欧姆社学习漫画

漫画统计学之因子分析

〔日〕高桥　信/著
〔日〕Inoue Iroha/漫画绘制
〔日〕株式会社TREND-PRO/漫画制作
张仲桓/译

科学出版社
www.sciencep.com

KindleDX出版署

✿前　言✿

本书介绍了有关因子分析以及主成分分析的统计学知识。

本书主要面向以下读者：

● 阅读完《漫画统计学》或者具备相同程度以上的统计学知识的读者

● 对因子分析感兴趣的读者

● 对主成分分析感兴趣的读者

● 对问卷调查感兴趣的读者

本书由以下几部分构成：

● 第 1 章　问卷调查的基础知识

● 第 2 章　调查问卷和问题设计

● 第 3 章　数学的基础知识

● 第 4 章　主成分分析

● 第 5 章　因子分析

各章又包括：

● 漫画部分

● 对漫画部分进行补充的文字说明

此外，本书还备有附录。附录中介绍了一些在《漫画统计学》和《漫画统计学之回归分析》里出现过的、著名的分析方法。

本书中的相关计算过程记录得相当详细，数学基础好的读者只需仔细地看一遍即可，数学基础稍差的读者则要用心揣摩，多加思考。但不管怎么说，即便是那些觉得“意思不太明白、计算起来也困难”的读者，无论如何都要按照书中的步骤把解求出来，这样做起码可以掌握大致的计算流程。没必要强迫自己马上就理解，要不焦不躁地坚持读到最后。在阅读过程中，请您一定要全神贯注地阅读。

作为本书主要内容的因子分析，其计算过程与《漫画统计学》和《漫画统计学之

回归分析》中的内容相比更为复杂。所以在看到有关计算过程的部分时，可能有不少读者会觉得过于复杂难懂，因而没有心情看下去。这时请您坚持住，不要气馁，继续努力。因子分子的计算确实比较复杂，但是只要具备大学入学考试水平（理科）的数学知识，就绝对不会感觉到“困难”。如此说来，对于数学基础不是很好的读者就不太好讲了，可能还是会觉得比较困难，但无论如何，都请您不焦不躁地读到最后。

在阅读的过程中，会存在读者自己的计算结果和书中的计算结果不一致的情况，这是由于四舍五入的原因。如带来不便，还请诸位读者多多包涵。

能够有这次执笔的机会，我要感谢株式会社欧姆社开发局的诸位；感谢将我的原稿制成漫画的株式会社 TREND-PRO 的诸位；感谢负责脚本的 re_akino 先生，以及负责绘画的 Inoue Iroha 先生。对于立教大学社会学系的酒折文武先生提出的诸多建议，在此表示深深的谢意。

高桥　信

✿目　录✿

出场人物介绍

高津　露儿

大一新生。天生的乐天派，喜怒哀乐变幻无常，但也有专一的一面，课余时间在诺伦茶餐厅兼职打工。

高津　博

露儿的父亲。任职于营销公司，担任部长之职。十分宠爱独生女，言行也略带孩子气。

山本　守

露儿父亲的下属，生活中总是不修边幅。曾经做过露儿的家庭教师，教她统计学知识。现在和露儿的关系在朋友和恋人之间徘徊。

五十岚　美羽

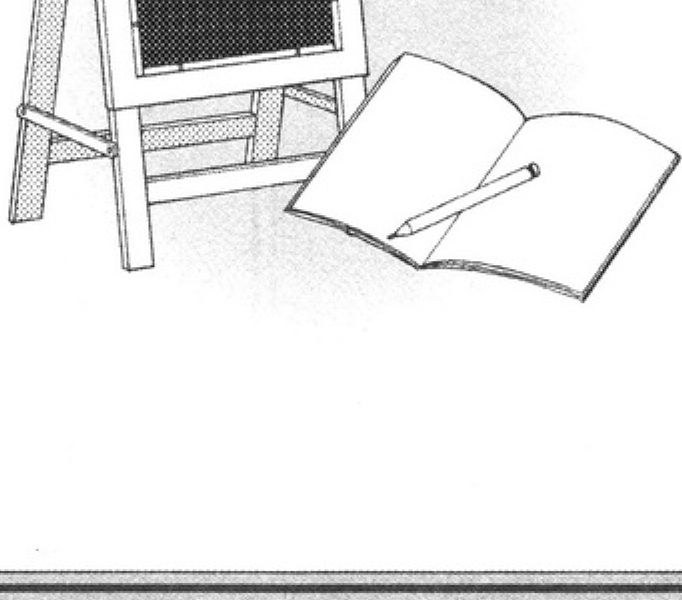
大学三年级学生，露儿打工时的伙伴。从前一直认为数学很难，但是，自从某个契机之后就开始对统计学着迷了。

宫野　理纱

研究生，露儿打工时的伙伴。性格直爽，颇有大姐姐风范。头脑聪颖，整日忙于研究学习。

序　言

你和我的因子分析

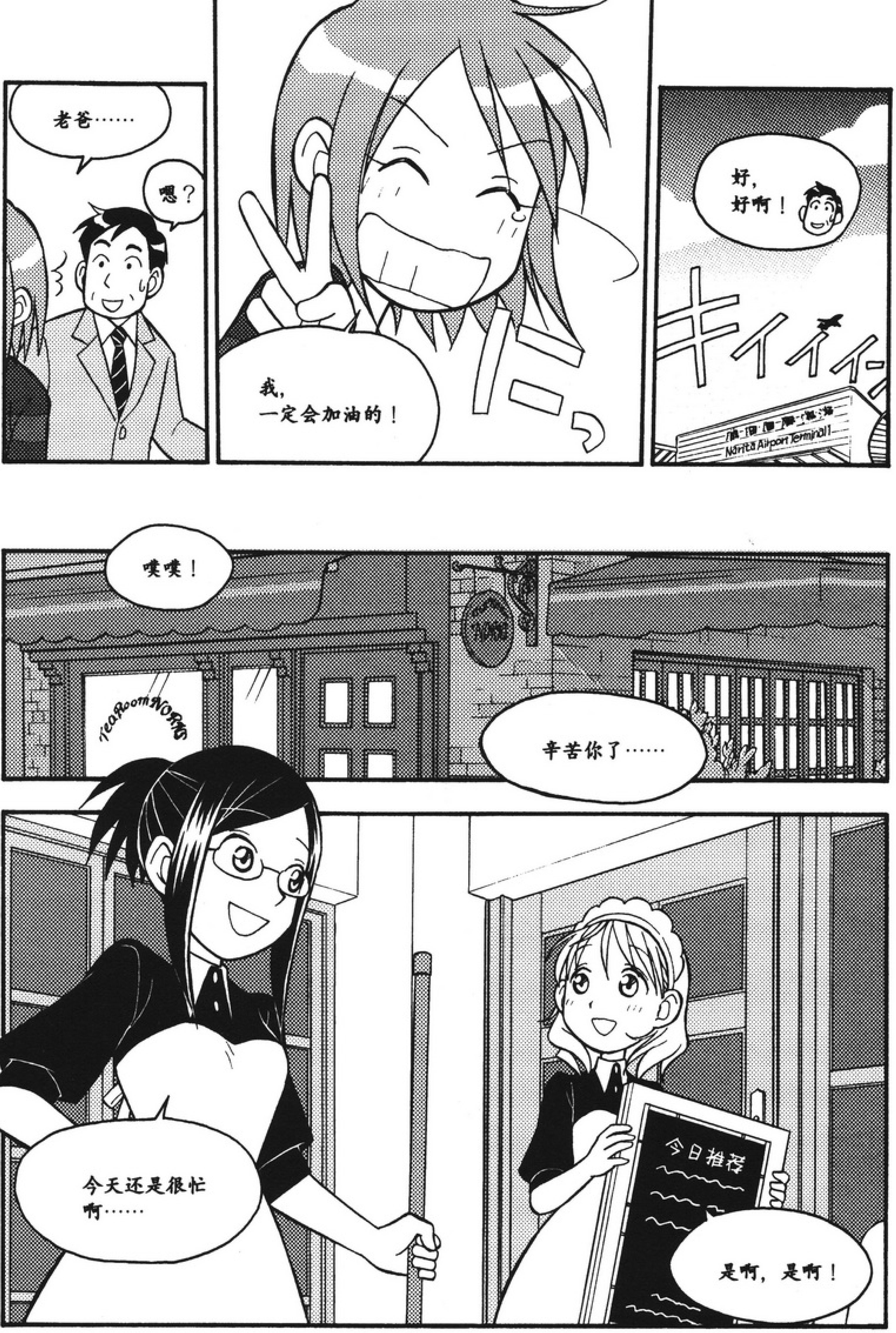


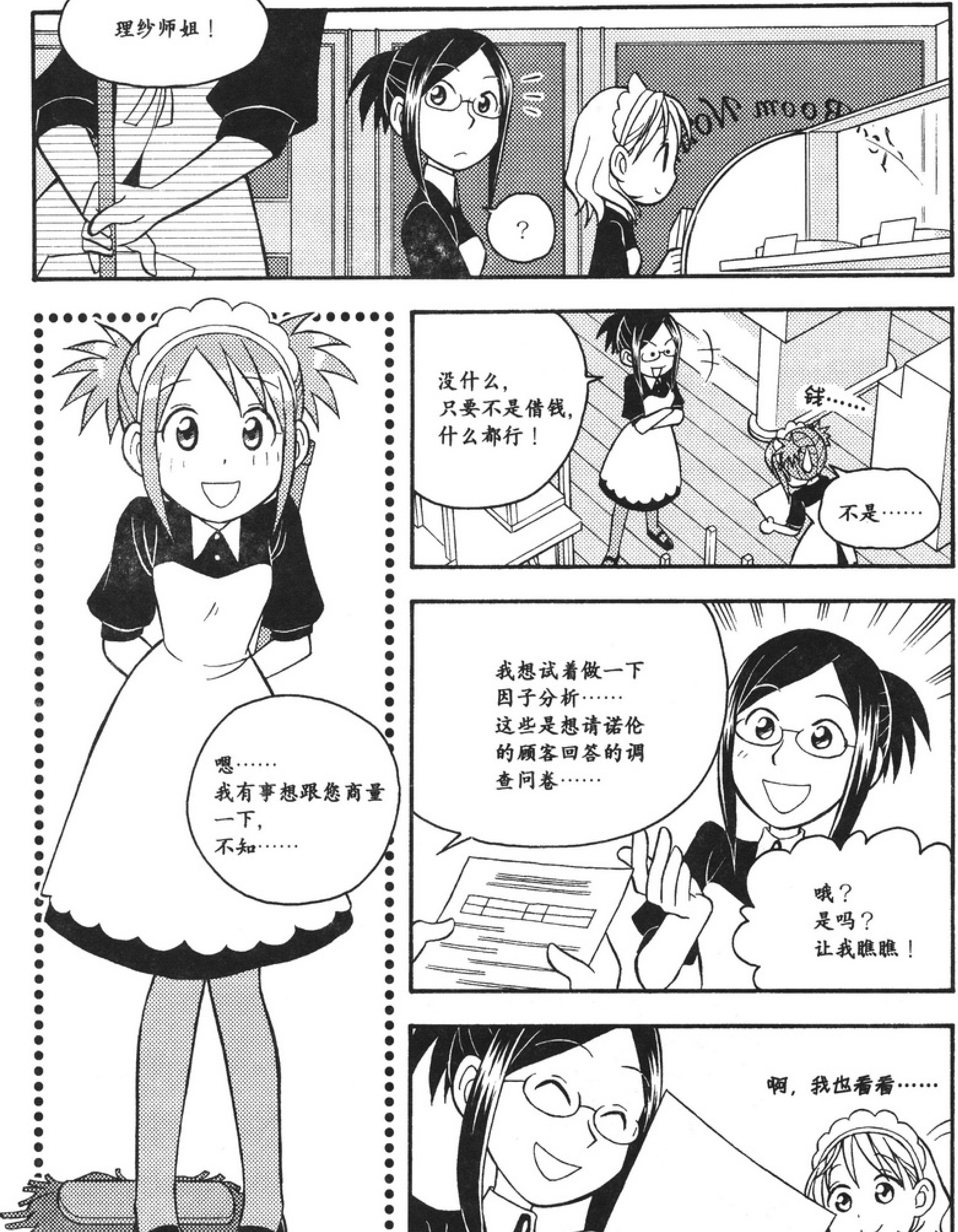

キィィーン
山本啊啊啊啊啊啊！！
山本要去国外工作了啊……
あああ
ああ
这也是没有办法的事啊！
露儿！
呜……呜……


多亏了山本这位
优秀的家庭教师，
露儿才顺利
地考上了大
学啊……
露儿，你在日本
努力完成大学的学业
就可以啦！
耶！
怎么……
我也要跟他留学去！
喂、
喂！
哎呀，不能去找山本，
他在那边也有很多正事
要办……
我不！
露儿，
你别去添乱了
行不行啊！？


咱们回去吧！
把山本的事情……
忘了吧！
添乱……
爸爸，山本去国外干什么啊？
你是他的上司，应该知道吧？
这个？
嗯……怎么说呢，无非就是市场营销方面的工作！
就是去分析调查问卷！
分析调查问卷……是什么啊？
就是……
比如说，因子分析之类的。
因子分析……
ポーン
呀！对于露儿来说，是不是听起来有些复杂啊？
バラバラ
因子分析……

バラ
山本！
小露儿！
因子分析
露儿为了成为山本的助手而发奋用功！
哈哈！终于成了因子分析的高手！
从此以后，我们一起奋斗吧！
嗯！
就这样！
グッ
露，露儿……


老爸……
嗯？
我，
一定会加油的！
好，
好啊！
Narita Airport Terminal1
噗噗！
辛苦你了……
今天还是很忙
啊……
今日推荐
是啊，是啊！



理纱师姐！
？
嗯……
我有事想跟您商量一下，
不知……
没什么，
只要不是借钱，
什么都行！
啫……
不是……
我想试着做一下
因子分析……
这些是想请诺伦
的顾客回答的调
查问卷……
哦？
是吗？
让我瞧瞧！
啊，我也看看……

诺伦茶餐厅调查问卷

参与此次调查的顾客将可获得我们赠送的优惠券，请在结账时索取！

性别	男、女	年　龄	岁
职业		年收入	万日元

Q. 您感觉店内的就餐环境如何？

非常不好　　不好　　一般　　好　　非常好

Q. 您感觉女服务生的制服和服务态度如何？

非常不好　　不好　　一般　　好　　非常好

Q. 您觉得我们的红茶味道如何？

非常不满意　　不满意　　一般　　满意　　非常满意

Q. 您觉得价格如何？

便宜　　适中　　贵

Q. 对于以下 4 种红茶，请您按照自己的喜好程度进行排序。

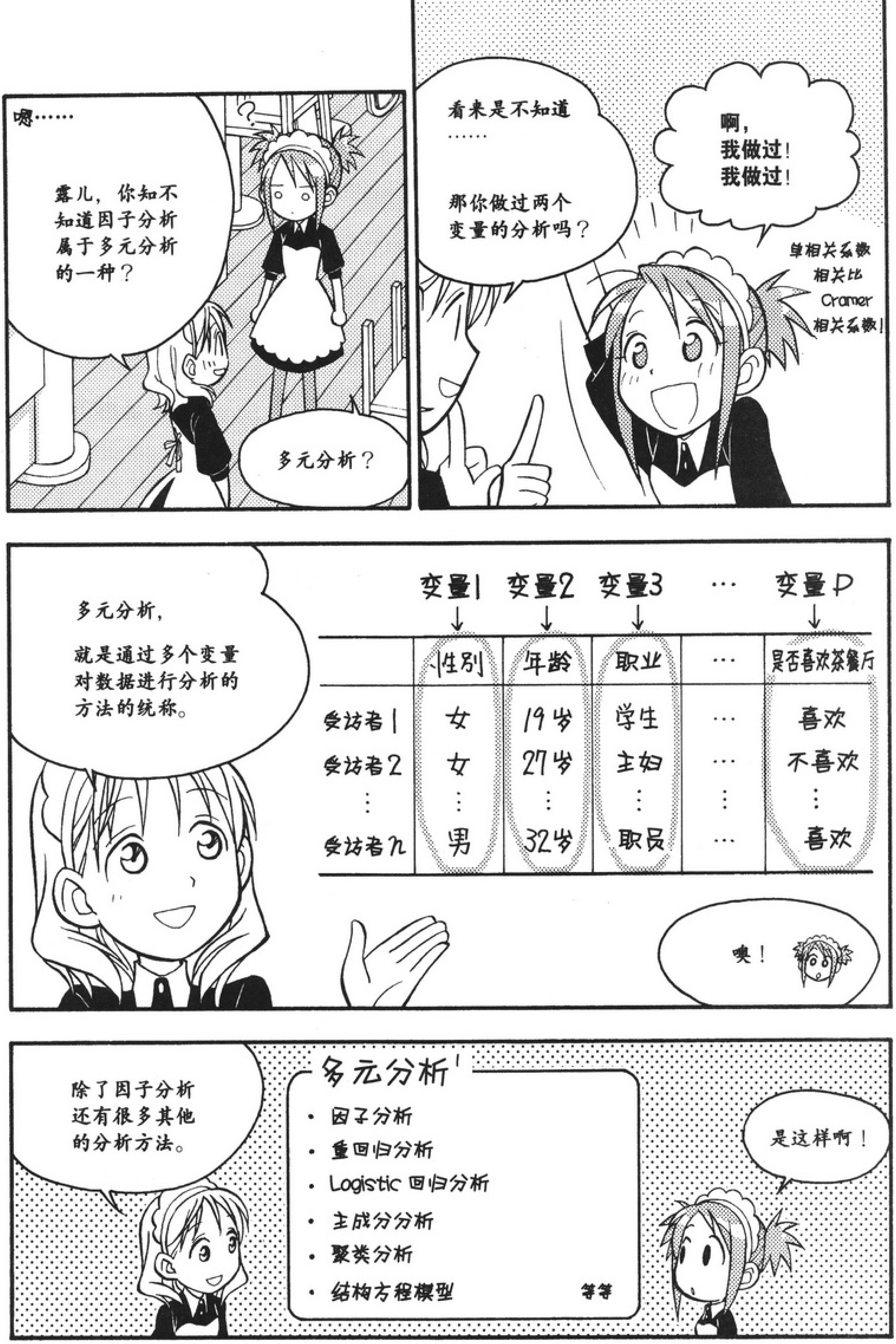
1. 原味红茶	→	第[　　]名
2. 柠檬茶	→	第[　　]名
3. 奶茶	→	第[　　]名
4. 玫瑰茶	→	第[　　]名

