1	-1,61
Л.	3	-0,62
М.	4	0,24
Н.	3	-0,64
О.	4	0,24
П.	2	-0,56

Если найти коэффициент корреляции для этих значений, то он окажется равным 0,79, то есть не будет совпадать с элементом $B_{11}=0,78$ в матрице факторных нагрузок со стр. 179. Причина заключается в том, что факторные значения, приведённые в таблице выше, являются не «точными значениями», а «предполагаемыми значениями»¹.

¹ Как отмечено на стр. 186, факторные значения можно вычислить либо регрессионным методом, либо методом Бартлетта, либо методом Андерсена-Рубина. Такое разнообразие методов означает, другими словами, отсутствие «единственного правильного» метода нахождения точных факторных значений.

10. МЕТОД ПРОМАКС

10.1. ОБЩИЕ СВЕДЕНИЯ О МЕТОДЕ ПРОМАКС

Как отмечалось ранее, самым известным наклонным вращением является *метод промакс*. Порядок его применения в общих чертах выглядит следующим образом:

1. Проводят вращение методом варимакс.
2. Предполагают вид «истинной матрицы факторных нагрузок», которая, «исходя из предшествующего опыта, непременно должна отражать положение дел в генеральной совокупности». Эту матрицу принято называть *целевой матрицей*.
3. Вращают координатные оси пункта 1 так, чтобы как можно больше приблизиться к предполагаемой матрице пункта 2.

Теперь о главном. Мы изо дня в день вынуждены испытывать трудности именно потому, что не знаем положения дел в генеральной совокупности. То есть о возможности «предположить вид целевой матрицы» не может быть и речи. Поэтому вращение методом промакс требует рассматривать матрицу C , расчёт которой для примера из данной главы приведён ниже, как целевую матрицу:

$$C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \\ c_{41} & c_{42} \\ c_{51} & c_{52} \\ c_{61} & c_{62} \end{pmatrix} = \begin{pmatrix} 0,98599 & 0,00005 \\ 0,99538 & 0,00001 \\ 0,99966 & 0,00000 \\ 0,00000 & 0,99966 \\ 0,00001 & 0,99270 \\ 0,00003 & 0,98987 \end{pmatrix}$$

$$\frac{\sqrt{b_{21}^2 + b_{22}^2}}{b_{22}} \times \left| \frac{b_{22}}{\sqrt{b_{21}^2 + b_{22}^2}} \right|^{k+1} = \frac{\sqrt{0,89^2 + 0,04^2}}{0,04} \times \left| \frac{0,04}{\sqrt{0,89^2 + 0,04^2}} \right|^{k+1} = 0,00001$$

Здесь fr_2^i , 2 и т.д. представляют собой значения нижеприведённой матрицы B , которая является матрицей факторных нагрузок после вращении методом варимакс.

$$B = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \\ b_{41} & b_{42} \\ b_{51} & b_{52} \\ b_{61} & b_{62} \end{pmatrix} = \begin{pmatrix} 0,78 & 0,07 \\ 0,89 & 0,04 \\ 0,99 & 0,01 \\ 0,01 & 0,94 \\ 0,05 & 0,86 \\ 0,06 & 0,79 \end{pmatrix}$$

κ (греч. буква «каппа») — это значение, которое аналитик должен задать сам. Это значение, которое обычно задают равным 2, 3 или 4, в этом примере выбрано равным 4.

Может быть, вышеизложенное вызовет у некоторых читателей ощущение какой-то недосказанности. Они могут почувствовать следующие сомнения:

- На чём основано утверждение, что «целевая матрица равна матрице C »?
- Допустимо ли, чтобы аналитик задавал значение κ на свой субъективный взгляд?
- Почему нужно проводить вращение методом варимакс, то есть ортогональное вращение, в начале применения метода промакс, который является наклонным вращением?

Я вполне понимаю чувства этих читателей, но хочу, чтобы они перестроили своё мышление. Ведь всё это: и утверждение «целевая матрица равна матрице C », и выбор значения κ на усмотрение аналитика, и вращение методом варимакс в самом начале, — и является методом промакс.

10.2. МАТРИЦА ФАКТОРНЫХ НАГРУЗОК, МАТРИЦА ФАКТОРНЫХ КОРРЕЛЯЦИЙ, МАТРИЦА ФАКТОРНОЙ СТРУКТУРЫ

Для примера этой главы матрица факторных нагрузок, матрица факторных корреляций и матрица факторной структуры находятся путём нижеследующих вычислений. Матрица факторных корреляций — это матрица, состоящая из коэффициентов корреляции общих факторов друг с другом.

В нижеприведённом описании встречается много матриц, поэтому для удобства я обозначил матрицы как P , Q и т.д.

■ МАТРИЦА ФАКТОРНЫХ НАГРУЗОК P

$$P = \begin{pmatrix} 0,78 & 0,07 \\ \vdots & \vdots \\ 0,06 & 0,79 \end{pmatrix} \begin{pmatrix} 1,11 & -0,05 \\ -0,05 & 1,15 \end{pmatrix} \begin{pmatrix} 0,90 & 0 \\ 0 & 0,88 \end{pmatrix} = \begin{pmatrix} 0,78 & 0,03 \\ 0,89 & 0,01 \\ 0,99 & -0,03 \\ -0,03 & 0,94 \\ 0,02 & 0,86 \\ 0,02 & 0,79 \end{pmatrix}$$

Матрица B с предыдущей страницы

Q

D

$$Q = \begin{bmatrix} (0,78 \dots 0,06) \\ (0,07 \dots 0,79) \end{bmatrix} \begin{pmatrix} 0,78 & 0,07 \\ \vdots & \vdots \\ 0,06 & 0,79 \end{pmatrix}^{-1} \begin{pmatrix} 0,78 & \dots & 0,06 \\ 0,07 & \dots & 0,79 \end{pmatrix} \begin{pmatrix} 0,98599 & 0,00005 \\ \vdots & \vdots \\ 0,00003 & 0,98987 \end{pmatrix} = \begin{pmatrix} 1,11 & -0,05 \\ -0,05 & 1,15 \end{pmatrix}$$

Матрица, транспонированная к B

B

Матрица, транспонированная к B

Матрица C со страницы 202

$$d = \begin{bmatrix} (1,11 \dots -0,05) \\ (-0,05 \dots 1,15) \end{bmatrix} \begin{pmatrix} 1,11 & -0,05 \\ -0,05 & 1,15 \end{pmatrix}^{-1} = \begin{pmatrix} 0,81 & 0,07 \\ 0,07 & 0,77 \end{pmatrix}$$

Матрица, транспонированная к Q

Q

$$D = \begin{pmatrix} \sqrt{0,81} & 0 \\ 0 & \sqrt{0,77} \end{pmatrix} = \begin{pmatrix} 0,90 & 0 \\ 0 & 0,88 \end{pmatrix}$$

Матрица, значения на главной диагонали которой равны корням значений из d , а остальные — нулю.