Q. 您喜欢在茶餐厅就餐吗？

喜欢　　不喜欢

这哪里是什么因子分析啊！更不会有人回答你的问题！
啊！？这可是我费了半天劲才弄出来的啊！
露儿啊，露儿……你到底知不知道什么叫因子分析啊？
这个……我只知道它的名字……
えへ！
我也算是学过统计学的，只不过……
你以为做个调查问卷，把这些数据收集起来，就可以做因子分析了？那你可就大错特错了！
因子分析可不是这么简单就能够完成的分析方法啊！
·那，它是什么呀!?
那么，理纱师姐，请您教教我吧……
您在研究生院研究统计学，所以一定是这方面的高手！
啊——
这种事似乎发生过啊\
噢
ばっ


ブツブツブツ
这么说的话，有一个比我还合适的人选。
哦？是谁呢？
就是……美羽！
就麻烦你给露儿讲讲因子分析吧！
要我讲吗？
咔？美羽师姐的统计学也学得很好吗？
真的吗？
那当然了，美羽的统计学很棒的哟！
那个……嗯，我会像理纱师姐那样……
她的哥哥在营销公司工作……
她又是我教出来的学生


嗯……
露儿，你知不知道因子分析属于多元分析的一种？
多元分析？
看来是不知道……
那你做过两个变量的分析吗？
啊，我做过！我做过！
单相关系数
相关比
Cramer相关系数！
多元分析，
就是通过多个变量对数据进行分析的方法的统称。
变量1 变量2 变量3 … 变量P
性别 年龄 职业 … 是否喜欢茶餐厅
受访者1 女 19岁 学生 … 喜欢
受访者2 女 27岁 主妇 … 不喜欢
受访者n 男 32岁 职员 … 喜欢
噢！
除了因子分析还有很多其他的分析方法。
多元分析1
· 因子分析
· 重回归分析
· Logistic回归分析
· 主成分分析
· 聚类分析
· 结构方程模型
等等
是这样啊！

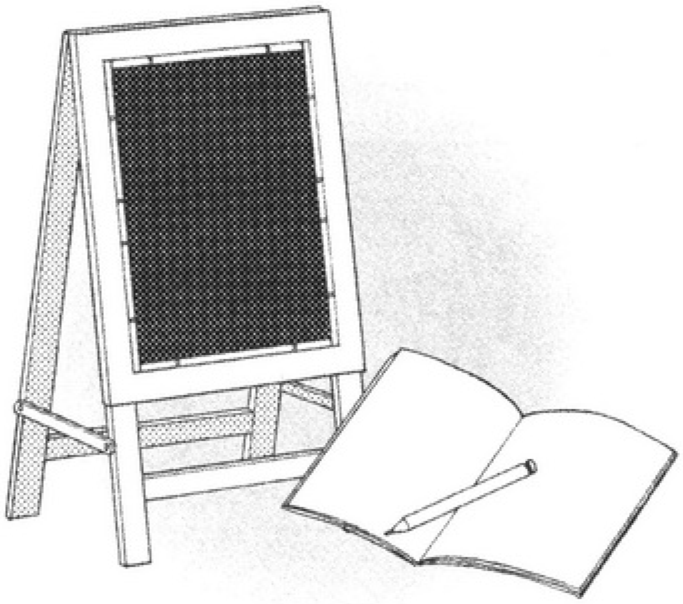
1. 详见附录（219 页）。

那么，所谓因子分析就是将数据背后的……
潜在共性因子挖掘出来的一种分析方法。
什么意思？
简单地说，就是通过问卷调查的结果，
将受访者的真实想法归纳出来的一种分析方法。
这么认为
那么认为
这种情况下有5个变量
调查问卷
Q1
Q2
Q3
Q4
Q5
A1
A2
A3
A4
A5
会有几种"想法"呢……
那些又会是什么"想法"呢……


嘿——嘿
原来是这样！
你看，美羽老师的课是不是讲得相当精彩啊？
是啊！
哎呀，老师可不敢当……
以前，美羽的统计学学得并不好……
正因如此，她更清楚哪些地方是难点，
所以我才认为她更适合做露儿的老师。
美羽师姐，请多多指教！
请您教我因子分析吧！
拜托您了！！

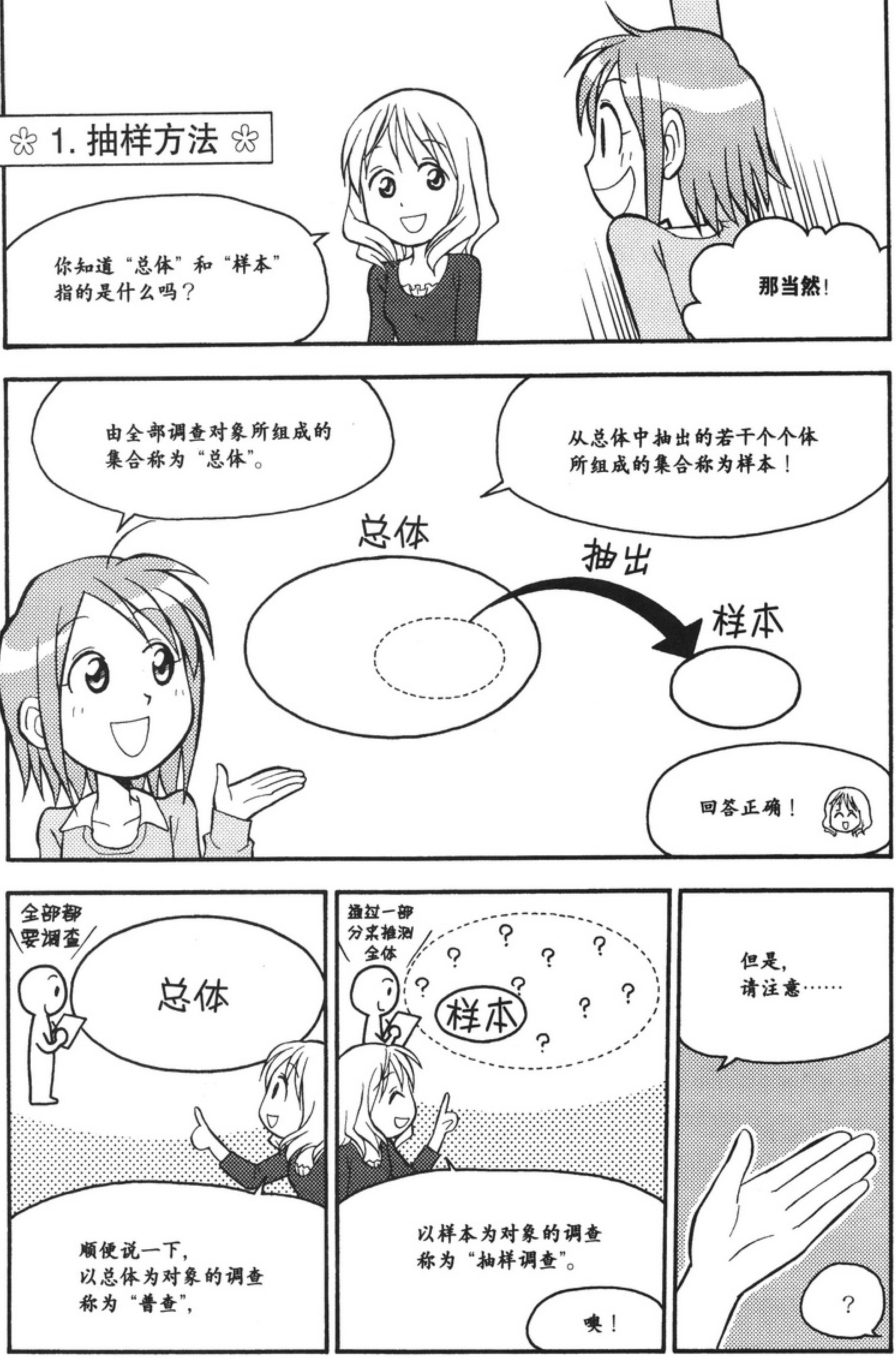

啊……
那好吧！
我们就
一起努力吧！
真不愧是
美羽！
太好啦！
请您多多关照！
喂，赶快换衣服
回家吧！
是！！
等着我啊，
山本！！

第1章
问卷调查的基础知识

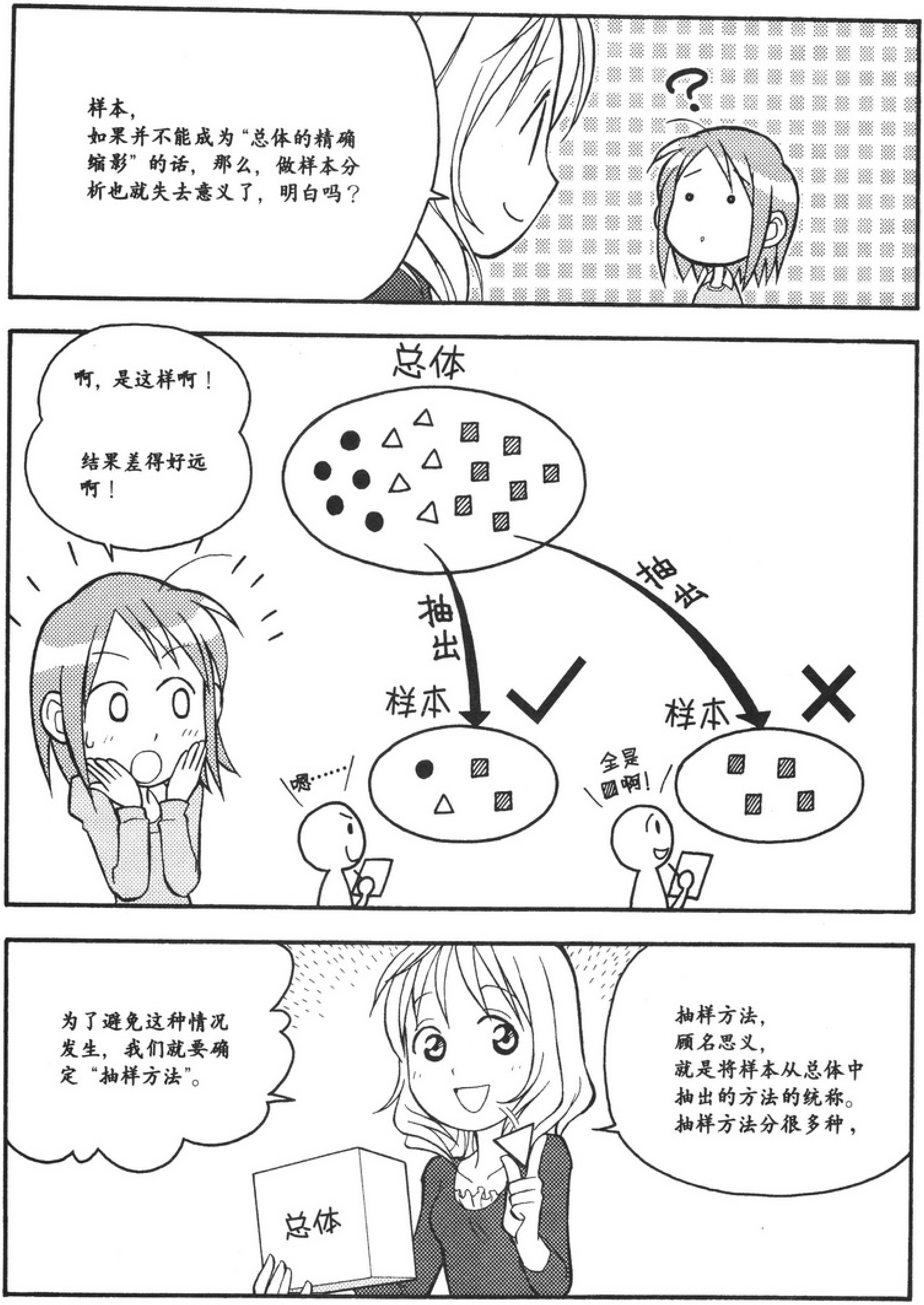


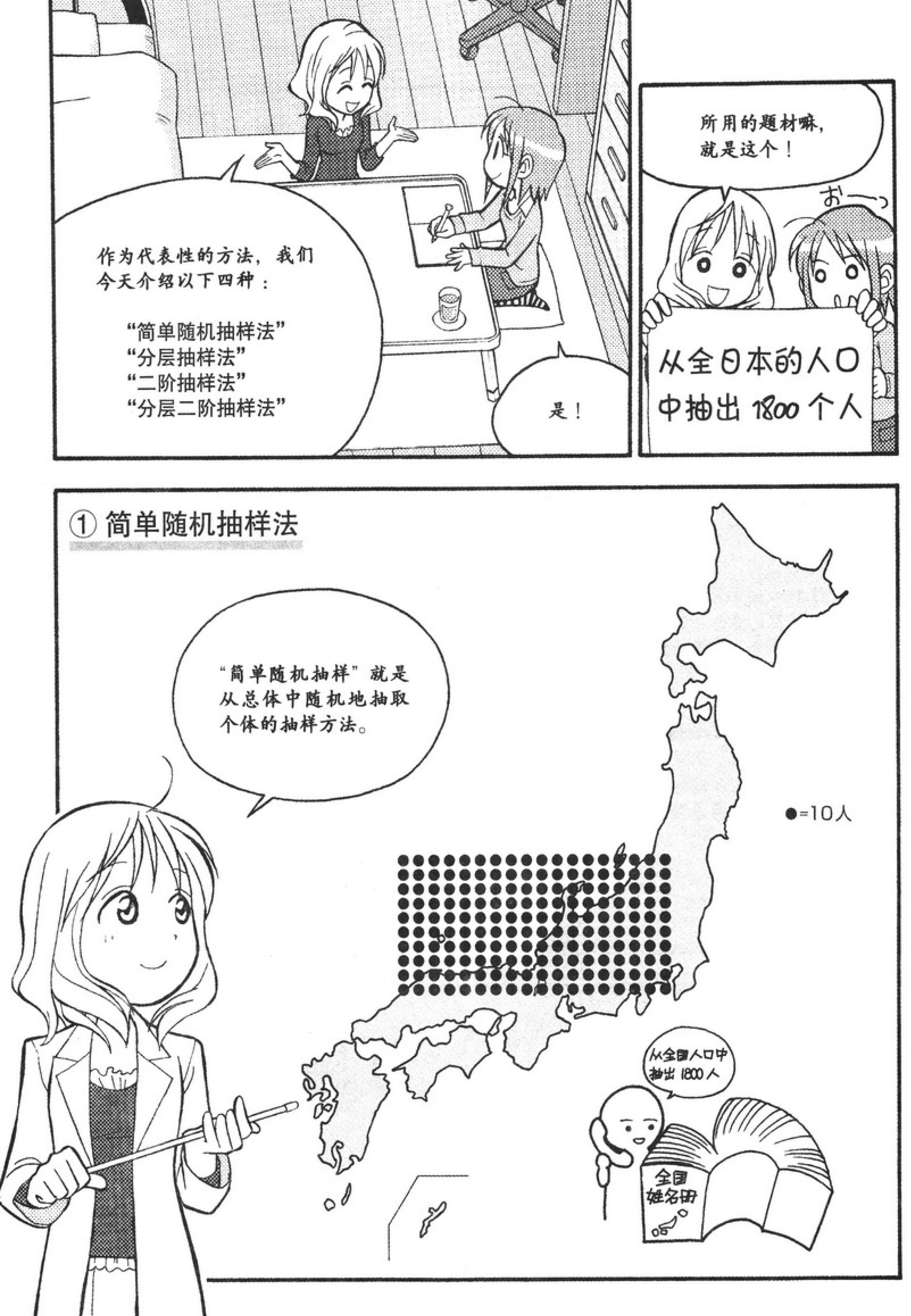
1. 抽样方法
2. 调查方法
3. 样本容量的标准
4. 随机抽样和定向抽样
5. 定量调查和定性调查
6. 数据分析的搭配方法

高津
那么，今天我们就来学习一下问卷调查的基础知识吧！
啊？我们不是学习因子分析吗？
因子分析大多是先通过问卷调查来收集数据，然后再进行分析的，
所以我们必须要先掌握问卷调查的相关知识。
哎呀，原来是这样啊！
好了，今天我们就先学“抽样方法”和“调查方法”吧！
ゴオオオ
辛苦你了！