■ МАТРИЦА ФАКТОРНЫХ КОРРЕЛЯЦИЙ L

$$L = \begin{pmatrix} 0,9991 & 0,0414 \\ 0,0421 & 0,9991 \end{pmatrix} \begin{pmatrix} 0,9991 & 0,0421 \\ 0,0414 & 0,9991 \end{pmatrix} = \begin{pmatrix} 1 & 0,08 \\ 0,08 & 1 \end{pmatrix}$$

T

Матрица, транспонированная к T

$$T = \left[\begin{pmatrix} 1,11 & -0,05 \\ -0,05 & 1,15 \end{pmatrix} \begin{pmatrix} 0,90 & 0 \\ 0 & 0,88 \end{pmatrix} \right]^{-1} = \begin{pmatrix} 0,9991 & 0,0414 \\ 0,0421 & 0,9991 \end{pmatrix}$$

Q

D

■ МАТРИЦА ФАКТОРНОЙ СТРУКТУРЫ S

$$S = \begin{pmatrix} 0,78 & 0,03 \\ 0,89 & 0,01 \\ 0,99 & -0,03 \\ -0,03 & 0,94 \\ 0,02 & 0,86 \\ 0,02 & 0,79 \end{pmatrix} \begin{pmatrix} 1 & 0,08 \\ 0,08 & 1 \end{pmatrix} = \begin{pmatrix} 0,78 & 0,10 \\ 0,89 & 0,08 \\ 0,99 & 0,05 \\ 0,05 & 0,94 \\ 0,09 & 0,86 \\ 0,09 & 0,79 \end{pmatrix}$$

P

L

10.3. ТОЧНОСТЬ РЕЗУЛЬТАТОВ АНАЛИЗА

Точность результатов анализа после вращения методом промакс нельзя проверить с помощью величин удельных вкладов или суммарного вклада. Вместо этого иногда используют проверку, основанную на таких критериях, как *вклад после исключения влияния других общих факторов*, а также *вклад с игнорированием других общих факторов*, а иногда вообще не проверяют точность.

Идея «вклада после исключения влияния других общих факторов» трудна для понимания, поэтому здесь я объясню только «вклад с игнорированием других общих факторов»¹.

«Вклад z-го общего фактора с игнорированием других общих факторов» вычисляется путём сложения возведённых в квадрат значений по столбцам матрицы факторной структуры. Для примера данной главы это выглядит так:

	Вклад с игнорированием других общих факторов
1-й общий фактор	$r_{1f_1}^2 + r_{2f_1}^2 + r_{3f_1}^2 + r_{4f_1}^2 + r_{5f_1}^2 + r_{6f_1}^2 =$ $= 0,78^2 + 0,89^2 + 0,99^2 + 0,05^2 + 0,09^2 + 0,09^2 =$ $= 2,40$
2-й общий фактор	$r_{1f_2}^2 + r_{2f_2}^2 + r_{3f_2}^2 + r_{4f_2}^2 + r_{5f_2}^2 + r_{6f_2}^2 =$ $= 0,10^2 + 0,08^2 + 0,05^2 + 0,94^2 + 0,86^2 + 0,79^2 =$ $= 2,28$

Чем больше значение, тем сильнее соответствующий общий фактор влияет на множество откликов. В отличие от метода варимакс, «вклад с игнорированием других общих факторов» не выражается в процентах.

«Вклад с игнорированием влияния других общих факторов» является не абсолютным, а относительным критерием. Другими словами, к нему надо относиться гибко, размышляя примерно так: «У этого общего фактора значение больше, чем у того, а вон у того больше, чем у этого». Поэтому не исключено, что этот критерий может разочаровать аналитика, потратившего силы на его нахождение.

¹ Объяснение я не привожу только ввиду трудности для понимания, а не потому, что считаю вклад с исключением влияния других общих факторов плохим критерием.

10.4. ФАКТОРНЫЕ ЗНАЧЕНИЯ

Для примера данной главы, нахождение факторных значений после вращения методом промакс проводится с использованием регрессионного метода путём нижеследующих вычислений.

Данные, нормированные по каждой переменной (см. стр. 155)

Матрица, обратная к матрице корреляции (см. стр. 166)

Матрица факторной структуры (см. стр. 205)

$$\begin{pmatrix} f_{A1} & f_{A2} \\ f_{B1} & f_{B2} \\ f_{B1} & f_{B2} \\ f_{T1} & f_{T2} \\ f_{Л1} & f_{Л2} \\ f_{E1} & f_{E2} \\ f_{Ж1} & f_{Ж2} \\ f_{З1} & f_{З2} \\ f_{И1} & f_{И2} \\ f_{К1} & f_{К2} \\ f_{Л1} & f_{Л2} \\ f_{M1} & f_{M2} \\ f_{И1} & f_{И2} \\ f_{O1} & f_{O2} \\ f_{И1} & f_{И2} \end{pmatrix} = \begin{pmatrix} 1,2 & 1,6 & 1,3 & 0,4 & 0,5 & -0,9 \\ 1,2 & 0,8 & 1,3 & -1,2 & -1,0 & -0,9 \\ 0,4 & 0,8 & 0,3 & 0,4 & 0,5 & 0,6 \\ -1,2 & 0,1 & 0,3 & -0,4 & -0,3 & -0,1 \\ -0,4 & 0,1 & -0,6 & -0,4 & 0,5 & -1,6 \\ 1,2 & 0,8 & 1,3 & -0,4 & -1,0 & -0,1 \\ 1,2 & 1,6 & 1,3 & 0,4 & 1,3 & 1,3 \\ -0,4 & -1,5 & -1,6 & 1,3 & 0,5 & 0,6 \\ 0,4 & -1,5 & -0,6 & -0,4 & -1,0 & -0,1 \\ -2,0 & -0,7 & -1,6 & -1,2 & -1,0 & -0,9 \\ -0,4 & -0,7 & -0,6 & -2,1 & -1,8 & -1,6 \\ 0,4 & 0,1 & 0,3 & 0,4 & -0,3 & 0,6 \\ -0,4 & -0,7 & -0,6 & 0,4 & 1,3 & 1,3 \\ 0,4 & 0,1 & 0,3 & 1,3 & 0,5 & 1,3 \\ -1,2 & -0,7 & -0,6 & 1,3 & 1,3 & 0,6 \end{pmatrix} \begin{pmatrix} 1 & 0,65 & 0,80 & 0,11 & 0,01 & 0,14 \\ 0,65 & 1 & 0,89 & 0,02 & 0,19 & 0,01 \\ 0,80 & 0,89 & 1 & 0,02 & 0,04 & 0,10 \\ 0,11 & 0,02 & 0,02 & 1 & 0,82 & 0,77 \\ 0,01 & 0,19 & 0,04 & 0,82 & 1 & 0,64 \\ 0,14 & 0,01 & 0,10 & 0,77 & 0,64 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0,78 & 0,10 \\ 0,89 & 0,08 \\ 0,99 & 0,05 \\ 0,05 & 0,94 \\ 0,09 & 0,86 \\ 0,09 & 0,79 \end{pmatrix}$$

$$= \begin{pmatrix} 1,39 & 0,29 \\ 1,22 & -1,14 \\ 0,30 & 0,50 \\ 0,33 & -0,29 \\ -0,46 & -0,36 \\ 1,17 & -0,51 \\ 1,25 & 0,83 \\ -1,56 & 0,96 \\ -0,70 & -0,56 \\ -1,65 & -1,17 \\ -0,71 & -2,00 \\ 0,25 & 0,30 \\ -0,61 & 0,80 \\ 0,28 & 1,13 \\ 0,51 & 1,22 \end{pmatrix}$$

11. ВЕРХНИЙ ПРЕДЕЛ ЗАДАВАЕМОГО ЧИСЛА ОБЩИХ ФАКТОРОВ

Запомните, что при факторном анализе, по крайней мере, с вращением по методам варимакс или промакс, существует верхний предел задаваемого числа общих факторов, который конкретно выглядит так:

$$\boxed{\text{Число общих факторов}} \leq \frac{2 \times \text{Число откликов} + 1 - \sqrt{8 \times \text{Число откликов} + 1}}{2},$$

На основе вышеприведённого неравенства я рассчитал верхние пределы задаваемого числа общих факторов, приведённые в таблице ниже. Используйте эти значения для справки.