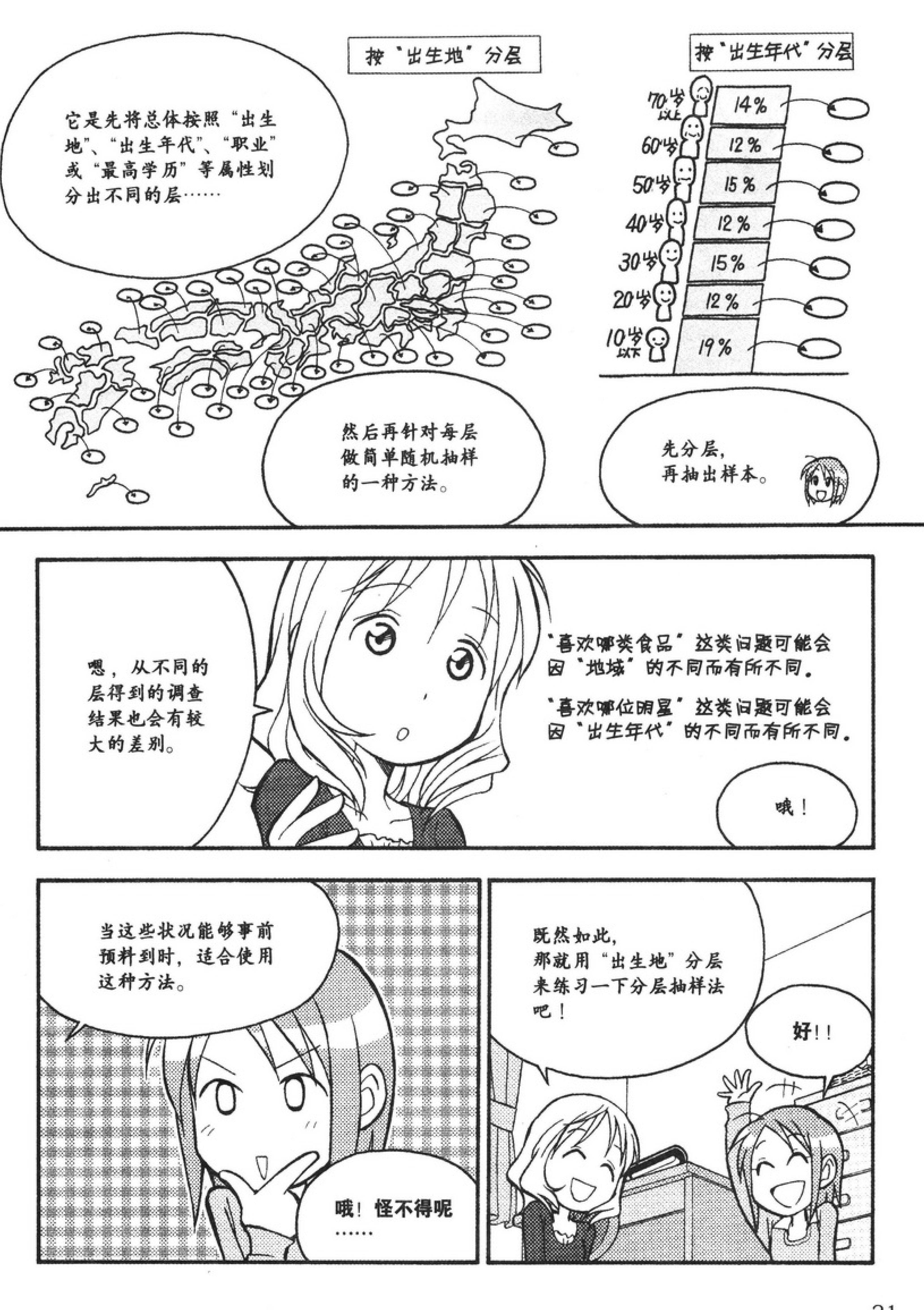
1. 抽样方法
你知道“总体”和“样本”指的是什么吗？
那当然！
由全部调查对象所组成的集合称为“总体”。
从总体中抽出的若干个个体所组成的集合称为样本！
总体
抽出
样本
回答正确！
全部都要调查
总体
顺便说一下，以总体为对象的调查称为“普查”，
通过一部分来推测全体
样本
以样本为对象的调查称为“抽样调查”。
噢！
但是，请注意……
？

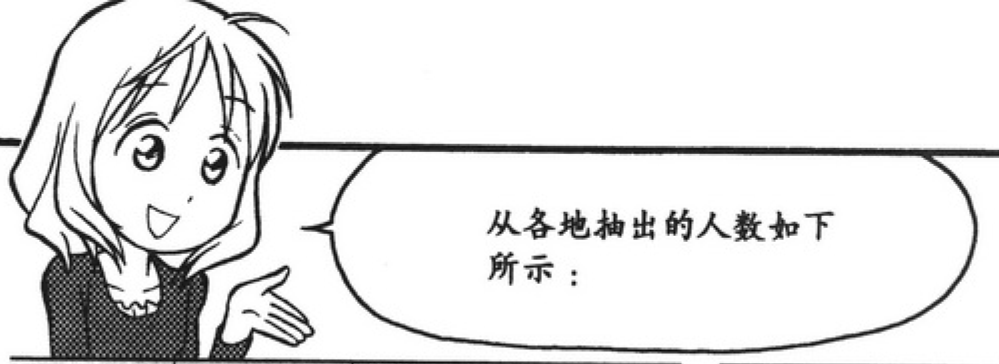

样本，
如果并不能成为“总体的精确缩影”的话，那么，做样本分析也就失去意义了，明白吗？
啊，是这样啊！
结果差得好远啊！
总体
抽出
抽出
样本
样本
嗯……
全是▨啊！
为了避免这种情况发生，我们就要确定“抽样方法”。
抽样方法，
顾名思义，
就是将样本从总体中抽出的方法的统称。
抽样方法分很多种，
总体



作为代表性的方法，我们今天介绍以下四种：
“简单随机抽样法”
“分层抽样法”
“二阶抽样法”
“分层二阶抽样法”
是！
所用的题材嘛，就是这个！
お一っ
从全日本的人口中抽出 1800 个人
① 简单随机抽样法
“简单随机抽样”就是从总体中随机地抽取个体的抽样方法。
●=10人
从全国人口中抽出 1800 人
全国姓名册


恰如其名，确实是很简单呐！
没错！
但是，如果全国姓名册没有将所有人包含在内的话，就不能进行抽取，所以实施起来有难度。
的确如此……
这种方法常常用来进行像“某大型企业的员工意识调查”这样的调查，
有时虽然可以拿到全体成员的名册，但是人数过多，不适合进行普查，这时可以用简单随机抽样的方法。
员工名单
嗯！
② 分层抽样法
分层抽样法
接下来是分层抽样法。


按“出生地”分层
它是先将总体按照“出生地”、“出生年代”、“职业”或“最高学历”等属性划分出不同的层……
按“出生年代”分层
70岁以上
60岁
50岁
40岁
30岁
20岁
10岁以下
14%
12%
15%
12%
15%
12%
19%
然后再针对每层做简单随机抽样的一种方法。
先分层，再抽出样本。
嗯，从不同的层得到的调查结果也会有较大的差别。
“喜欢哪类食品”这类问题可能会因“地域”的不同而有所不同。
“喜欢哪位明星”这类问题可能会因“出生年代”的不同而有所不同。
哦！
当这些状况能够事前预料到时，适合使用这种方法。
哦！怪不得呢……
既然如此，那就用“出生地”分层来练习一下分层抽样法吧！
好!!

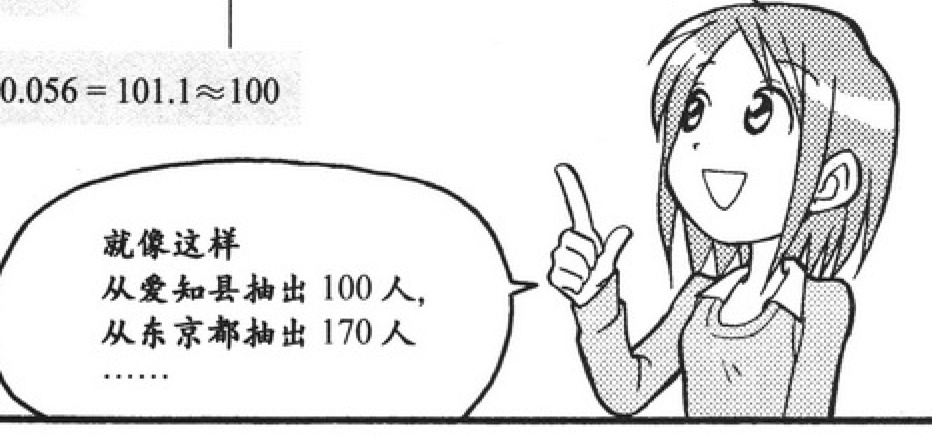


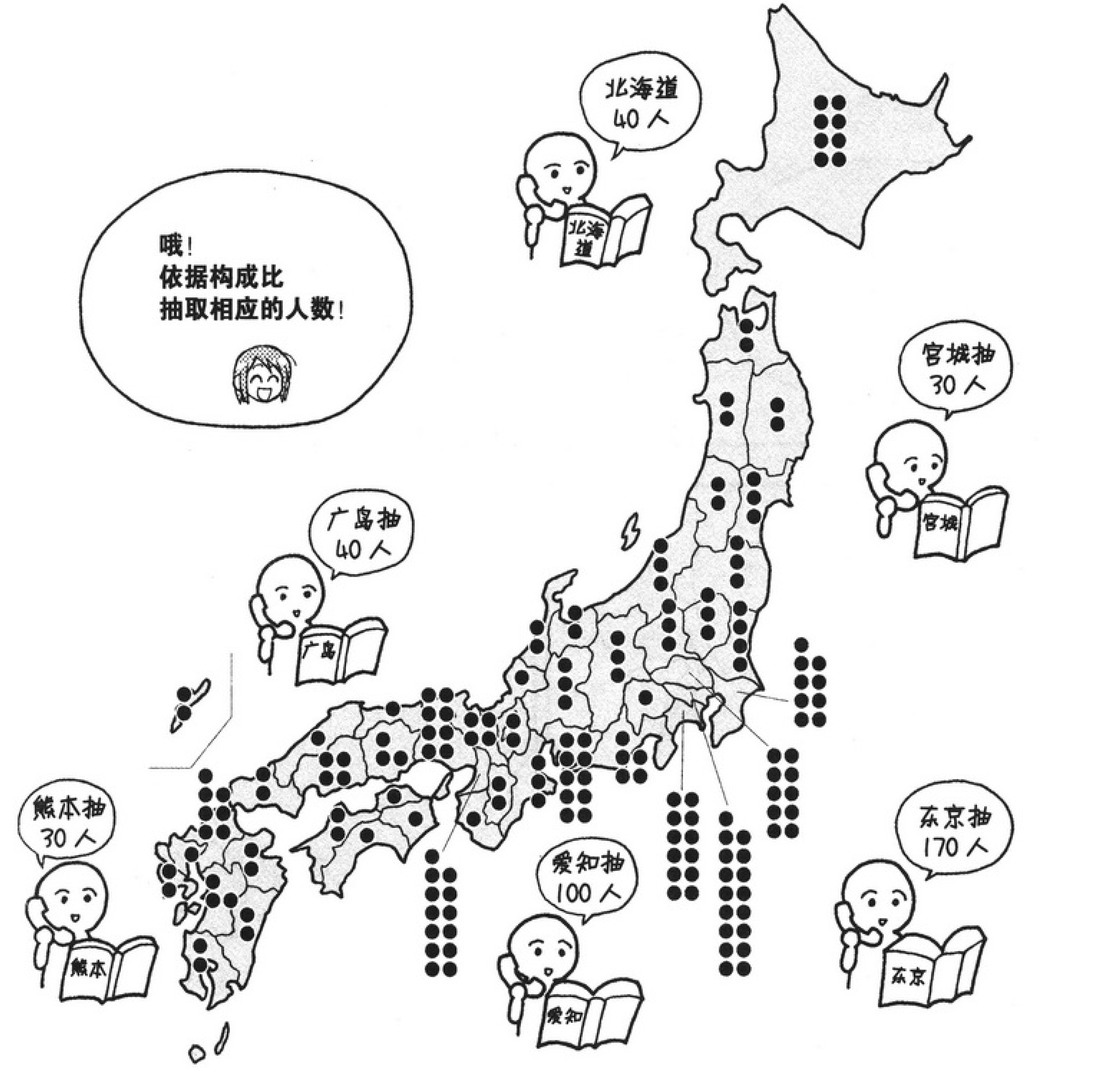
	人口	构成比	抽取人数
北海道	5 600 000	0.045	80
青森县	1 400 000	0.011	20
岩手县	1 400 000	0.011	20
宫城县	2 300 000	0.018	30
秋田县	1 100 000	0.009	20
山形县	1 200 000	0.010	20
福岛县	2 100 000	0.017	30
茨城县	2 900 000	0.023	40
枥木县	2 000 000	0.016	30
群马县	2 000 000	0.016	30
埼玉县	6 900 000	0.055	100
千叶县	5 900 000	0.047	90
东京都	12 000 000	0.096	170
神奈川县	8 400 000	0.067	120
新泻县	2 400 000	0.019	30
富山县	1 100 000	0.009	20
石川县	1 100 000	0.009	20
福井县	800 000	0.006	10
山梨县	800 000	0.006	10
长野县	2 200 000	0.018	30
岐阜县	2 100 000	0.017	30
静冈县	3 700 000	0.030	50
爱知县	7 000 000	0.056	100
三重县	1 800 000	0.014	30
滋贺县	1 300 000	0.010	20
京都府	2 600 000	0.021	40
大阪府	8 700 000	0.070	130
兵库县	5 500 000	0.044	80
奈良县	1 400 000	0.011	20
和歌山县	1 000 000	0.008	10
鸟取县	600 000	0.005	10
岛根县	700 000	0.006	10
冈山县	1 900 000	0.015	30
广岛县	2 800 000	0.022	40
山口县	1 500 000	0.012	20
德岛县	800 000	0.006	10
香川县	1 000 000	0.008	10
爱媛县	1 400 000	0.011	20
高知县	800 000	0.006	10
福冈县	5 000 000	0.040	70
佐贺县	800 000	0.006	10
长崎县	1 500 000	0.012	20
熊本县	1 800 000	0.014	30
大分县	1 200 000	0.010	20
宫崎县	1 100 000	0.009	20
鹿儿岛县	1 700 000	0.014	20
冲绳县	1 300 000	0.010	20
总计	124 600 000	1	1800

（以上数据参考2006年度日本总务省统计局资料制成）

$$\frac{\text{当地人口数}}{\text{总人口}} = \frac{7\,000\,000}{124\,600\,000} = 0.056$$

$$\text{总抽取人数} \times \text{构成比} = 1800 \times 0.056 = 101.1 \approx 100$$




北海道
40人
北海道
哦！
依据构成比
抽取相应的人数！
宫城抽
30人
宫城
广岛抽
40人
广岛
熊本抽
30人
熊本
爱知抽
100人
爱知
东京抽
170人
东京


不过……
这种方法是不是也
一定要有全员名单
呢？
嗯，没错！


而且，名册中还一
定要登记分层时所
依据的信息。
噢！
是这样啊！
年龄 住址
21 神奈川县
23 东京都
19 东京都
27 东京都
姓名 性别 年龄 住址
高津露儿 女 19 东京都
五十岚美羽 女 21 东京都
宫野理纱 女 23 东京都
山本守 男 27 东京都
男 48 东京都

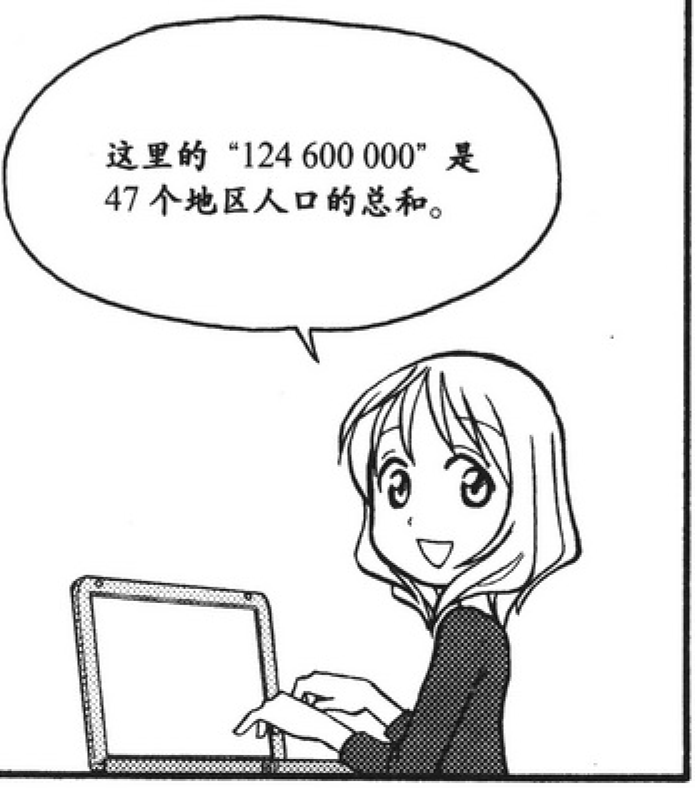

③ 二阶抽样法
“二阶抽样法”就是分2个阶段抽取数据的方法！
两个阶段？
拿我们现在做的例子来说，作为第1阶段，按照出生地原则进行抽取，
实际上，除出生地外，其他属性也可以，
在第2阶段，对已抽出的每一区域进行简单随机抽样，
嗯嗯！
现在来看看它的详细过程吧！
好的！