◆ Таблица 5.2. Верхние пределы задаваемого числа общих факторов

Число откликов	→	Верхний предел задаваемого числа общих факторов	Число откликов	→	Верхний предел задаваемого числа общих факторов
1	→	0	21	→	15
2	→	0	22	→	15
3	→	1	23	→	16
4	→	1	24	→	17
5	→	2	25	→	18
6	→	3	26	→	19
7	→	3	27	→	20
8	→	4	28	→	21
9	→	5	29	→	21
10	→	6	30	→	22
11	→	6	31	→	23
12	→	7	32	→	24
13	→	8	33	→	25
14	→	9	34	→	26
15	→	10	35	→	27
16	→	10	36	→	28
17	→	11	37	→	28
18	→	12	38	→	29
19	→	13	39	→	30
20	→	14	40	→	31

12. СТОИТ ЛИ СЧИТАТЬ МЕТОД ГЛАВНЫХ ФАКТОРОВ И МЕТОД ВАРИМАКС УСТАРЕВШИМИ?

Данный параграф необходимо прочитать даже тем, кто впервые узнал о факторном анализе из этой книги, хотя некоторые места здесь будут для них не очень понятны.

В примере данной главы в качестве метода расчёта факторных нагрузок был рассмотрен метод главных факторов, а в качестве метода вращения — метод варимакс. В действительности, и к методу главных факторов, и к методу главных компонент начинают относиться как к «устаревшим» и «плохим».

По нижеизложенным причинам, я не уверен в правильности подобного отношения.

- Метод главных факторов по большому счёту представляет собой спектральный анализ¹, то есть метод вычислений, который не может быть «новым или старым», «хорошим или плохим».
- Метод главных факторов, в отличие от метода максимального правдоподобия, не налагает такого большого ограничения, как «подчинение генеральной совокупности многомерному нормальному распределению».
- Метод главных факторов на первый взгляд кажется сложным, но в действительности вычисления в нём даже проще, чем в методе максимального правдоподобия².
- Метод промакс, рассматриваемый как замена методу варимакс, тоже не идеален: вспомните довольно странный способ задания целевой матрицы, значения k .
- Хотя постулат метода варимакс о том, что «коэффициенты корреляции между любыми различными общими факторами равны 0» и не лишён натяжки, но представляется непоследовательным сначала признавать этот метод по причинам, далёким от поиска истины: «относительной лёгкости вычислений», «недостаточной мощности компьютеров», а потом клеймить его как «устаревший» и «плохой».
- Считать метод главных факторов и метод варимакс «устаревшими» и «плохими» равнозначно мнению: «Результаты исследований с использованием этих методов бесполезны и не заслуживают внимания. Давайте их игнорировать». Однако в подавляющем большинстве исследований, проведённых до настоящего времени³, использовалась комбинация «метод главных факторов + метод варимакс», поэтому окажется, что практически не существует материалов, на которые можно опереться⁴.

¹ Описанное на стр. 76-78 можно, в очень широком смысле, рассматривать как спектральный анализ.

² Это личное мнение автора.

³ Эти строки были написаны в 2006 году.

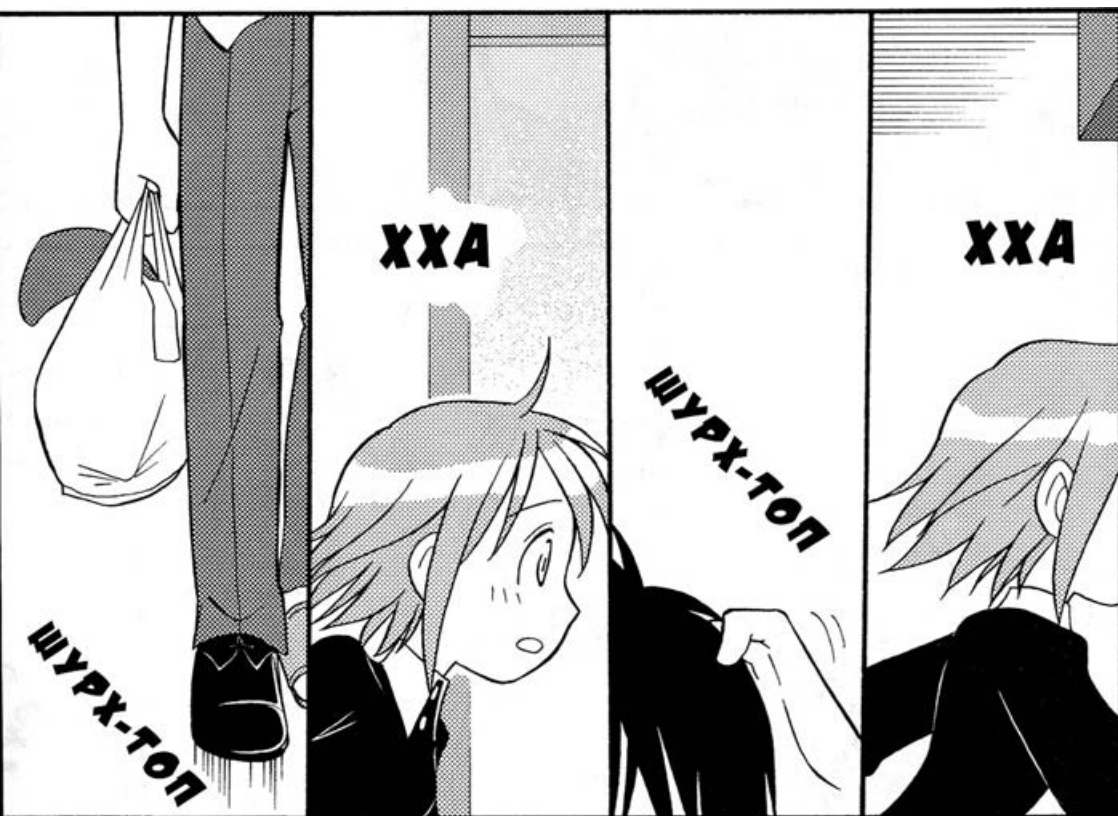
⁴ Мне кажется, что если заявлять коллеге: «Ты что, делаешь факторный анализ с помощью метода главных факторов и метода варимакс?! Ты устарел!», — то в конечном итоге сам окажешься в дураках.

Что касается способа расчёта факторных нагрузок, я считаю, что метод главных факторов и метод максимального правдоподобия надо сравнивать не как «устаревший и современный» или «плохой и хороший», а просто как разные обычаи разных «школ». Что касается метода вращения, действительно, постулат метода варимакс: «значения коэффициентов корреляции между любыми различными общими факторами равны 0» не лишён натяжки, но метод промакс тоже вызывает непреодолимые сомнения. Мне даже кажется, что метод варимакс, по сравнению с методом промакс, не так уж и плох.

13. ТЕРМИНОЛОГИЯ ФАКТОРНОГО АНАЛИЗА

В данной книге я постоянно называл переменные отклика факторного анализа словом «отклики», однако их часто называют также *наблюдаемыми переменными*.

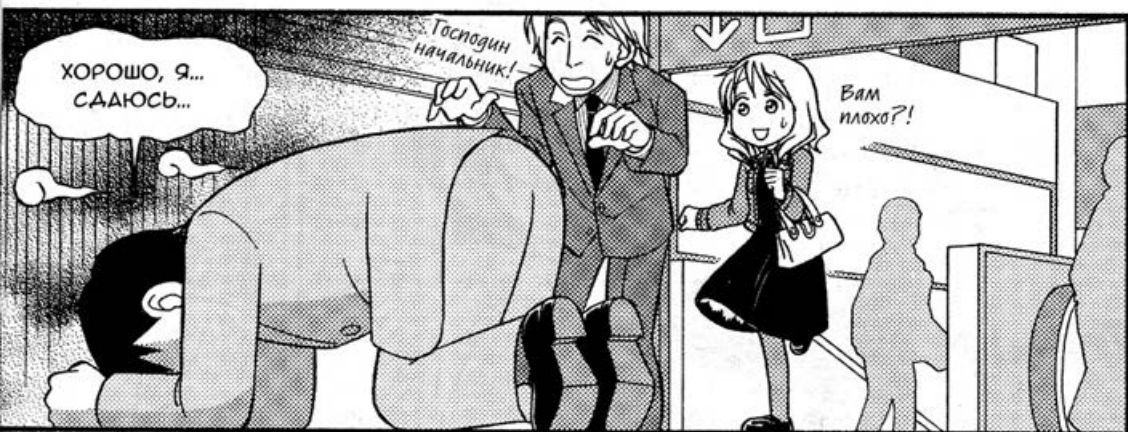
Общие факторы иногда называют также *скрытыми переменными (ненаблюдаемыми переменными)*. Факторные нагрузки иногда называют также *путевыми коэффициентами*.