	人口	累积人口	累积人口下限值	累积人口上限值
北海道	5 600 000	5 600 000	1	5 600 000
青森县	1 400 000	7 000 000	5 600 001	7 000 000
岩手县	1 400 000	8 400 000	7 000 001	8 400 000
宫城县	2 300 000	10 700 000	8 400 001	10 700 000
秋田县	1 100 000	11 800 000	10 700 001	11 800 000
山形县	1 200 000	13 000 000	11 800 001	13 000 000
福岛县	2 100 000	15 100 000	13 000 001	15 100 000
茨城县	2 900 000	18 000 000	15 100 001	18 000 000
栃木县	2 000 000	20 000 000	18 000 001	20 000 000
群马县	2 000 000	22 000 000	20 000 001	22 000 000
埼玉县	6 900 000	28 900 000	22 000 001	28 900 000
千叶县	5 900 000	34 800 000	28 900 001	34 800 000
东京都	12 000 000	46 800 000	34 800 001	46 800 000
神奈川县	8 400 000	55 200 000	46 800 001	55 200 000
新泻县	2 400 000	57 600 000	55 200 001	57 600 000
富山县	1 100 000	58 700 000	57 600 001	58 700 000
石川县	1 100 000	59 800 000	58 700 001	59 800 000
福井县	800 000	60 600 000	59 800 001	60 600 000
山梨县	800 000	61 400 000	60 600 001	61 400 000
长野县	2 200 000	63 600 000	61 400 001	63 600 000
岐阜县	2 100 000	65 700 000	63 600 001	65 700 000
静冈县	3 700 000	69 400 000	65 700 001	69 400 000
爱知县	7 000 000	76 400 000	69 400 001	76 400 000
三重县	1 800 000	78 200 000	76 400 001	78 200 000
滋贺县	1 300 000	79 500 000	78 200 001	79 500 000
京都府	2 600 000	82 100 000	79 500 001	82 100 000
大阪府	8 700 000	90 800 000	82 100 001	90 800 000
兵库县	5 500 000	96 300 000	90 800 001	96 300 000
奈良县	1 400 000	97 700 000	96 300 001	97 700 000
和歌山县	1 000 000	98 700 000	97 700 001	98 700 000
鸟取县	600 000	99 300 000	98 700 001	99 300 000
岛根县	700 000	100 000 000	99 300 001	100 000 000
冈山县	1 900 000	101 900 000	100 000 001	101 900 000
广岛县	2 800 000	104 700 000	101 900 001	104 700 000
山口县	1 500 000	106 200 000	104 700 001	106 200 000
德岛县	800 000	107 000 000	106 200 001	107 000 000
香川县	1 000 000	108 000 000	107 000 001	108 000 000
爱媛县	1 400 000	109 400 000	108 000 001	109 400 000
高知县	800 000	110 200 000	109 400 001	110 200 000
福冈县	5 000 000	115 200 000	110 200 001	115 200 000
佐贺县	800 000	116 000 000	115 200 001	116 000 000
长崎县	1 500 000	117 500 000	116 000 001	117 500 000
熊本县	1 800 000	119 300 000	117 500 001	119 300 000
大分县	1 200 000	120 500 000	119 300 001	120 500 000
宫崎县	1 100 000	121 600 000	120 500 001	121 600 000
鹿儿岛县	1 700 000	123 300 000	121 600 001	123 300 000
冲绳县	1 300 000	124 600 000	123 300 001	124 600 000
总计	124 600 000	————	————	————

（以上数据参考2006年度日本总务省统计局资料制成）

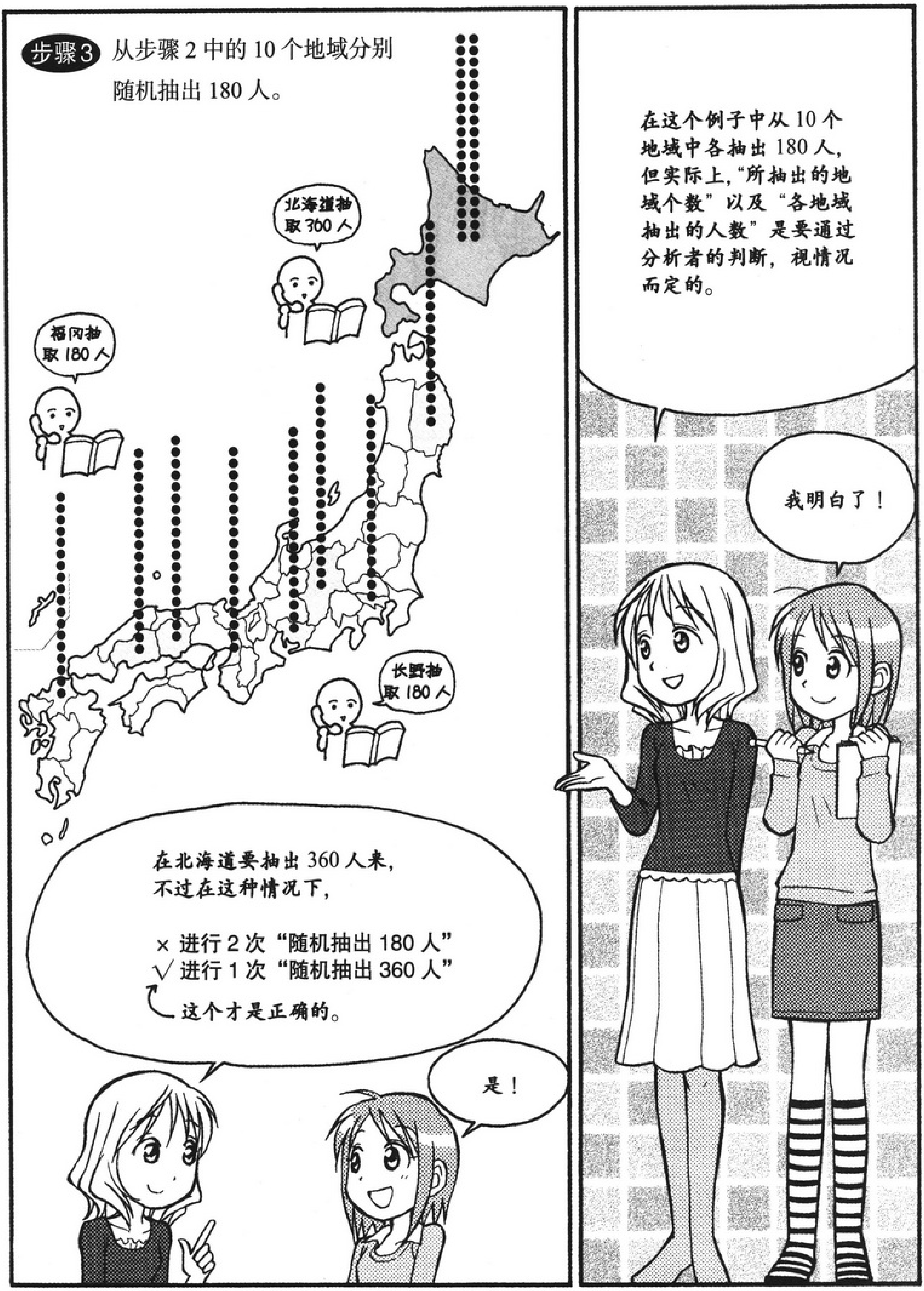
步骤1 使用 Excel 中的“Rand”等函数，在 1 到 124 600 000 中求出 10 个随机数。

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

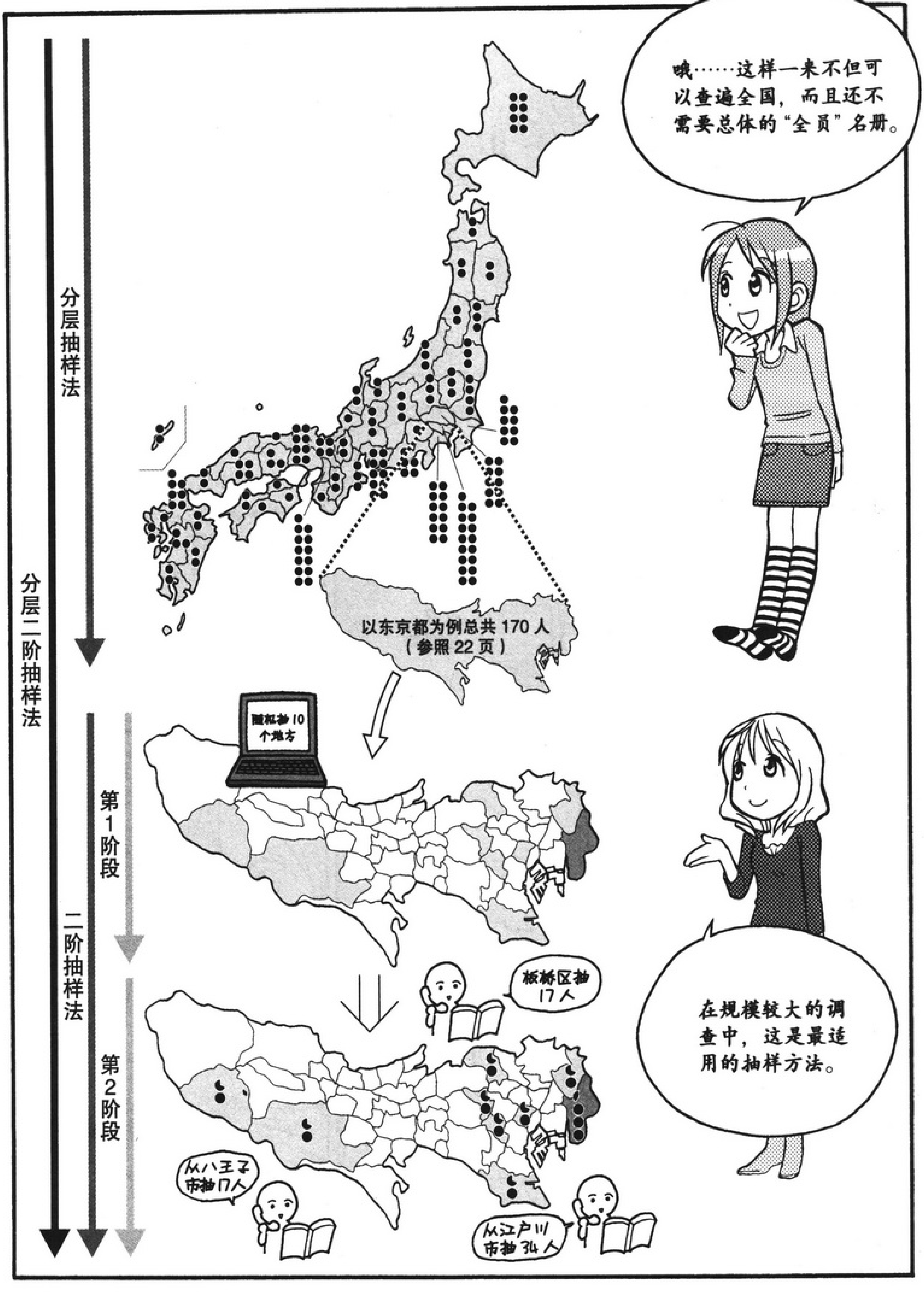


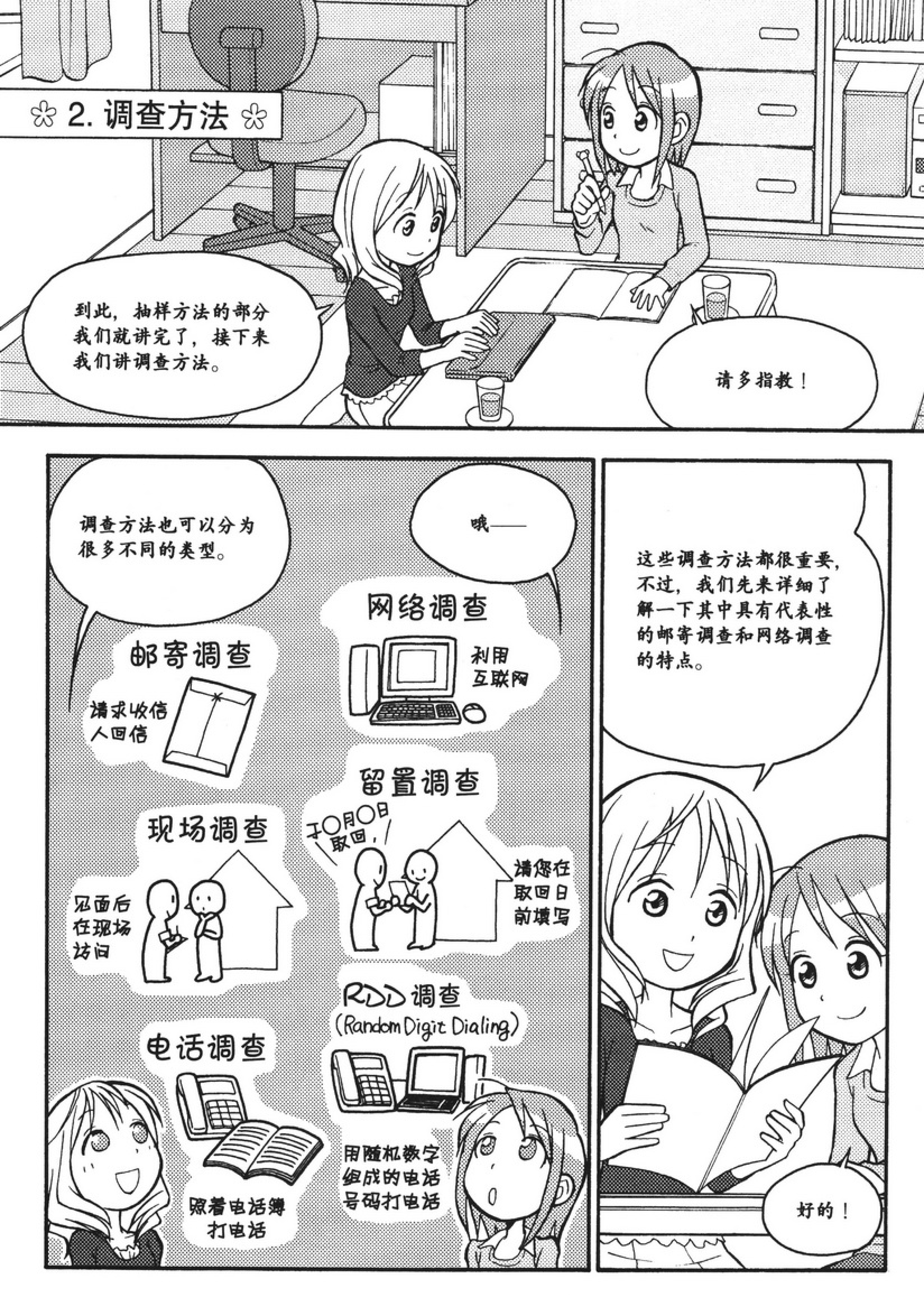
步骤2 步骤 1 中求出的随机数将会介于某个地域的“累积人口上限值”和“累积人口下限值”之间，找出这一区域。

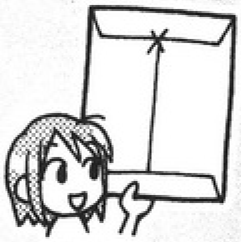
随机数1	104 333 307	→	广岛县
随机数2	8 007 588	→	岩手县
随机数3	35 224 073	→	东京都
随机数4	72 352 247	→	爱知县
随机数5	3 951 586	→	北海道
随机数6	114 308 209	→	福冈县
随机数7	3 724 893	→	北海道
随机数8	100 701 197	→	冈山县
随机数9	62 591 858	→	长野县
随机数10	89 167 908	→	大阪府

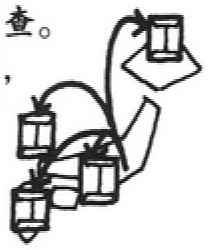

步骤3 从步骤2中的10个地域分别随机抽出180人。
北海道抽取360人
福冈抽取180人
长野抽取180人
在北海道要抽出360人来，不过在这种情况下，
× 进行2次“随机抽出180人”
√ 进行1次“随机抽出360人”
这个才是正确的。
是！
在这个例子中从10个地域中各抽出180人，但实际上，“所抽出的地域个数”以及“各地域抽出的人数”是要通过分析者的判断，视情况而定的。
我明白了！

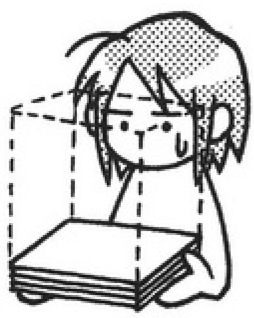
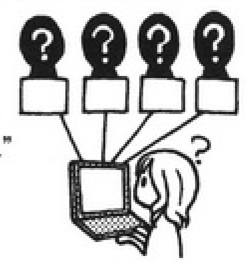
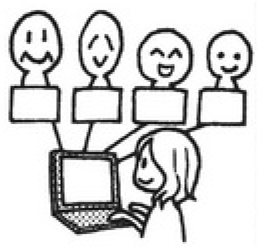

那么，你来比较一下，二阶抽样法同简单随机抽样法或分层抽样法吧！
怎么说呢？
在第1阶段中，只要有被抽取地域的名单就可以了……
哈哈，没错！
但是，还有几个问题……
首先，我们必须知道第1阶段中各地域的准确人数，否则就无法进行调查，是吧？
这是为了步骤1。
此外，在第1阶段中，没有被抽到的地域的人们，他们的意见，将被我们忽视。
对呀，没错！
北陆呢？
四国呢？
④分层二阶抽样法
最后是“分层二阶抽样法”！
分层二阶抽样法
顾名思义，就是分层抽样法和二阶抽样法相结合而成，我用图大致地讲解一下。

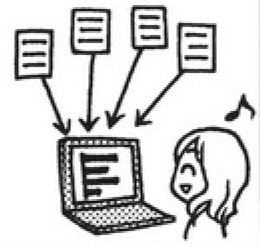

哦……这样一来不但可以查遍全国，而且还不需要总体的“全员”名册。
分层抽样法
分层二阶抽样法
以东京都为例总共 170 人
（参照 22 页）
随机抽 10 个地方
第 1 阶段
二阶抽样法
第 2 阶段
板桥区抽 17人
从八王子市抽17人
从江户川市抽34人
在规模较大的调查中，这是最适用的抽样方法。

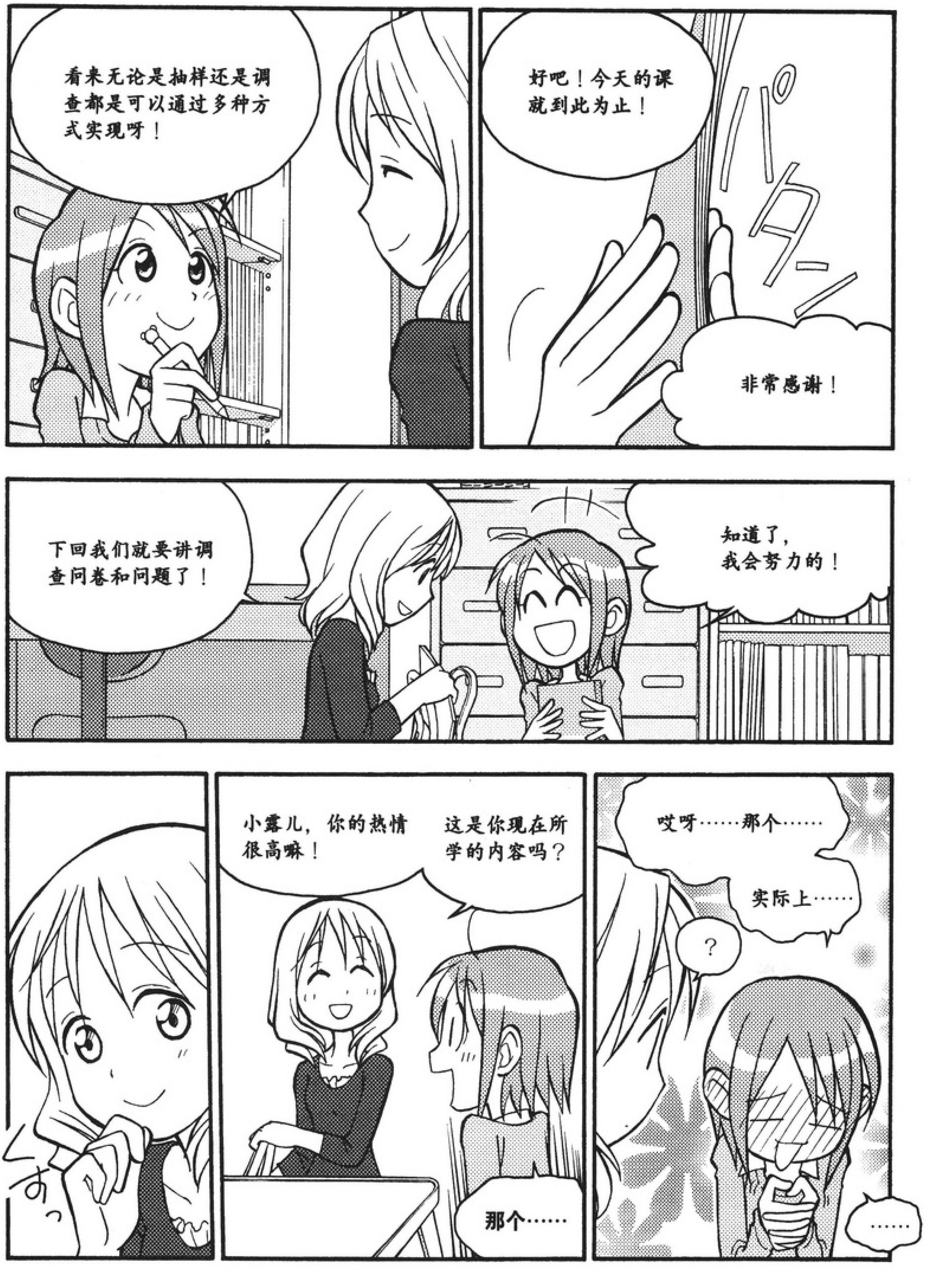

2. 调查方法
到此，抽样方法的部分我们就讲完了，接下来我们讲调查方法。
请多指教！
调查方法也可以分为很多不同的类型。
哦——
网络调查
利用互联网
邮寄调查
请求收信人回信
留置调查
于○月○日取回，
请您在取回日前填写
现场调查
见面后在现场访问
RDD调查
(Random Digit Dialing)
电话调查
照着电话簿打电话
用随机数字组成的电话号码打电话
这些调查方法都很重要，不过，我们先来详细了解一下其中具有代表性的邮寄调查和网络调查的特点。
好的！

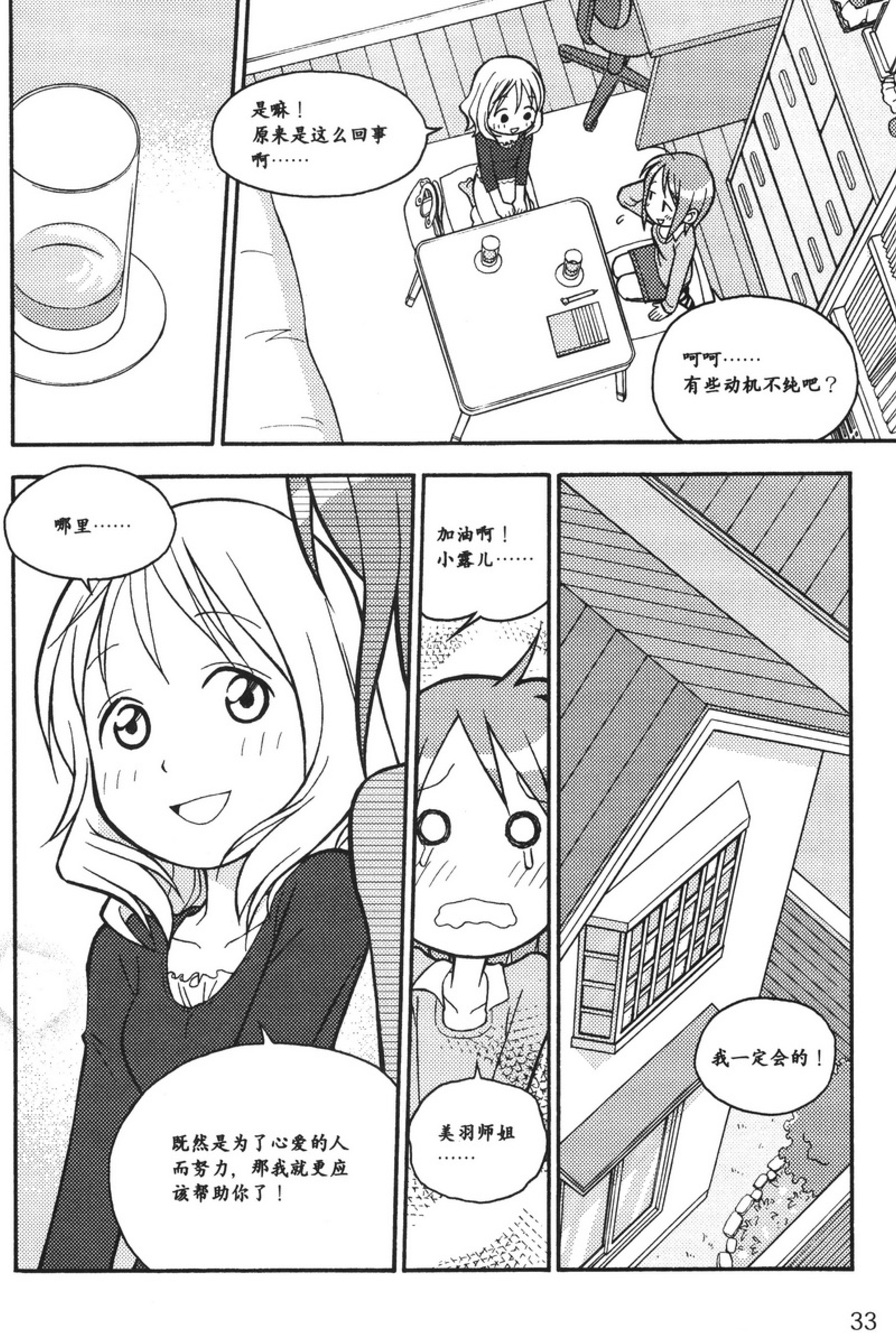
<table>
<tr><td></td><td>
邮寄调查
将调查问卷寄给受访者，请受访者完成问卷后寄回的方法。</td><td>

网络调查
开放型
不限定参与人数，通过互联网来收集访问结果的方法。
监控型
雇佣监控员（monitor），通过互联网来收集访问结果的方法。</td></tr>
<tr><td>地域和名单</td><td>能够广泛地对各地人群进行调查。为了使调查问卷可以准确寄出，必须掌握名单中人员的地址等信息。
</td><td>能够广泛而全面地对各地人群进行调查。受访者自愿参与，无需掌握受访者地址等信息。
</td></tr>
<tr><td>访问量</td><td>由于受访者无须在固定的连续时间内完成，所以可以根据各自的情况决定回答的时间。如此一来，便可以设计较多的问题。
</td><td>受访者回答问题时需要一气呵成，基本上不允许中断。所以，不能设计过多的问题。

</td></tr>
<tr><td>回答的可信度</td><td>通常，只用20%~30%的调查问卷能够得以回收。虽然分析者的初衷是建立尽可能完全反映总体特征的样本，从而精心挑选受访者，并寄出调查问卷的，但是寄回问卷的构成比，却不能保证该样本完全地反映了总体特征。
</td><td>开放型
由受访者单方面决定是否参与，所以分析者在“总体”和“样本”的选择上进行设定。

监控型
虽然样本是经过监控员的选择而产生的，但并不能说这些样本完全反映了总体特征。由于监控员在进行抽样时总会带有自己的偏好，所以很难说被抽取样本可以代表“普通”人群的总体情况。
</td></tr>
<tr><td>数据收集时间</td><td>较长。</td><td>较短。</td></tr>
<tr><td>数据的录入</td><td>由分析者（或其他人员）录入。

</td><td>由受访者录入。也就是说，受访者在回答的同时便将数据自动录入，因此省去了“录入”这一工序。
</td></tr>
</table>


看来无论是抽样还是调查都是可以通过多种方式实现呀！
好吧！今天的课就到此为止！
パタン
非常感谢！
下回我们就要讲调查问卷和问题了！
知道了，我会努力的！
くすっ
小露儿，你的热情很高嘛！
这是你现在所学的内容吗？
那个……
哎呀……那个……
实际上……
？
……


是嘛！
原来是这么回事啊……
呵呵……
有些动机不纯吧？
哪里……
既然是为了心爱的人而努力，那我就更应该帮助你了！
加油啊！
小露儿……
美羽师姐……
我一定会的！

✿ 3. 样本容量的标准 ✿

样本中的个体数目称为样本容量。

样本容量越大，就越接近总体的数量，但不会超过总体的数量。因此，想获得大一些的样本容量，就要尽可能多地收集数据。

话虽如此，但数据的收集时间和资金预算也是我们要考虑的现实问题，所以扩大样本容量也不是一件容易的事。这样一来，我们便产生了这样一种想法："只要对样本容量达到 ×× 程度的数据进行分析，就可以得到理想的结果，即可以较好地代表总体数据的特征"。也就是说，我们期待着存在一个"统计学最低标准"的值，但遗憾的是并没有这样的值存在。

"样本容量的统计学最低标准"是不存在的。但是，在问卷调查领域中，通常都会有一个"约 400"的最低标准,不过这个标准也不能说是合理的。"约 400"只是"出于某种考虑"而设定的值，并不能无条件地令人信服。但有不少人相信存在着这样一个"虽不确切，但是在统计学上仍然能够值得信赖的值"。那么，这个"约 400"是怎样得出来的呢，我们简要地在下面做些解释。同时也给出注意事项。

A 报社想在一个月后做如下问卷调查。

Q. 您支持△△内阁吗？（只能选择其中一项）

1. 支持	2. 不支持

为了避免问题复杂化，我们只做简单说明。实际上，统计学中在进行问卷调查之前，有这样的前提存在。

作为总体的"全体选民"对内阁的支持率为 P，虽然无法通过统计学的手段得出其具体的结果，但这一结果一定会出现在下面这一范围内：

$$p-1.96\times\sqrt{\frac{p\times(1-p)}{n}}<P<p+1.96\times\sqrt{\frac{p\times(1-p)}{n}}$$

※ P 为总体对内阁的支持率，p 为样本对内阁的支持率，n 为样本容量。

我们一般假定这个预测的置信度是 95%。顺便说一下，所谓置信度就是“上面方框中内容的可信性”，又称置信水平或置信系数。

请再观察一下之前方框中的内容。

$1.96\times\sqrt{\dfrac{p\times(1-p)}{n}}$ 的值越小，P 的范围就越窄，说服力也就越强。因此，

> · 对于有说服力的结果，$1.96\times\sqrt{\dfrac{p\times(1-p)}{n}}$ 的值应为 p 的 $\dfrac{1}{10}$ 以下。
>
> · 当处于必须再做一次问卷调查才能得出结论的情况时，p 的值恰好在 0 和 1 的正中间 0.5 处，即 $p=0.5$。

如果无条件地接受上述说法，那么就可以推导出以下结果：

$$1.96\times\sqrt{\frac{0.5\times(1-0.5)}{n}}\leqslant 0.5\times\frac{1}{10}$$

$$1.96\times\sqrt{\frac{0.5\times(1-0.5)}{n}}\leqslant 0.05$$

$$\frac{1.96}{0.05}\times\sqrt{\frac{0.5\times(1-0.5)}{n}}\leqslant 1$$

$$\left(\frac{1.96}{0.05}\right)^2\times\frac{0.5\times(1-0.5)}{n}\leqslant 1^2$$

$$\left(\frac{1.96}{0.05}\right)^2\times 0.5\times(1-0.5)\leqslant n$$

$$384.2\leqslant n$$

这个 384.2，就是之前所说的“约 400”。

对于这个“约 400”，需要注意以下 4 点：

第一，不要抱有这种想法:“只要收集约 400 份数据就可以得到一个准确的结果”。例如,就“对于宪法第 9 条持什么看法？”这个问题,调查“400 名 A 报纸的购买者”，所得出的结论也许可以确定 A 报购买者的整体意见，但是其结果却无法反映全体国民的意愿。

第二,不要抱有这种想法:“如果数量不足400就不能得到足够可信的结果”。例如，就“对于宪法第 9 条持什么看法？”这个问题，如果是对“A 报纸的 400 名购买者”进行访问，反而不如对“A 报纸、B 报纸、C 报纸、D 报纸、E 报纸各 50 人”进行调查所得的结果更能说明问题。

第三，我们再回头去看“约 400”的推导过程。当只有一个问题“您支持 ×× 议案吗？”以及两个选项“支持”、“不支持”时，可知，所谓的“约 400”是在“仅有 1 个 2 选项问题”的情况下，所得出的结论，但如果是“10 个 5 选项问题”就不能得出这个结论了。即这个值并不是在所有情况下都适用。

第四，读者阅读时可能会觉得奇怪，前一页的方框内怎么会有打阴影的地方。实际上，阴影部分的取值也不是绝对的，分析者可以根据自己的情况进行定义。比如，不用“$\frac{1}{10}$以下”改用“$\frac{1}{50}$以下”、不用“p=0.5”改用“p=0.273”,全由分析者自己掌握。这样一来，由于阴影部分的定义不同，就会导致计算结果和所谓的“约 400”产生大不相同的情况。

✿ 4. 随机抽样和定向抽样 ✿

抽样方法大致可以分为随机抽样和定向抽样两种。随机抽样，就是将构成样本的个体从总体中等概率地抽取出来的方法。第 19 页至第 29 页所讲过的：

· 简单随机抽样法
· 分层抽样法
· 二阶抽样法
· 分层二阶抽样法

这些都是抽样的一种。定向抽样，就是随机抽样以外的抽样方法，也就是将构成样本的个体从总体中非等概率地抽取出来的方法。

下表中列出了几种定向抽样法。

介绍法[1]	以熟人或朋友为调查对象，获取样本的方法。
征召法	以读者意见反馈卡等方式招募调查对象，获取样本的方法。
拦截法	在商业街或街角等处寻找调查对象，获取样本的方法。

通过定向抽样法获得的样本，一定不会是“总体的精确缩影”。这样说，有些读者或许会认为定向抽样法是一种不太好的方法，但事实并非如此，其原因在下一节中再做说明。

1. 又称**机缘法**。以熟人或朋友为调查对象，再由他们不断地介绍新的调查对象，我们将其称为**滚雪球法**。

✿ 5. 定量调查和定性调查 ✿

至此，我们已经讨论了许多不同的抽样方法和调查方法之间的区别，但是在此之前，最初的调查大致被分为**定量调查**和**定性调查**两种，如下图所示。

◆图1.1　调查、抽样方法以及调查方法

所谓**定量分析**，大体上讲，就是以“问卷调查收集的数据”或“官方的统计数据”为基础对事物进行考察的调查方法。在本书的讲解中，除了本节以外，均为定量调查。所谓**定性调查**，大体上讲，就是以少数人作为调查对象，就是人们常说的“采访”。

定量分析具有以下优点：

· 所得结果客观

· 结果具备广泛适用性

· 重复率高

另一方面，它也难免存在缺点，如不能从受访者处得到更深层面的信息。也就是说，所获得的总体信息也只能达到“情况大致如此”的程度。我们再来看**定性分析**，它和定量分析的性质刚好相反。总的来说，由于是采访的缘故，所以就可以获得受访者更深层面的信息，这也正是定性分析的优点。但另一方面，它还存在如下缺点：

· 所得结果不够客观
· 结果不具备广泛适用性
· 重复率不高

如此说来，你可能会觉得定性分析不是一个好的分析方法。但是，千万不要这样认为。假如，您是某个公司的一名职员，那么请先设想一下，以下两项关于该公司产品的调查。

调查 1	随机抽取 1000 人，通过调查问卷，征求其对现有产品的意见。（注意，这 1000 人中还包含很多并不认为本公司产品优于其他同类产品的人）
调查 2	找来 10 位主动要求为本公司产品进行评价的用户，进行圆桌会议，对现有产品进行一次彻底的调查访问。

前者就是“通过随机抽样法进行的定量调查”，而后者则是“通过定向抽样法进行的定性调查”。您觉得怎么样？当然，我们还要看调查的目的是什么，不能就表面情况一概而论。但是，我们也不能盲目地断定，前者就是一个好的调查方法，后者也有其实际的应用价值。

并且，就定量调查而言，对使用随机抽样法获得的样本进行分析自然是无可厚非的，不过对使用定向抽样法获得的样本进行分析也是可以的。就定性分析而言，绝大多数情况下，都是对使用定向抽样法获得的样本进行分析，但是话虽如此，对随机抽样法获得的样本进行定性分析也不是不可以的。