ХОРОШО, Я... СДАНЬСЯ...

Господин начальник!

Вам плохо?!



ЗНАЕШЬ, ЯМАМОТО-САН.

АА, Я СЛУШАЮ.



Я ХОТЕЛА ПОЕХАТЬ ЗА ТОБОЙ И ТАМ ПОМОГАТЬ ТЕБЕ,...

...И ДЛЯ ЭТОГО КАЖДЫЙ ДЕНЬ ИЗУЧАЛА ТАКИЕ ШТУКИ, КАК ФАКТОРНЫЙ АНАЛИЗ.

ЧТО? ФАКТОРНЫЙ АНАЛИЗ?

Здорово!



ОДНАКО ТЫ ВЕРНУЛСЯ РАНЬШЕ.



НО... Я ВЕДЬ ТОЖЕ... ЭТО,...



...ХОТЕЛ ПОСКОРЕЕ УВИДЕТЬ ТЕБЯ,...



...ПОЭТОМУ СТАРАЛСЯ В РАБОТЕ.



С ВОЗВРАЩЕНИЕМ
ТЕБЯ,
ЯМАМОТО-САН!

ПРЫГ



РУЦ,
Я РАДА, ЧТО
У ВАС ВСЁ
ХОРОШО!

Я ещё
никогда
не видел
и него
такого
лицца!



ПОСЛУШАЙ,
ЯМАМОТО!
ПОПРОБУЙ
ТОЛЬКО
ОБИДЕТЬ
МОЮ ДОЧЬ!

НЕТ... НЕТ,
Я НЕ БУДУ ЕЁ
ОБИЖАТЬ.

Пана.



ГОСПОДА!
Я ПРЕДЛАГАЮ
УСТРОИТЬ В КАФЕ
NORNS ВЕЧЕРНКУ
ПО СЛУЧАЮ
ВОЗВРАЩЕНИЯ!

ХОРОШАЯ
ИДЕЯ!



ПОЙДАЁМТЕ!
И ВЫ ВЕДЬ
ТОЖЕ
С НАМИ?

НЕТ!
Я НЕ ПОЙДУ!

НИКУДА ВЫ
НЕ ДЕНЕТЕСЬ!
МЫ ВАС СИЛОЙ
ОТВЕДЕМ!



Понял, понял!
Только не тол-
китесь!

Ха-ха-ха

ПОЙДЕМ,
ЯМАМОТО-
САН!

ХОРОШО.

НАЧИНАЯ
С СЕГОДНЯШНЕГО
ДНЯ...

...МЫ БУДЕМ
УЧИТЬСЯ
ВМЕСТЕ!

Приложение

РАЗЛИЧНЫЕ МЕТОДЫ АНАЛИЗА

1. МНОГОМЕРНЫЙ АНАЛИЗ

1.1. ОБЩИЕ СВЕДЕНИЯ О МНОГОМЕРНОМ АНАЛИЗЕ

1.2. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

1.3. ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

1.4. КЛАСТЕРНЫЙ АНАЛИЗ

1.5. АНАЛИЗ СООТВЕТСТВИЙ И КВАНТИФИКАЦИЯ ХАЯШИ МЕТОДОМ III

1.6. МОДЕЛИРОВАНИЕ СТРУКТУРНЫМИ УРАВНЕНИЯМИ

2. ДРУГОЕ

2.1. ПРОВЕРКА СТАТИСТИЧЕСКОЙ ГИПОТЕЗЫ

2.2. МЕТОД КАПЛАНА-МЕЙЕРА

Здесь я опишу несколько известных методов анализа, включая те, о которых упоминалось в книгах «Манга о статистике» и «Манга о статистике. Регрессионный анализ». Я хочу, чтобы вы получили общее представление о следующем:

- Какие методы анализа существуют?
- Что представляют из себя эти методы анализа?
- Что позволяют узнать эти методы анализа?

Кроме того, в отличие от книг «Манга о статистике» и «Манга о статистике. Регрессионный анализ», порядок расчёта в Excel конкретных примеров данной книги я не привожу, так как рассчитывать их в Excel не так уж просто.

1. МНОГОМЕРНЫЙ АНАЛИЗ

1.1. ОБЩИЕ СВЕДЕНИЯ О МНОГОМЕРНОМ АНАЛИЗЕ

Как было сказано на стр. 11, *многомерный анализ* — это общее название для методов анализа данных, состоящих из множества переменных, как показано в таблице ниже.

	Переменная 1	Переменная 2	...	Переменная p
Респондент 1	34	1	...	171,7
Респондент 2	27	0	...	156,8
:	:	:	:	:
Респондент n	19	1	...	178,3



В этой главе я ознакомлю вас со следующими методами анализа из представленных на рисунке выше:

- множественная регрессия;
- логистическая регрессия;
- кластерный анализ;
- квантификация Хаяши методом III.

В дополнение я опишу также *анализ соответствий* и *моделирование структурными уравнениями*.

1.2. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Множественная регрессия — это прогноз значений на основе множества объясняющих переменных.

■ КОНКРЕТНЫЙ ПРИМЕР

В таблице ниже приведены данные сети кондитерских магазинов под названием «Казами-Вакегу».

Название магазина	Площадь магазина, цубо*	Удаление от ближайшей железнодорожной станции, м	Месячная выручка, десятки тысяч йен
Юмэно-ока	10	80	469
Тэраи	8	0	366
Сонэ	8	200	371
Хасимото	5	200	208
Кикё	7	300	246
Юбин	8	230	297
Суйдо	7	40	363
Рокузё	9	0	436
Вакаба	6	330	198
Мисато	9	180	364

* Японская традиционная мера площади равная примерно 3,3 м².

Предположив, что переменные связаны между собой следующим образом:



и проведя множественную регрессию, мы получим следующее уравнение:

$$y = 41,5x_1 - 0,3x_2 + 65,3$$

↑ ↑ ↑
 Месячная Площадь Удаление от ближайшей
 выручка магазина железнодорожной станции

Подставляя в это уравнение различные значения x_1 и x_2 , мы можем моделировать значения y .

ЗАМЕЧАНИЕ Читателям, интересующимся множественной регрессией, я рекомендую обратиться к моей книге «Занимательная статистика. Регрессионный анализ». В ней на основе приведённого здесь примера подробно описывается множественная регрессия.

1.3. ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

Логистическая регрессия — это прогноз вероятности на основе множества объясняющих переменных.

■ КОНКРЕТНЫЙ ПРИМЕР

В таблице ниже приведено состояние продаж «спецменю NORNS», приготавливаемого не более 1 раза в день в кафе NORNS, где подрабатывает Руи.

	Среда, суббота или воскресенье	Максимальная температура, °С	Число продаж спецменю NORNS
05.08 (пон.)	0	28	1
06.08 (втр.)	0	24	0
07.08 (срд.)	1	26	0
08.08 (чтв.)	0	24	0
09.08 (птн.)	0	23	0
10.08 (суб.)	1	28	1
11.08 (век.)	1	24	0
12.08 (пон.)	0	26	1
13.08 (втр.)	0	25	0
14.08 (срд.)	1	28	1
15.08 (чтв.)	0	21	0
16.08 (птн.)	0	22	0
17.08 (суб.)	1	27	1
18.08 (век.)	1	26	1
19.08 (пон.)	0	26	0
20.08 (втр.)	0	21	0
21.08 (срд.)	1	21	1
22.08 (чтв.)	0	27	0
23.08 (птн.)	0	23	0
24.08 (суб.)	1	22	0
25.08 (век.)	1	24	1

1.4. КЛАСТЕРНЫЙ АНАЛИЗ

Кластерный анализ — это классификация анализируемых объектов на несколько кластеров на основе расстояний между этими объектами.

Некоторые из читателей, может быть, подумают: «Если применить такие методы, как анализ главных компонент или факторный анализ, то экземпляры и переменные классифицируются сами, поэтому необязательно проводить ещё какой-то кластерный анализ». Это действительно так.

Но понимаете ли вы, что анализ главных компонент, факторный анализ дают лишь грубую классификацию, что-то вроде: «Взгляд на точечный график позволяет предположить, что эти два респондента сравнительно похожи». Кластерный анализ нужен для более точной классификации.

■ КОНКРЕТНЫЙ ПРИМЕР

В таблице ниже приведены результаты тестирования учеников третьего класса средней школы, занимающихся в одном из центров внешкольного образования.

Ученики	Родной язык	Обществоведение	Естествознание	Иностранный язык	Математика
A.	93	100	100	84	77
B.	100	98	98	95	86
C.	84	84	84	85	100
D.	70	73	73	66	77
E.	70	72	72	66	75
F.	66	68	68	57	82
G.	74	70	70	93	88
H.	74	75	75	70	79
I.	76	77	77	78	83
J.	79	88	88	86	100

Предположив, что число кластеров равно 2, и проведя кластерный анализ, получим следующую классификацию.

Кластер № 1	Кластер №2
Ученик А.	Ученик D.
Ученик В.	Ученик E.
Ученик С.	Ученик F.
Ученик G.	Ученик H.
Ученик J.	Ученик I.

ЗАМЕЧАНИЕ Содержание этого конкретного примера, наверное, заставит многих читателей покачать головой. В действительности, в кластерном анализе число кластеров не выясняется после проведения анализа, а должно быть задано аналитиком до проведения анализа, аналогично числу общих факторов в факторном анализе.

Кластерный анализ — это не что иное, как математический метод классификации объектов анализа на основе расстояний между ними. Он не поможет узнать, например, особенности людей, из которых состоит кластер №1. Особенности каждого кластера интерпретирует аналитик субъективно и «задним умом», то есть после того, как в меру своих возможностей проведёт кластерный анализ. Он рассуждает, например, так: «Похоже, что кластер № 1 — это люди, способные к учёбе, а кластер № 2 — неспособные».

В кластерном анализе существует множество методов вычислений, которые можно даже назвать «школами». Их в действительности довольно много.

Пробегите ещё раз глазами по тексту данной колонки. Видите ли вы, что кластерный анализ содержит немало мест, не позволяющих ему избежать упреков в произвольности. Я решил затронуть здесь эту тему потому, что в другой литературе об этом не говорится достаточно ясно.

1.5. АНАЛИЗ СООТВЕТСТВИЙ И КВАНТИФИКАЦИЯ ХАЯШИ МЕТОДОМ III

Среди методов анализа, очень похожих на *квантификацию Хаяши методом III* есть метод под названием *анализ соответствий*. Основная цель данного параграфа — описание квантификации Хаяши методом III, но сначала я расскажу об анализе соответствий, так как его атмосферу почувствовать легче. Замечу, что анализ соответствий обычно не относят к методам многомерного анализа.

Анализ соответствий — это метод построения точечного графика на основе перекрёстной таблицы. Или, давая более полное определение, это метод, в котором каждой категории перекрёстной таблицы ставятся в соответствие значения, содержащие в достаточной мере информацию из перекрёстной таблицы. Этот метод, наверное, можно назвать «грубой аэрофотосъёмкой» перекрёстной таблицы.

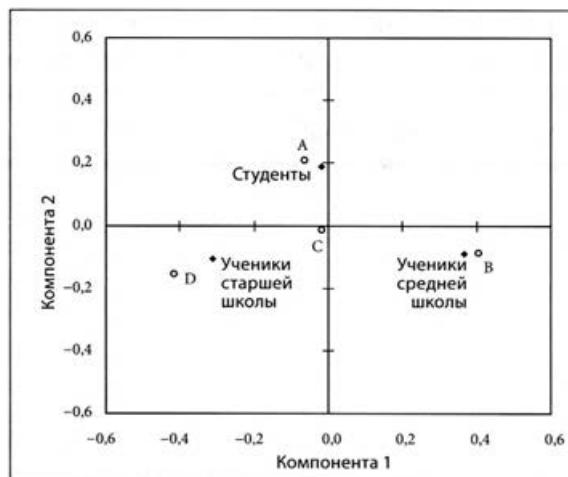
■ КОНКРЕТНЫЙ ПРИМЕР

Ниже приведена перекрёстная таблица на основе результатов одного анкетирования в отношении учащихся средних школ, старших школ и студентов университетов.

		Самый любимый артист				Итого
		A	B	C	D	
Принадлежность респондентов	Ученики средней школы	10	19	13	5	47
	Ученики старшей школы	13	8	15	16	52
	Студенты	18	11	14	8	51
Всего		41	38	42	29	150

Применение: Единицы измерения - число человек.

Анализ соответствий для вышеприведённой таблицы даёт результат, показанный на графике ниже.



Один взгляд на это позволяет понять, что ученикам средней школы нравится артист В, ученикам старшей школы нравится артист D, а студентам нравится артист А.

Итак, переходим к основной теме. *Квантификация Хаяши методом III* — это, можно сказать, анализ соответствий в отношении исходных данных. Можно сказать также, что это:

- метод построения точечного графика на основе исходных данных;
- метод постановки значений, содержащих достаточно информации исходных данных, в соответствие экземплярам (респондентам) и переменным исходных данных;
- метод «грубой аэрофотосъёмки» исходных данных.

Объект анализа соответствий

↓

перекрёстная таблица

	A	B	C	D
I				
II				
III				

Объект квантификации Хаяши методом III

↓

исходные данные

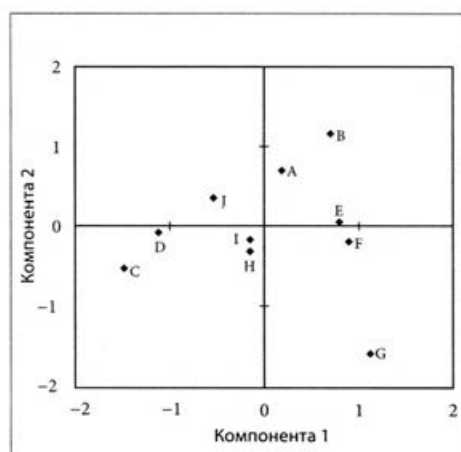
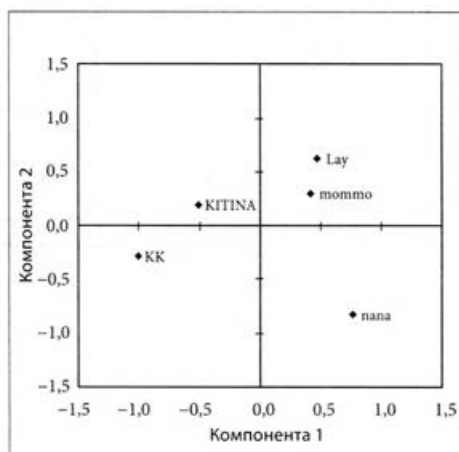
	Пере- менная 1	Пере- менная 2	Пере- менная 3	Пере- менная 4
A				
B				
C				
D				
E				
F				
G				
H				
I				
J				

■ КОНКРЕТНЫЙ ПРИМЕР

В таблице ниже приведены результаты опроса женщин от 20 до 30 лет о том, какие журналы им нравятся.