✿ 6. 数据分析的搭配方法 ✿

以下要讲的内容也是相当重要的，但是同之前所讲的内容有所不同，请调整思路再做阅读。

数据分析的搭配方法有两种类型："探索型"和"验证型"。

"探索型"的数据分析流程

① 收集手头资料。
② 试着运用各种分析方法，进行全面分析。
③ 如"事后诸葛"般恍然大悟"原来世上还有这样的事"。
④ 向周围的人公布自己的分析结果。

"验证型"的数据分析流程

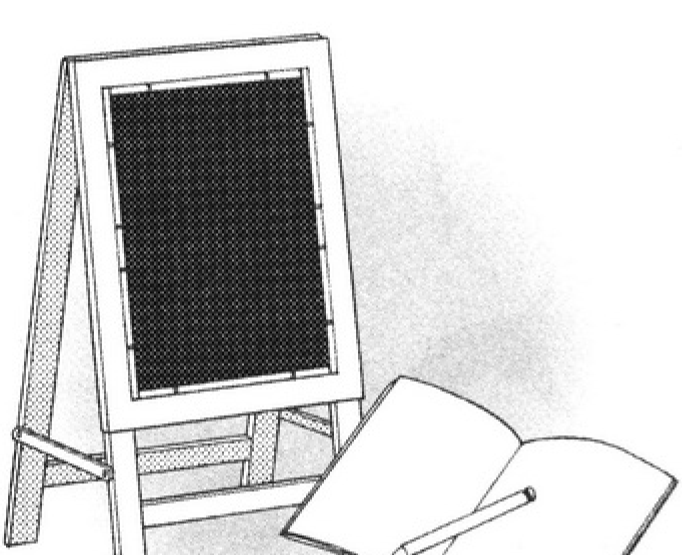
① 建立假设。
② 为了确认假设是否成立，收集资料，并进行分析。
③ 得出结论，即假设成立与否。
④ 向周围的人公布自己的分析结果。

"探索型"的数据分析，只需要手头的数据即可，其优点在于能够简单、快速地完成分析。另一方面，它也存在弊端，如数据被随意加工、变量间被强加上想当然的因果关系，从而可以让"事后诸葛"们乱说一气。更有甚者，便会"为所欲为"了。因此，纵然是费了一番工夫得来的分析结果，周围的人还是有可能会觉得这一结果缺乏说服力。

"验证型"的数据分析，一定要在调查的最初阶段建立假设，所以着手分析时会比较困难，这也正是它的缺点。但另一方面，假设一旦建立，接下来便是收集数据和分析数据。经过一番分析，如果假设成立，就可以得到一个非常有说服力的结果，获得周围人的认同。即便假设不成立，我们同样可以得到一个事实"至少这个假设不成立"，这对指导今后的研究是很有帮助的，所以绝不能说这是一个没有意义的数据分析。

第 2 章

调查问卷和问题

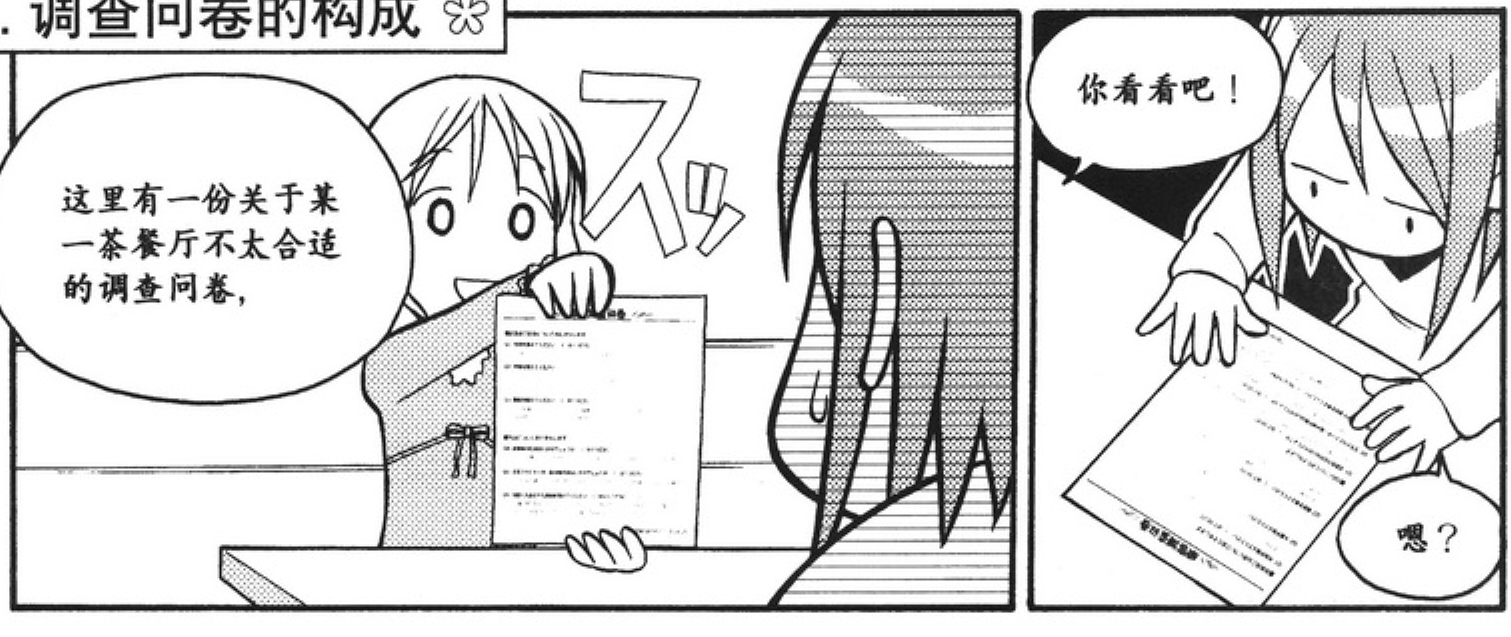


1. 调查问卷的构成
2. 问题的分类
3. 应当避免的问题
4. 应当避免的问题（续）
5. “中值”的存在性

我走啦！
ノバタ
咦？
ムグ！
虽然见不到山本，
可露儿还是很精神
啊！
不要再提
那家伙！
露儿还是个学生，
她的任务是踏踏
实实地完成学业。
是啊，是
啊，没错！
快把睡
衣换下
来吧！
Cafe
mimir


今天我们要讲调查问卷和问题！
请您多多指教……
ハハ———
这里的蛋糕真好吃啊！
喂！调查调查
キャッ
キャッ
无论是调查问卷的构成还是询问的问题都没有一个所谓的“最佳法则”，但是……
还是有一些规则能使问卷和问题的设计更科学合理一些。
噢！

1. 调查问卷的构成


这里有一份关于某一茶餐厅不太合适的调查问卷，
スッ
你看看吧！
嗯？

顾客调查问卷

■ 顾客情况调查

Q1. 请问您的性别是（只能选 1 项）

1. 男性	2. 女性

Q2. 请问您的年龄是

☐ 岁

Q3. 请问您的职业是（只能选 1 项）

1. 职员	2. 个体经营	3. 学生
4. 主妇	5. 其他（	）

■ 关于本店的问题

Q4. 您觉得工作人员的态度如何？（只能选 1 项）

1. 很差	2. 差	3. 一般	4. 好	5. 很好

Q5. 您对所点的蛋糕及饮品的口味是否满意？（只能选 1 项）

1. 很不满意	2. 不满意	3. 一般	4. 满意	5. 很满意

Q6. 您为何光临本店？（可选多项）

1. 在杂志或广告中见过	2. 访问过主页	3. 听朋友或熟人介绍
4. 碰巧路过	5. 喜欢店面的设计	6. 其他（ ）

十分感谢您的合作！

顾客调查问卷

■ 关于本店的问题

Q1. 您觉得工作人员的态度如何？（只能选 1 项）

1. 很差	2. 差	3. 一般	4. 好	5. 很好

Q2. 您对所点的蛋糕及饮品的口味是否满意？（只能选 1 项）

1. 很不满意	2. 不满意	3. 一般	4. 满意	5. 很满意

Q3. 您为何光临本店？（可选多项）

1. 在杂志或广告中见过	2. 访问过主页	3. 听朋友或熟人介绍
4. 碰巧路过	5. 喜欢店面的设计	6. 其他（　　　　　　）

■ 顾客情况调查

Q4. 请问您的性别是（只能选 1 项）

1. 男性	2. 女性

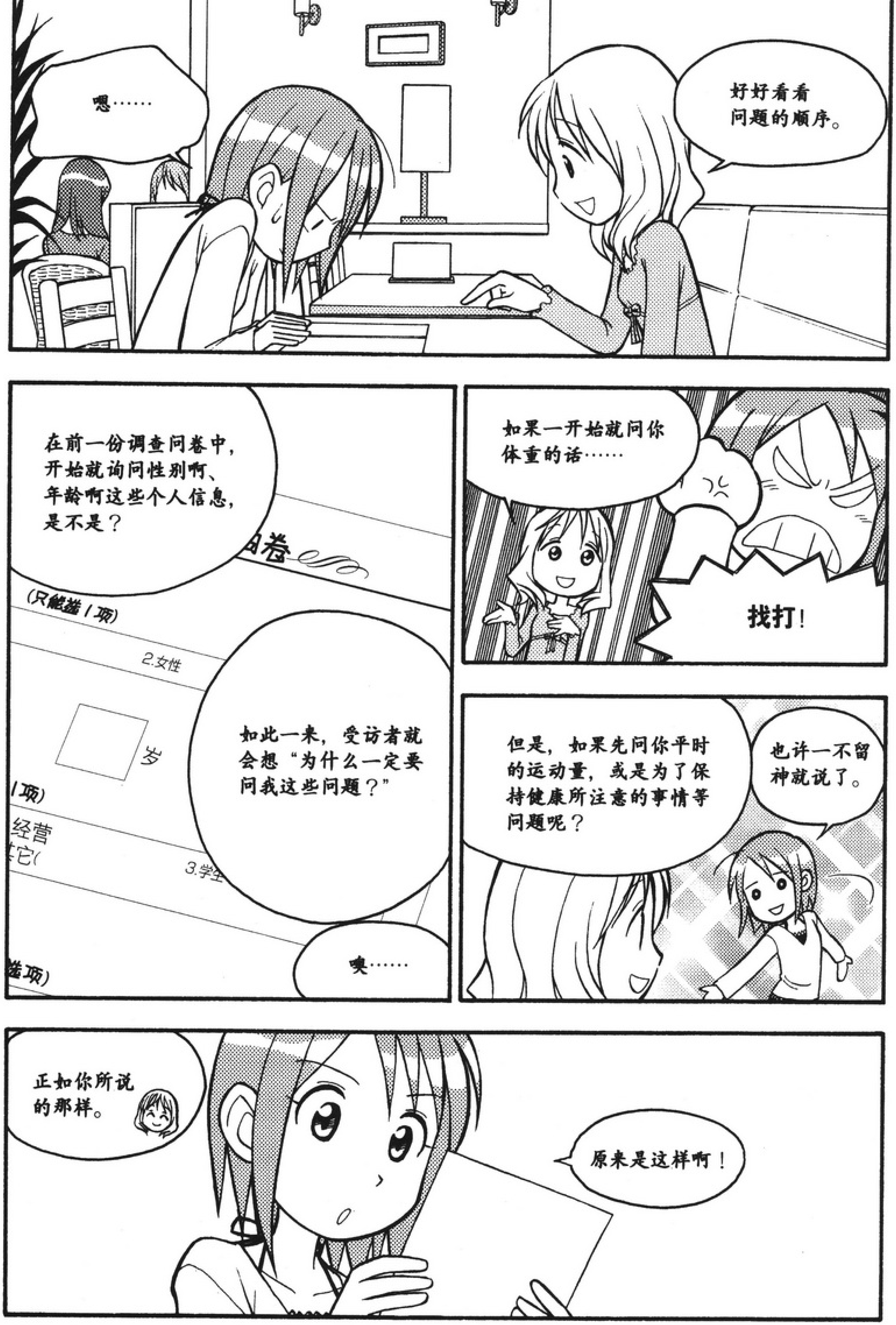
Q5. 请问您的年龄是

☐ 岁

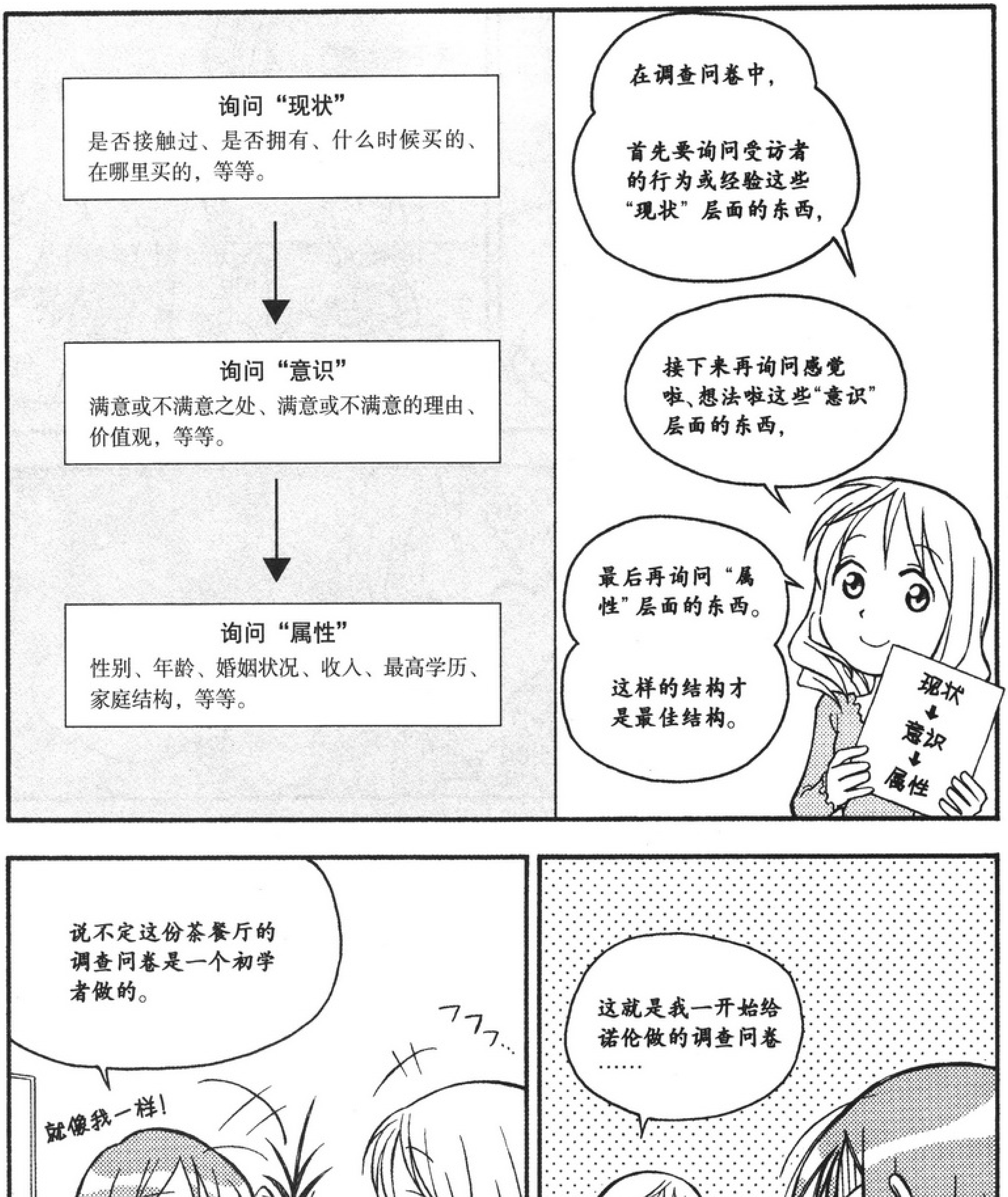
Q6. 请问您的职业是（只能选 1 项）

1. 职员	2. 个体经营	3. 学生
4. 主妇	5. 其他（　　　　　　　　　　）	

十分感谢您的合作！



嗯……
好好看看
问题的顺序。
在前一份调查问卷中，开始就询问性别啊、年龄啊这些个人信息，是不是？
(只能选1项)
2.女性
岁
如此一来，受访者就会想“为什么一定要问我这些问题？”
噢……
如果一开始就问你体重的话……
找打！
但是，如果先问你平时的运动量，或是为了保持健康所注意的事情等问题呢？
也许一不留神就说了。
正如你所说的那样。
原来是这样啊！



询问“现状”
是否接触过、是否拥有、什么时候买的、在哪里买的，等等。
询问“意识”
满意或不满意之处、满意或不满意的理由、价值观，等等。
询问“属性”
性别、年龄、婚姻状况、收入、最高学历、家庭结构，等等。
在调查问卷中，
首先要询问受访者的行为或经验这些“现状”层面的东西，
接下来再询问感觉啦、想法啦这些“意识”层面的东西，
最后再询问“属性”层面的东西。
这样的结构才是最佳结构。
现状
意识
属性
说不定这份茶餐厅的调查问卷是一个初学者做的。
就像我一样！
フフフ…
这就是我一开始给诺伦做的调查问卷……


2. 问题的分类
真不好意思啊！
没关系，没关系！今天我请客！
那我就不客气了……
接下来我们来讲讲调查问卷中所提的问题。
好的。
调查问卷中的问题大致可分为“单项选择题”、“多项选择题”、“数量题”、“文字题”等。
这样啊。
下面举个例子介绍一下。

单项选择题

首先是单项选择题。
也就是只允许选择一个选项的问题。

Q. 请问下列哪种蛋糕是您最喜爱的？（限选一项）

1. 油酥蛋糕　　2. 奶酪蛋糕　　3. 巧克力蛋糕　　4. 蒙布朗

Q. 在您选择单人住房时，是否重视与最近车站的距离？（限选一项）

1. 不重视　　2. 不太重视　　3. 一般　　4. 有些重视　　5. 重视

当几个问题的选项相同时，为了节约纸面空间，
我建议采取这种方式。

Q. 在您选择单人住房时，对以下事项的重视程度如何？（限选一项）

	不重视	不太重视	一般	有些重视	重视
a. 距最近车站的距离	1	2	3	4	5
b. 房间朝阳	1	2	3	4	5
c. 储物空间	1	2	3	4	5
d. 房租	1	2	3	4	5