Респонденты	Журналы				
	КК	папа	mommo	Lay	KITINA
A	0	0	1	1	1
B	0	0	0	1	0
C	1	0	0	0	0
D	1	0	0	0	1
E	0	1	1	1	0
F	0	1	0	1	0
G	0	1	0	0	0
H	1	1	1	0	1
I	1	1	0	1	1
J	1	0	0	1	1

Квантификация Хаяши методом III для вышеприведённой таблицы даёт результат, показанный на графиках ниже.



Сравнение этих графиков позволяет понять, что журнал «КК» нравится респондента С, а журнал «папа» нравится респондентке G».

ЗАМЕЧАНИЕ Читателям, у которых есть интерес к квантификации Хаяши методом III и анализу соответствий, я рекомендую обратиться к моей книге «Изучение анализа соответствий с помощью Excel» (изд. Ohmsha).

1.6. МОДЕЛИРОВАНИЕ СТРУКТУРНЫМИ УРАВНЕНИЯМИ

На нижеприведённых рисунках изображены *диаграммы путей*.

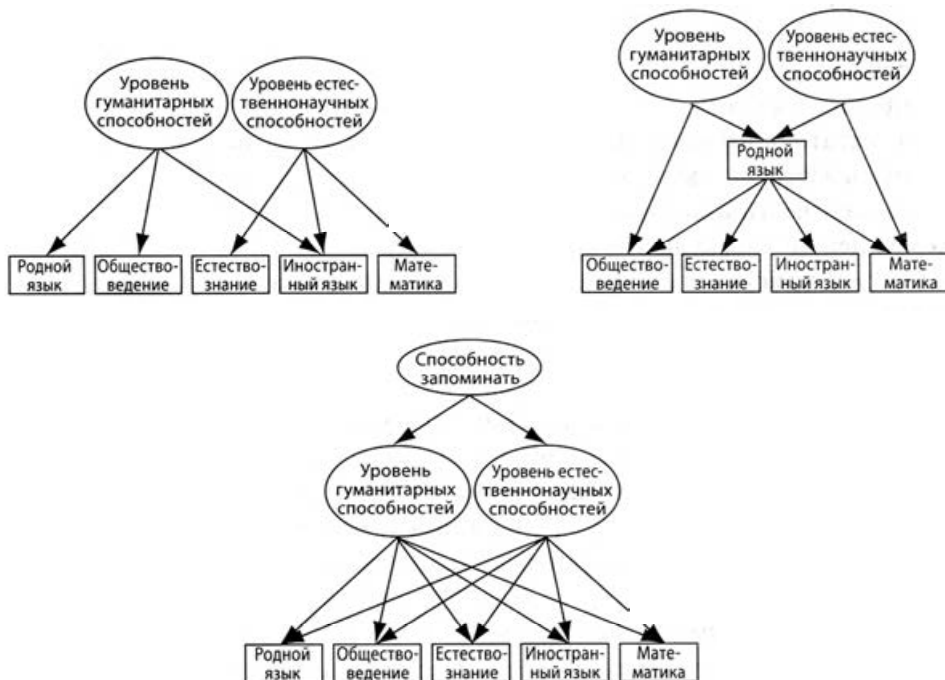


Диаграмма путей — это графическое выражение «субъективной» гипотезы аналитика о структуре, описывающей положение дел в какой-либо сфере жизни. Переменные, обозначенные прямоугольниками, являются наблюдаемыми переменными, а обозначенные овалами — скрытыми переменными.

Моделирование структурными уравнениями — это метод анализа, предназначенный для проверки на истинность диаграммы путей, которую начертил аналитик. Другими словами, это метод проверки на истинность субъективной гипотезы аналитика о структуре, описывающей положение дел. Это также и метод нахождения *путевых коэффициентов* (конкретных значений для каждой стрелки, аналогичных факторным нагрузкам в факторном анализе).

Термин «моделирование структурными уравнениями» является переводом английского названия «Structural Equation Modeling», поэтому для обозначения этого метода используется сокращение SEM.

У этого метода есть также другое название — *анализ ковариационной структуры*, которое, по моим наблюдениям, является более употребительным.

■ КОНКРЕТНЫЙ ПРИМЕР

К сожалению, конкретный пример мне пришлось опустить.

ЗАМЕЧАНИЕ Моделирование структурными уравнениями — это метод анализа, соответствующий «подтверждающему подходу», описанному на стр. 40. Другими словами, порядок проведения анализа следующий:

- 1) построить гипотезу;
- 2) собрать данные;
- 3) провести анализ.

Без достаточного понимания этого вы придёте в замешательство:

- мне неизвестно, сколько есть скрытых переменных;
- мне неизвестны связи между скрытыми и наблюдаемыми переменными;
- мне непонятно, куда нужно проводить стрелки.

Поэтому, приступая к анализу, нужно ясно осознавать, что аналитик, проводящий моделирование структурными уравнениями, обязан сам размышлять над всем этим неизвестным на этапе 1.

Меня очень радует тот факт, что для моделирования структурными уравнениями создано много удобных компьютерных программ. Однако их удобство имеет и отрицательную сторону: у меня создаётся впечатление, что пользуясь ими, многие начинают думать: «Моделирование структурными уравнениями — это так просто». Но на самом деле это не так: по моему мнению, моделирование структурными уравнениями в различных смыслах сложнее, чем даже факторный анализ. Кроме того, в моделировании структурными уравнениями невозможность найти решение часто означает, попросту говоря, неудачу анализа. Поэтому мне кажется, что не стоит (особенно если вы сотрудник компании) легкомысленно заявлять (особенно вашим клиентам): «На этот раз анализ мы попробуем провести с помощью SEM».

2. ДРУГОЕ

2.1. ПРОВЕРКА СТАТИСТИЧЕСКОЙ ГИПОТЕЗЫ

Проверка статистической гипотезы — это предположительная оценка на основе данных выборки: верна или ошибочна выдвинутая аналитиком гипотеза о генеральной совокупности. Её часто называют просто «проверкой».

Это определение не означает, что при проверке статистической гипотезы можно рассуждать так: «Хоть я не очень хорошо понимаю её смысл, но если значение этого p мало, то есть получена какая-то там “значимость”, значит, с точки зрения математики, всё в порядке. Это такая удобная вещь, дающая математическую индульгенцию.» Такое отношение к проверке статистической гипотезы, похоже, довольно распространено, поэтому не только начинающие изучать статистику, но и уже изучавшие её должны отнестись к этому с вниманием.

Проверка статистической гипотезы — это не название какого-либо одного метода, а общее название. Есть множество различных типов такой проверки, например:

- критерий разности средних арифметических в генеральных совокупностях (так называемый t -критерий Стьюдента);
 - критерий независимости (так называемый критерий «хи-квадрат»);
 - критерий разности долей в генеральных совокупностях;
 - критерий отношения дисперсий в генеральных совокупностях;
 - критерий Уилкоксона;
- и многие другие.

■ КОНКРЕТНЫЕ ПРИМЕРЫ

Критерий разности средних арифметических в генеральных совокупностях

Оценим, отличаются или нет «среднее арифметическое месячной суммы на текущие расходы среди всех служащих в городе Токио» и «среднее арифметическое месячной суммы на текущие расходы среди всех служащих в городе Осака».