多项选择题

接下来是多项选择题。
也就是同一道题可以选择多个选项的问题。

Q. 请问您在选择单人住房时比较看重哪些条件？
（可选多项）

1. 距最近车站的距离	2. 房间朝阳	3. 储物空间
4. 房租	5. 周边设施	

另外，还有下面这种询问方式，但是并不推荐大家使用。

Q. 请问您在选择单人住房时比较看重哪些条件？
（最多选两项）

1. 距最近车站的距离	2. 房间朝阳	3. 储物空间
4. 房租	5. 周边设施	

请注意！

为什么呢？

与“可选多项”相比，这种设计要求受访者必须先将全部选项浏览一遍之后才能作答，这样无形中会增加受访者的负担。

数量题

接下来是数量题。
也就是需要回答具体数值的问题。
这时要用线将每一位数字隔开，以免书写时出现错误。

Q. 您的家庭月消费平均是多少？

万	千	百	十	个	
					元

文字题

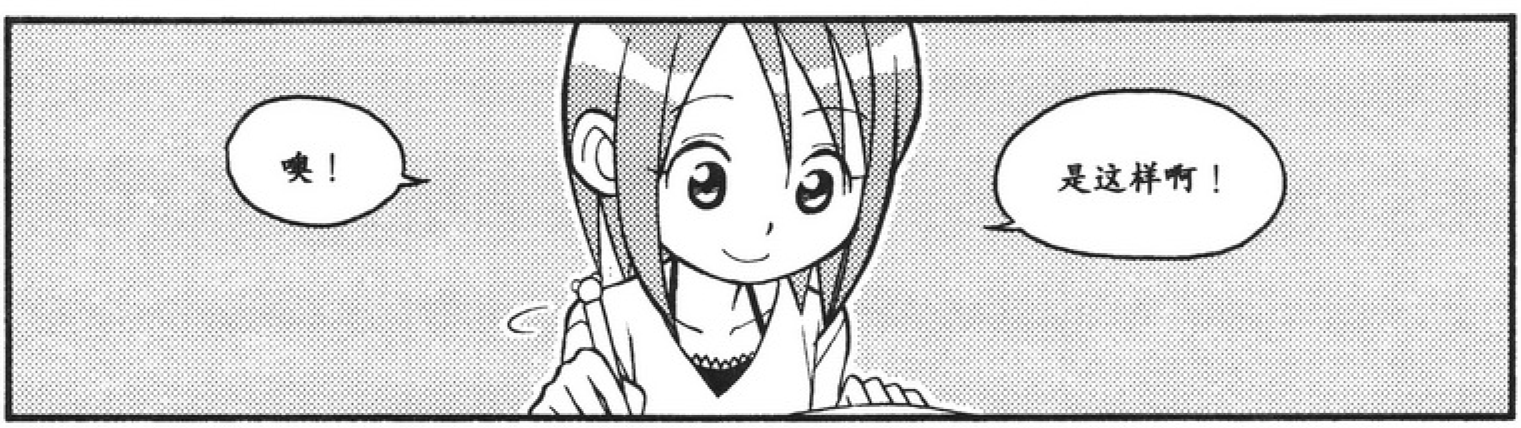
最后是文字题。
也就是不需要选择选项，而是需要自由做答的问题。

Q. 请列举出 1 位您最喜欢的艺人。

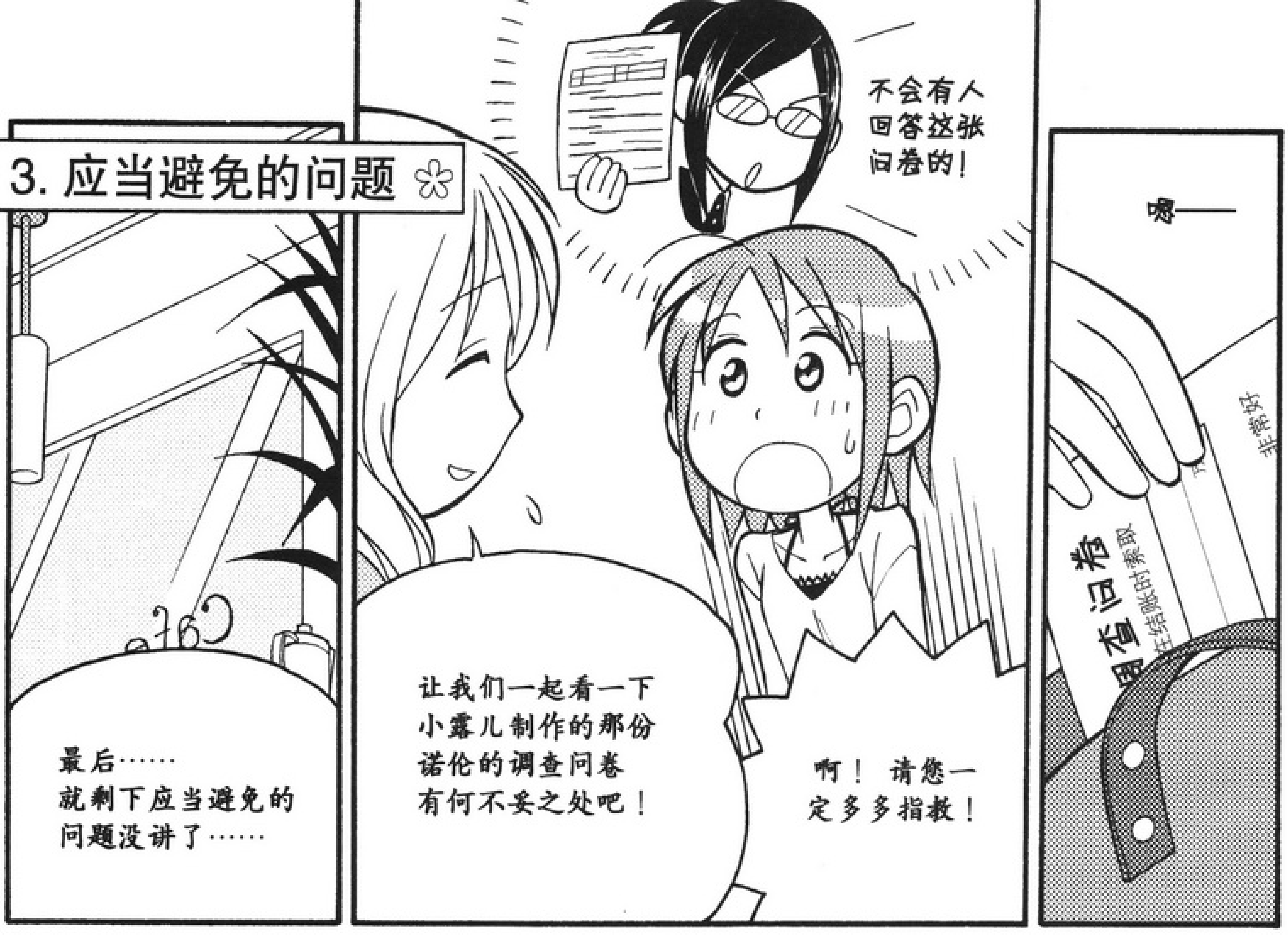
答题栏	

Q. 如果您对本店还有什么意见或要求，请提出来。

（　　　　　　　　　　　　　　　　　　　）



3. 应当避免的问题



诺伦茶餐厅调查问卷

参与此次调查的顾客将获得我们赠送的优惠券，请在结账时索取！

性别	男、女	年龄	岁
职业		年收入	万日元

Q. 您感觉店内的就餐环境如何？

非常不好　不好　一般　好　非常好

Q. 您感觉女服务生的制服和服务态度如何？

非常不好　不好　一般　好　非常好

Q. 您觉得我们的红茶如何？

非常不满意　不满意　一般　满意　非常满意

Q. 您觉得价格如何？

便宜　适中　贵

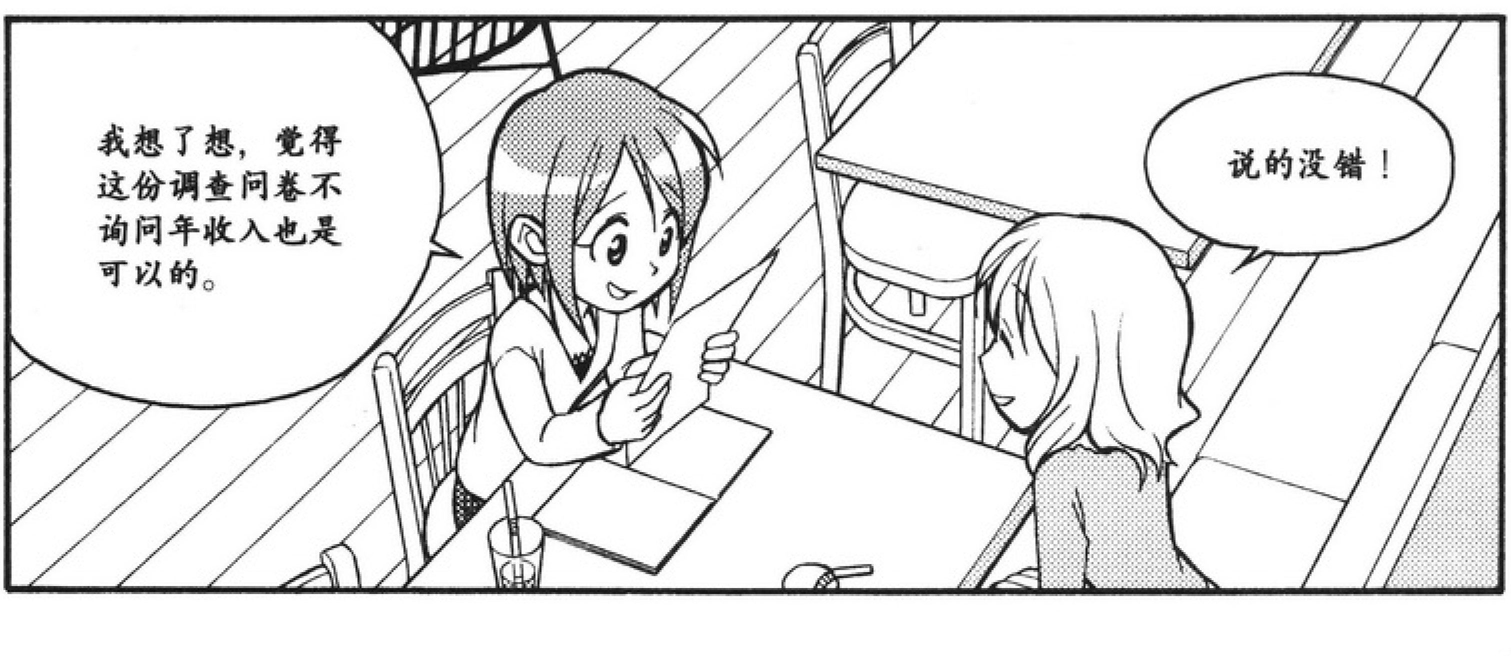
Q. 对于以下 4 种红茶，请您按照自己的喜好程度进行排序。

1. 原味红茶	→	(　　) 位
2. 柠檬茶	→	(　　) 位
3. 奶茶	→	(　　) 位
4. 玫瑰茶	→	(　　) 位

Q. 您喜欢在茶餐厅就餐吗？　喜欢　不喜欢


年龄 岁
年收入 万日元
对过于隐私的问题问得太具体。
与你的调查问卷相比，人们会更忌讳这种年收入的问法。
？

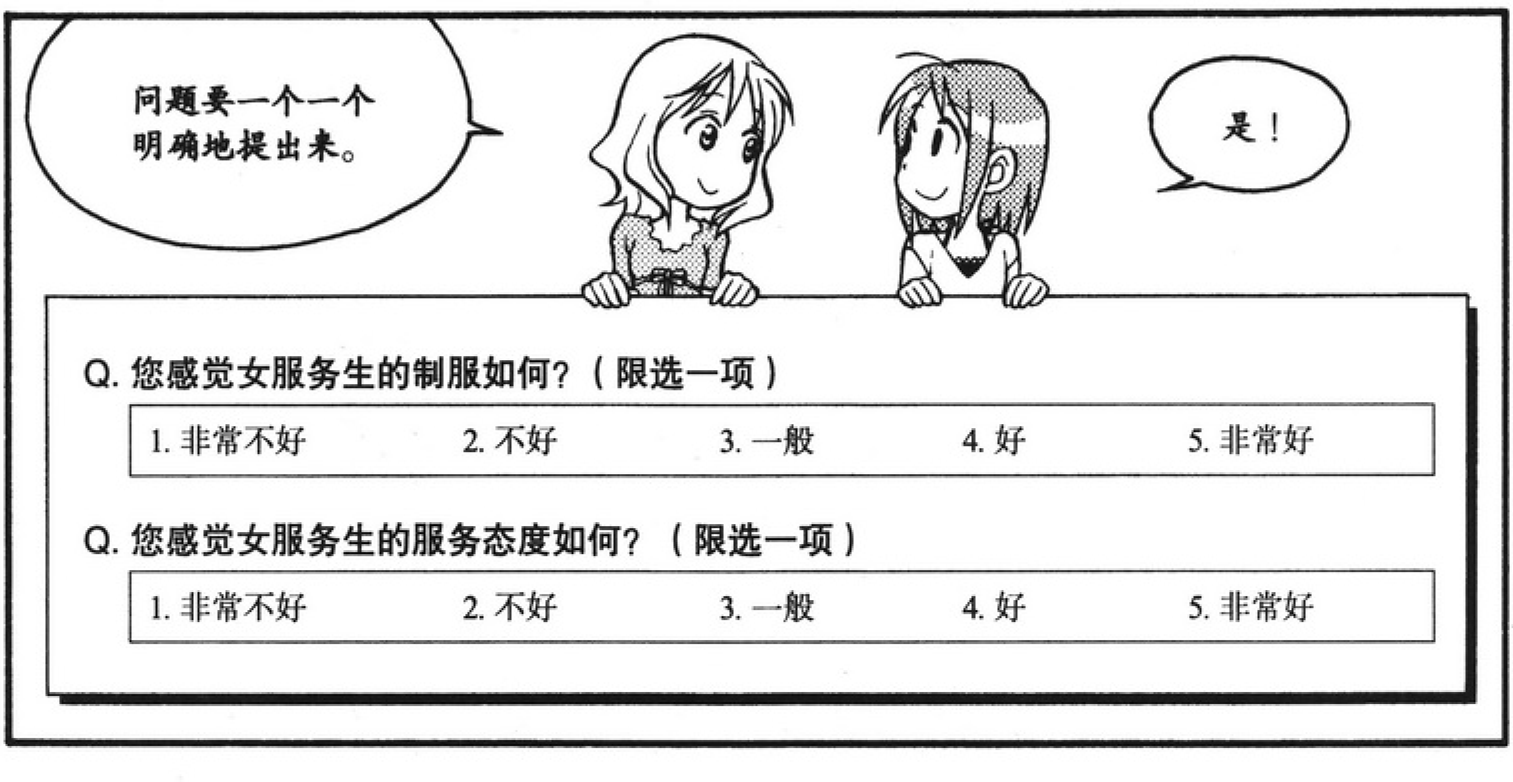

如果一定要询问具体值的话，就要另想办法。
这种方法行不通可以换一种方法进行询问，会让受访者感到更为亲切。
原来如此……
Q. 您的年收入是？（限选一项）
1. 400 万日元及以下
2. 400 万日元以上　600 万日元及以下
3. 600 万日元以上　800 万日元及以下
4. 800 万日元以上　1000 万日元及以下
5. 1000 万日元以上


我想了想，觉得这份调查问卷不询问年收入也是可以的。
说的没错！

Q. 您感觉店内的就餐环境如何？
非常不好 不好 一般 好 非常好
Q. 您感觉女服务生的制服和服务态度如何？
非常不好 不好 一般 好 非常好
Q. 您觉得我们的红茶如何？
非常不满意 不满意 一般 满意 非常满意
Q. 您觉得价格如何？
便宜 适中 贵
存在表达不明确的问题
这里的问题问得有点模糊。
嗯，嗯，确实……
例如“价格如何”这道题，
不知道你指的是红茶的价格呢？还是蛋糕的价格？
这样啊……
还是尽量询问得具体些为好！
我明白了！
Q. 您觉得您所点的红茶价格如何？（限选一项）
1. 便宜 2. 适中 3. 贵

你看这里，这是1个问题，却包含两层意思？
Q. 您感觉女服务生的制服和服务态度如何？
非常不好 不好 一般 好 非常好
同一问题包含两层以上的意思。
啊！

倘若受访者认为“服务态度好，而制服不好”，那他就不知该如何作答了！
嗯，的确存在这种问题。
你这是要寻个什么样子啊！
呜——



问题要一个一个明确地提出来。
是！
Q. 您感觉女服务生的制服如何？（限选一项）
1. 非常不好 2. 不好 3. 一般 4. 好 5. 非常好
Q. 您感觉女服务生的服务态度如何？（限选一项）
1. 非常不好 2. 不好 3. 一般 4. 好 5. 非常好

Q. 对于以下 4 种红茶，请您按照喜好程度进行排序。
1. 原味红茶 → (　　)位
2. 柠檬茶 → (　　)位
3. 奶茶 → (　　)位
4. 玫瑰茶 → (　　)位
排序问题
这种排序的询问方式……
啊？
排序也不可以吗！？
也不是绝对不可以。
那么为什么不可以设计这样的问题呢？
排序基本上是在做“联合分析”（Conjoint Analysis）时才会用到……
比如说，认为“两者同样喜欢”的受访者会不知如何作答？
噢，还真是……
还有，没有尝过“玫瑰茶”的人，也会苦于它的位次问题。
玫瑰茶？
原来如此啊……
嗯，要放到第几位呢……

	原味红茶	柠檬茶	奶茶	玫瑰茶
顾客 A	1	0	1	0
顾客 B	1	1	0	0
顾客 C	1	1	0	0
:	:	:	:	:
	↓	↓	↓	↓
比例（%）	84	65	73	52


如果没有直接的代替方案这样问也是可以的。
哦！
Q1. 请在下列红茶中，选出您所喜爱的种类（可选多项）
1. 原味红茶 2. 柠檬茶 3. 奶茶 4. 玫瑰茶
Q2. 请在下列红茶中，选出您最喜爱的种类（限选一项）
1. 原味红茶 2. 柠檬茶 3. 奶茶 4. 玫瑰茶
至于其他的建议嘛，
要给问题和选项标明序号，
还要添上“(限选一项)”这样的限制性说明。
这样不就便于回答了吗？
性别
职业
男
Q. 您觉得店内的就餐环境如何？（限选一项）
非常不好
不好
一般
好
Q.您感觉女服务生的制服
Q.您觉得我们的红茶如何
年收入
我明白了！
需要注意的地方这么多，还真是麻烦啊！
也没有那么严重。
关键在于要从“受访者角度”出发！
受访者角度……
对！


如果你是受访者的话，做这份调查问卷会有什么感受……
如果你身边的人是受访者的话，他们会协助你完成调查吗……
如果你已经考虑到这些的话，就没问题了。
……对呀！
你这么说我一下子就明白了！
美羽师姐，真的是太感谢了！
呵呵，不用客气！
那——我们就为诺伦调查调查吧 ♡
这个是我的最爱啊!
啊？
还吃啊？
从这到这我都要 ♡
老天……
美羽师姐还真是一个甜食动物啊！?