Служащие	Город	Месячная сумма на текущие расходы (йен)
A	Токио	42500
B	Токио	40800
C	Токио	39400
D	Токио	42800
E	Токио	39800
F	Осака	38700
G	Осака	40000
H	Осака	38500
I	Осака	42100
J	Осака	37000

*CA — среднее арифметическое

Критерий независимости

Оценим, равно или нет нулю значение коэффициента корреляции Крамера (в данной книге не объясняется) для переменных «принадлежность» и «самый любимый артист». Другими словами, оценим, есть ли связь между этими переменными.

		Самый любимый артист				Итого
		A	B	C	D	
Принадлежность респондентов	Ученики средней школы	10	19	13	5	47
	Ученики старшей школы	13	8	15	16	52
	Студенты	18	11	14	8	51
Всего		41	38	42	29	150

Примечание: Единицы измерения - число человек.

ЗАМЕЧАНИЕ Нельзя думать: «Данные говорят о том, что служащие в Токио больше получают на текущие расходы». Необходимо помнить о том, что в таблице приведена информация выборки, а не генеральной совокупности. Повторю ещё раз: проверка статистической гипотезы — это предположительная оценка на основе данных выборки: верна или нет выдвинутая аналитиком гипотеза о генеральной совокупности.

Проверка статистической гипотезы — это хорошо известная процедура, однако она вовсе не так проста, как, похоже, принято считать.

Читателям, интересующимся проверкой статистической гипотезы, я рекомендую обратиться к моей книге «Манга о статистике» (изд. Ohmsha).

2.2. МЕТОД КАПЛАНА-МЕЙЕРА

Метод Каплана-Мейера — это один из способов предположительной оценки выживаемости. Каплан и Мейер — это фамилии исследователей: Эдварда Каплана (Edward Kaplan) и Пола Мейера (Paul Meier).

Особенность метода Каплана-Мейера в том, что если, например, объектом интереса является выживаемость больных раком лёгких, то можно предположительно оценить выживаемость, включив также данные следующих пациентов:

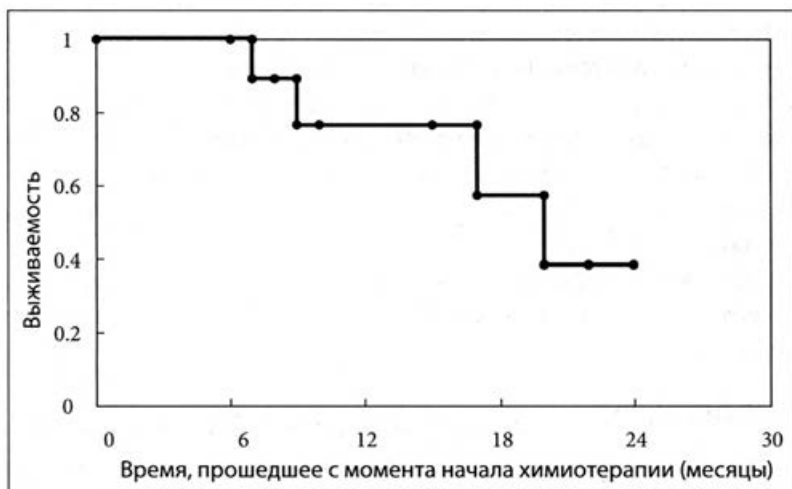
- пациенты, которые в период наблюдения умерли по не связанным с раком лёгких причинам: например, погибли в ДТП;
- пациенты, которые в период наблюдения были переведены в другую клинику;
- пациенты, которые оставались живы до момента окончания наблюдений, заданного исследователем.

■ КОНКРЕТНЫЕ ПРИМЕРЫ

В таблице ниже приведены результаты наблюдения каждого больного раком лёгких в терминальной стадии с момента начала химиотерапии.

Пациенты	Время, прошедшее с момента начала химиотерапии (месяцы)	Исход	
A	17	1	<- Умер от рака лёгких
B	10	0	<- Был жив до момента конца наблюдений
C	15	0	<- Умер по причине, не связанной с раком лёгких
D	20	1	<- Умер от рака лёгких
E	7	1	<- Умер от рака лёгких
F	6	0	<Г Был переведён в другую клинику
G	9	1	↻ Умер от рака лёгких
H	22	0	<- Был жив до момента конца наблюдений
I	8	0	↻ Умер по причине, не связанной с раком лёгких
J	24	0	<- Был жив до момента конца наблюдений

Применение метода Каплана-Мейера даёт результаты, показанные на нижеприведённом графике.



ЗАМЕЧАНИЕ Применять метод Каплана-Мейера только для предположительной оценки выживаемости в какой-либо одной группе, как это было сделано в вышеприведённом примере — это, конечно, совсем не плохо. Однако мне кажется, что это немного расточительно.

Вам следует знать, что можно поступить и по-другому: сначала предположительно оценить с помощью метода Каплана-Мейера выживаемость в каждой из трёх групп: «пациенты, получавшие препарат X», «пациенты, получавшие препарат Y», «пациенты, не получавшие никакого препарата», а затем проверить с помощью *логарифмического рангового критерия* (в этой книге не объясняется), определяются ли различия между ними.

СПИСОК ЛИТЕРАТУРЫ

- 📖 Син Такахаси. «Занимательная статистика.» М.: Додэка, 2010. 224 с.
- 📖 Син Такахаси, Иноуэ Ироха. «Занимательная статистика. Регрессионный анализ.» М.: ДМК Пресс, 2014. 222 с.
- 📖 Мана Такахаси, Сёко Адзума. «Занимательное программирование. Базы данных.» М.: ДМК Пресс, 2014. 240 с.
- 📖 Хироюки Кодзима, Син Тогами. «Занимательная математика. Производные и интегралы.» М.: ДМК Пресс, 2014. 240 с.
- 📖 Митио Сибуя, Хироки Харусэ. «Занимательная математика. Анализ Фурье.» М.: ДМК Пресс, 2014. 256 с.
- 📖 Хидео Нитта, Кейта Такацу «Занимательная физика. Механика.» М.: ДМК Пресс, 2014, 240 с.
- 📖 Казухиро Фудзитаки, Мацуда «Занимательная физика. Электричество.» М.: ДМК Пресс, 2014. 224 с.
- 📖 Киёси Кавабата, Кэндзи Исикава, Ютака Хиираги. «Занимательная астрономия. Вселенная.» М.: ДМК Пресс, 2014. 256 с.

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

А

Анализ

- исследовательский факторный — 96
- кластерный — 224
- ковариационной структуры — 229
- многомерный — 11, 220
- подтверждающий факторный — 196
- проверочный факторный — 196
- совместный (conjoint analysis) — 56
- соответствий — 220, 226
- спектральный — 209

Анкета — 31

- бланк анкеты — 31

В

Вклад

- суммарный — 112
- после исключения влияния других общих факторов — 206
- с игнорированием других общих факторов — 206
- удельный — 112

Вопросы

- с единичным выбором — 48, 49
- с количественным ответом — 48, 51
- с литеральным ответом — 48, 51
- с множественным выбором — 48, 50

Вращение — 70

- наклонное — 174
- ортогональное — 174

Выбор

- двухступенчатый — 19, 24
- простой случайный — 19
- расслоенный — 19, 20
- расслоенный двухступенчатый — 19, 28

Выборка — 17

Г

Генеральная совокупность — 17

Главная компонента — 95

График каменистой осыпи (график Кеттела) — 194

Д

Диаграммы путей — 229

Дисперсия — 87

- несмещённая — 87

Доверительная вероятность — 35

Доверительный уровень — 35

З

Значения главной компоненты — 95

И

Исследование

- выборочное — 17
- качественные — 38
- количественные — 38
- сплошное — 17

К

Квантификация Хаяши методом III — 226, 227

Коэффициент доверия — 35

Критерий независимости — 232

Критерий разности средних арифметических в генеральных совокупностях — 231

М

Матрица

- единичная — 68
- корреляции — 67
- обратная — 84
- симметричная — 76
- транспонированная — 85
- факторного отображения — 172, 200
- факторной структуры — 200
- факторных корреляций — 204
- факторных нагрузок — 172, 200
- целевая — 202

определитель — 198
произведение — 81
сложение — 81

Метод
Андерсена-Рубина — 186
Бартлетта — 186
биквартимакс — 174
биквартимин — 174
варимакс — 174
варимакс без нормализации — 178
варимакс нормализованный — 178
главных компонент — 91
главных факторов — 154
знакомых — 37
Каплана-Мейера — 233
квартимакс — 174
квартимин — 174
коваримин — 174
максимального правдоподобия — 154, 198
множителей Лагранжа — 107
набора — 37
наименьших квадратов — 198
наименьших квадратов (обобщенный) — 198
перехвата — 37
промакс — 174, 202
регрессионной оценки — 186
связей — 37
снежного кома — 37
удобного случая — 37

Методы
взятия выборок — 16, 18
опроса — 16, 30
преднамеренного выбора — 36
случайного выбора — 36

Многомерное нормальное распределение — 198

Моделирование структурными уравнениями (SEM) — 220, 229

Модель
неортогональных факторов — 159
ортогональных факторов — 159

О

Общий фактор (i -й) — 138
Общности — 164
Опрос
- RDD — 30
- до востребования — 30
- интервью-опрос — 30
- интернет-опрос — 30, 31
- почтовый — 30, 31
- телефонный — 30
Опросный лист — 31
Отклик — 98

П

Переменные
- наблюдаемые — 210
- объясняющие — 98
- скрытые (ненаблюдаемые) — 210
Подтверждающий подход — 40
Проверка — 231
- статистической гипотезы — 231
Путевые коэффициенты — 210, 229

Р

Размер выборки — 34
Регрессия
- логистическая — 222
- множественная — 221

С

Собственные векторы и значения — 73
Специфические факторы — 144
Среднеквадратичное отклонение — 87
Сумма квадратов отклонений — 87

Т

Тест согласованности — 195, 198

У

Уравнение множественной регрессии — 167

Ф

Фактор i (i -й) — 138
Факторные значения — 138, 186
Факторные нагрузки — 149

ОБ АВТОРЕ

Син Такахаси родился в 1972 году в префектуре Ниигата, закончил аспирантуру Факультета исследований дизайна Университета дизайна Кюсю (в настоящее время — Университета Кюсю) по специализации «Передача информации». В настоящее время, после работы в частных компаниях, где Син занимался анализом данных и проведением семинаров, посвятил себя писательской деятельности.



Книги издательства «ДМК Пресс» можно заказать в торгово-издательском холдинге «Планета Альянс» наложенным платежом, выслав открытку или письмо по почтовому адресу:

115487, г. Москва, 2-й Нагатинский пр-д, д. 6А

При оформлении заказа следует указать адрес (полностью), по которому должны быть высланы книги; фамилию, имя и отчество получателя.

Желательно также указать свой телефон и электронный адрес.

Эти книги вы можете заказать и в интернет-магазине: www.ailians-kniga.ru.

Оптовые закупки: тел. (499) 782-38-89

Электронный адрес: books@alians-kniga.ru.

Син Такахаси (автор), Иноуэ Ироха (художник)

Занимательная статистика. Факторный анализ. Манга

Издательство выражает благодарность *В. О. Панфилову*

Главный редактор *Д. А. Мовчан*

dmkpress@gmail.com

Перевод с японского *А. Б. Клионский*

Научный редактор *И. А. Сенников*

Верстальщик *А. Ю. Анненков*

Корректор *Г.Б.Абудеева*

Формат 70x100/16. Бумага офсетная.

Печать офсетная. Объем 15 п. л. Уел. п. л. 17,5. Тираж 1000 экз. Заказ 7643.

Веб-сайт издательства ДМК Пресс: www.dmkpress.com



Отпечатано в ОАО «Можайский полиграфический комбинат».
143200, г. Можайск, ул. Мира, 93.

www.oaompk.ru, www.oaompk.ru тел.: (495) 745-84-28, (49638) 20-685

ОБРАЗОВАТЕЛЬНАЯ МАНГА

ЗАНИМАТЕЛЬНАЯ ФИЗИКА



ПРОСТОЙ,
ПОСЛЕДОВАТЕЛЬНЫЙ,
НАГЛЯДНЫЙ И
НЕОБРЕМЕНИТЕЛЬНЫЙ ПУТЬ
ИЗУЧЕНИЯ
ШКОЛЬНОЙ И
ИНСТИТУТСКОЙ
ПРОГРАММЫ!

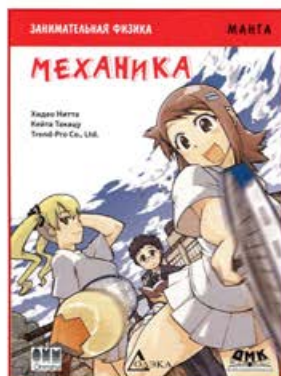
- ПОЧЕМУ
ЯПОНЦЫ
ТАКЖЕ УМНЫЕ?



ЧИТАЙ КОМИКСЫ И
СТАНОВИСЬ ОТЛИЧНИКОМ!



- ОНИ УЧАТСЯ
ПО МАНГЕ!



ВСЕ КНИГИ ПРОВЕРЕНЫ
ПРЕПОДАВАТЕЛЯМИ
И ПРОФЕССОРАМИ
ВЕДУЩИХ ПРОФИЛЬНЫХ
ВУЗОВ
ДЛЯ ТОЧНОГО ПЕРЕВОДА
И АДАПТАЦИИ.



КНИГИ СЕРИИ "ОБРАЗОВАТЕЛЬНАЯ МАНГА"
МОЖНО ЗАКАЗАТЬ НА САЙТЕ ИЗДАТЕЛЬСТВА
WWW.DMKPRESS.COM ИЛИ WWW.AMK.RF

**Нет кризису! Весь 2015 год при заказе на сайте
цена 349 руб + доставка**

ОБРАЗОВАТЕЛЬНАЯ МАНГА

ЗАНИМАТЕЛЬНАЯ МАТЕМАТИКА

ЗАНИМАТЕЛЬНОЕ ПРОГРАММИРОВАНИЕ



ПРОСТОЙ,
ПОСЛЕДОВАТЕЛЬНЫЙ,
НАГЛЯДНЫЙ И
НЕОБРЕМЕНИТЕЛЬНЫЙ ПУТЬ
ИЗУЧЕНИЯ
ШКОЛЬНОЙ И
ИНСТИТУТСКОЙ
ПРОГРАММЫ!

ЯПОНЦЫ
ТАКЖЕ УМНЫЕ,
ПОТОМУ ЧТО
УЧАТСЯ
ПО МАНГЕ!



ЧИТАЙ КОМИКСЫ И
СТАНОВИСЬ ОТЛИЧНИКОМ!



ВСЕ КНИГИ ПРОВЕРЕНЫ
ПРЕПОДАВАТЕЛЯМИ
И ПРОФЕССОРАМИ
ВЕДУЩИХ ПРОФИЛЬНЫХ
ВУЗОВ
ДЛЯ ТОЧНОГО ПЕРЕВОДА
И АДАПТАЦИИ.

ISBN 978-5-97060-270-6



КНИГИ СЕРИИ "ОБРАЗОВАТЕЛЬНАЯ МАНГА"
МОЖНО ЗАКАЗАТЬ НА САЙТЕ ИЗДАТЕЛЬСТВА
WWW.DMKPRESS.COM ИЛИ WWW.AMK.PF

**Нет кризису! Весь 2015 год при заказе на сайте
цена 349 руб + доставка**