✿ 4. 应当避免的问题（续）✿

上一节中，我们列举出了以下四类应当避免的问题：

· 对过于隐私的事情问得太具体
· 表达不明确的问题
· 同时含有 2 层以上意思的问题
· 排序的问题

而应当避免的问题还不止这些，这里我们再介绍几个。

■ 诱导回答的问题

Q. 由于日本的资源紧缺，所以在 21 世纪，关于科学技术的教育越发显得重要。因此，您对今后的初中理科教育有何看法？（限选一项）

1. 应当再丰富一些	2. 保持现状就好

面对这种提问方式，很多人都是出于不得已而选择"1. 应当再丰富一些"了。

■ 程度等级太多的问题

Q. 在您选择就业单位时，更看重哪些因素？（每题限选一项）

	极为不重视	非常不重视	不重视	不太重视	一般	有些重视	重视	非常重视	极为重视
a. 企业知名度	1	2	3	4	5	6	7	8	9
b. 充分的入职教育	1	2	3	4	5	6	7	8	9
c. 积极地为新人派发工作	1	2	3	4	5	6	7	8	9
:	:	:	:	:	:	:	:	:	:

或许是调查者觉得"用 5 个程度等级还不足以分清受访者心理的细微差异"，所以才设计出了程度等级如此之多的问题。正是出于这种想法，在实际操作中，才会有人认为等级越多越好。但是如果等级数目不断增加的话，就会适得其反，导致受访者在回答的过程中开始觉得"7 也行、8 也行、多少都行"。

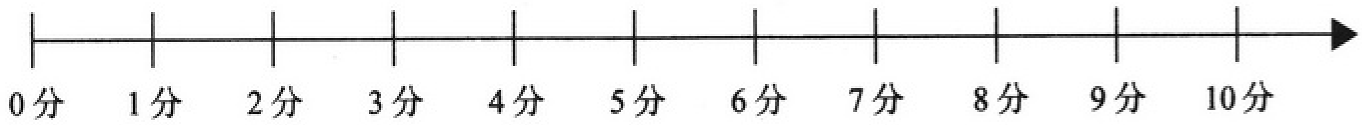
笔者认为等级程度最多有 7 项即可，不过这一观点并未得到学术上的证实，纯属笔者本人的看法。

■ 评分的问题

Q. 请您对这种曲奇饼口味的喜好程度进行评分，从 0 分到 10 分您会给

□ 分

问题中并没有将下图一并给出，而且分数间的间隔也没有定义，所以令受访者很难作答。



此外，再想想，“0 分”要如何解释，是要解释成为“难吃”，还是要解释成“既不好吃也不难吃”，读者在判断时会出现这样的困扰。

■ 征集自由答案的问题

Q. 请您列举出 1 位您最喜爱的艺人。

解答栏	

征集自由答案的问题有如下优点。

・从受访者的角度看，如题所讲可以自由回答。
・从调查者的角度看，可能得到意想不到的、很有趣的答案。

乍一看，优点很明显。但实际情况中，并不能如此随心所欲地进行评价。

首先，请考虑一下“如果你是受访者的话”。如果你是小学生或中学生的话，突然被问到“最喜爱的艺人”，你能立刻回答出来吗？也就是说，当被问到不是那么关心的问题时，你能立刻回答出来吗？其次，请考虑一下“如果你是调查者”，而这是邮寄调查的调查结果。

	A	B
1		最喜爱的艺人
2	受访者1	猪濑幸人
3	受访者2	猪濑幸人
4	受访者3	平河樱
5	受访者4	松山Arisu
6	受访者5	
7	受访者6	

所有的数据录入都需要你自己进行。当拿到千奇百怪的答案时，你就会意识到您所肩负的工作是多么沉重。

先通过一个预备调查征集自由答案，将其中的前 5 名像下图那样做成选项。可以用这种方式进行调查。

Q. 请您列举出 1 位您最喜爱的艺人。（限选一项）

1. 猪濑幸人　2. 松山 Arisu　3. 矢野爱梨　4. 并木和人　5. 平河樱

✿ 5. “中值”的存在性 ✿

在进行程度等级的评价时，无论采用哪种方式，都会有“包含‘中值’”和“不包含‘中值’”两种情况。

■ 包含“中值”

	不重视	不太重视	一般	有些重视	重视
a. 距最近车站的距离	1	2	3	4	5
b. 房间朝阳	1	2	3	4	5

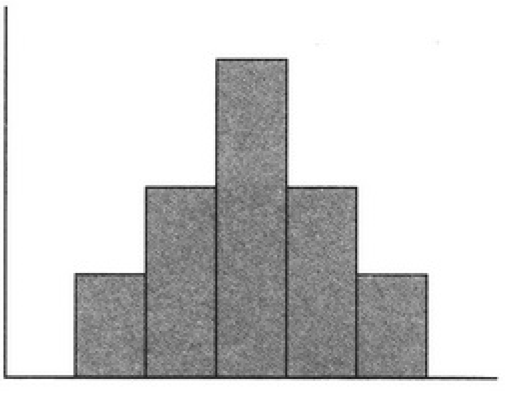
■ 不包含“中值”

	不重视	不太重视	有些重视	重视
a. 距最近车站的距离	1	2	3	4
b. 房间朝阳	1	2	3	4

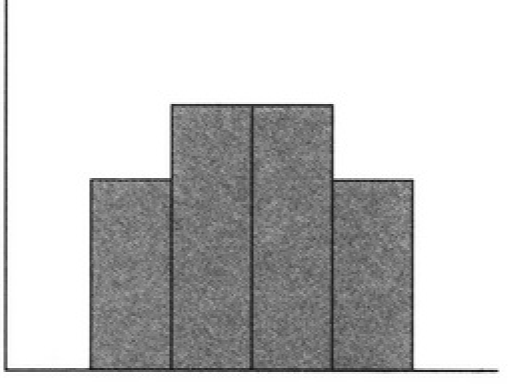
笔者认为包含或不包含“中值”都可以，但是不包含的话会出现这样的问题：

· 必须明确地回答“是”或者“否”，这会增加受访者的回答难度。
· 与含有“中值”的情况相比，其直方图与正态分布的相似度也较低。

包含“中值”



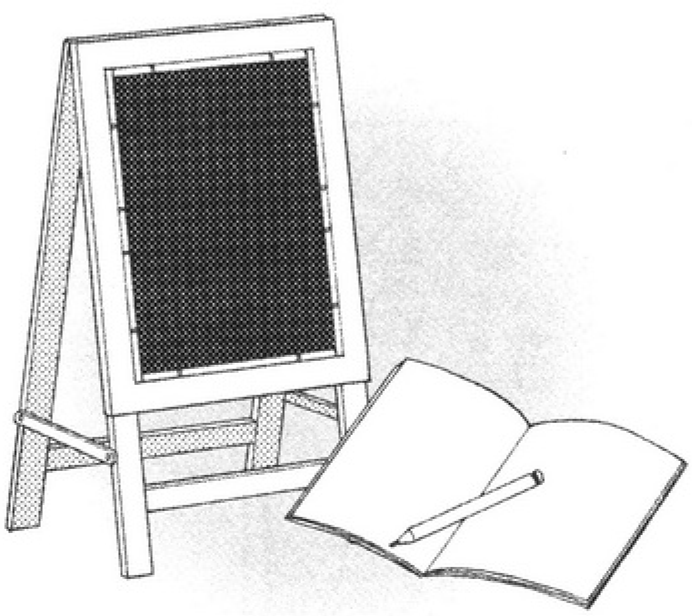
不包含“中值”



如果为不知应当选择哪种方式而苦恼，那么可以以周围的人为对象进行一次预备调查试试看。

第 3 章

基础数学知识



1. 相关矩阵
2. 单位矩阵
3. 旋转
4. 特征值和特征向量
5. 对称矩阵
6. 矩阵的补充
7. 离差平方和、方差、标准差

ピンポーン
您好！
啊！
快请进！快请进！
哇啊！
今天上课，
太好了！！
呵呵，要是
小露儿的话，
……
那么今天，为了理解
因子分析我们有必要
先来学习一下数学的
基础知识。
请您多多指教！
嗯……
我想今天我们先不必理会计
算过程等问题，
只是讲讲一些事实性问题。
权且当作“就是这么
回事”来理解吧！
明白了！

1. 相关矩阵

	语文	社会	理科	英语	数学
学生 1	92	83	77	156	38
学生 2	97	82	68	114	33
学生 3	100	100	93	176	44
学生 4	89	77	100	158	46
学生 5	95	79	75	140	37
学生 6	99	96	84	174	42
学生 7	97	87	98	190	49
学生 8	93	77	73	132	35
学生 9	89	75	72	132	35
学生 10	98	93	70	186	37

单相关系数

	语文	社会	理科	英语	数学
语文	语与语	语与社	语与理	语与英	语与数
社会	社与语	社与社	社与理	社与英	社与数
理科	理与语	理与社	理与理	理与英	理与数
英语	英与语	英与社	英与理	英与英	英与数
数学	数与语	数与社	数与理	数与英	数与数

r_{ij} 和 r_{ji} 的值相同

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

无论 i 为何值，r_{ii} 的值都是 1

2. 单位矩阵

接下来讲“单位矩阵”。
$$\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$
单位矩阵就是这样的矩阵。
主对角线（左上至右下）上的值全为1，其余的值全为0。
同单位矩阵相乘并不会对原来的矩阵产生影响。
为什么会这样呢？
例如，同$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$，相乘……
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}=\begin{bmatrix} 1\times a_1+0\times a_2 \\ 0\times a_1+1\times a_2 \end{bmatrix}=\begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$
瞧！
シャッ
真的！
サラサラ
也就是说这个矩阵像数字里的1一样！
1×7=7
1×A=A
哦……

再看看其他单位矩阵的例子吧！

$$\cdot\ \overset{2行2列}{\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}}\overset{2行1列}{\begin{bmatrix}a_1\\a_2\end{bmatrix}}=\overset{2行1列}{\begin{bmatrix}1\times a_1+0\times a_2\\0\times a_1+1\times a_2\end{bmatrix}}=\begin{bmatrix}a_1\\a_2\end{bmatrix}$$

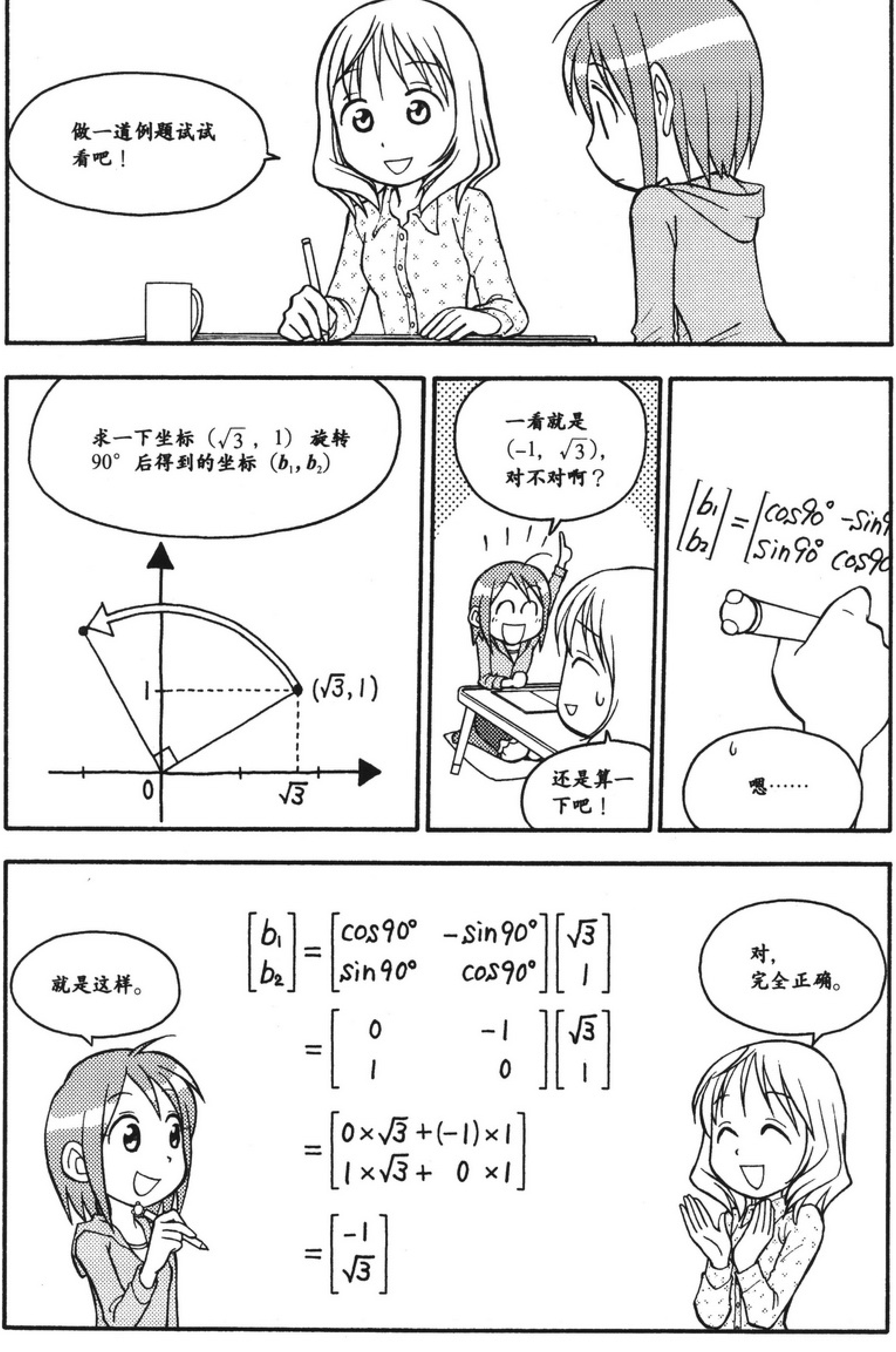
$$\cdot\ \overset{p行p列}{\begin{bmatrix}1 & 0 & \cdots & 0\\0 & 1 & \cdots & 0\\\vdots & \vdots & \ddots & \vdots\\0 & 0 & \cdots & 1\end{bmatrix}}\overset{p行1列}{\begin{bmatrix}a_1\\a_2\\\vdots\\a_p\end{bmatrix}}=\overset{p行1列}{\begin{bmatrix}1\times a_1+0\times a_2+\cdots+0\times a_p\\0\times a_1+1\times a_2+\cdots+0\times a_p\\\vdots\\0\times a_1+0\times a_2+\cdots+1\times a_p\end{bmatrix}}=\begin{bmatrix}a_1\\a_2\\\vdots\\a_p\end{bmatrix}$$

$$\cdot\ \overset{2行2列}{\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}}\overset{2行p列}{\begin{bmatrix}a_{11} & a_{21} & \cdots & a_{p1}\\a_{12} & a_{22} & \cdots & a_{p2}\end{bmatrix}}=\overset{2行p列}{\begin{bmatrix}1\times a_{11}+0\times a_{12} & 1\times a_{21}+0\times a_{22} & \cdots & 1\times a_{p1}+0\times a_{p2}\\0\times a_{11}+1\times a_{12} & 0\times a_{21}+1\times a_{22} & \cdots & 0\times a_{p1}+1\times a_{p2}\end{bmatrix}}=\begin{bmatrix}a_{11} & a_{21} & \cdots & a_{p1}\\a_{12} & a_{22} & \cdots & a_{p2}\end{bmatrix}$$

$$\cdot\ \overset{p行2列}{\begin{bmatrix}a_{11} & a_{12}\\a_{21} & a_{22}\\\vdots & \vdots\\a_{p1} & a_{p2}\end{bmatrix}}\overset{2行2列}{\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}}=\overset{p行2列}{\begin{bmatrix}a_{11}\times 1+a_{12}\times 0 & a_{11}\times 0+a_{12}\times 1\\a_{21}\times 1+a_{21}\times 0 & a_{21}\times 0+a_{21}\times 1\\\vdots & \vdots\\a_{p1}\times 1+a_{p2}\times 0 & a_{p1}\times 0+a_{p2}\times 1\end{bmatrix}}=\begin{bmatrix}a_{11} & a_{12}\\a_{21} & a_{22}\\\vdots & \vdots\\a_{p1} & a_{p2}\end{bmatrix}$$

单位矩阵确实对原来的矩阵不会产生任何影响！


3. 旋 转
接下来讲坐标的“旋转”。
旋转啊……
(a_1, a_2)
0
某坐标（a_1, a_2）……
(b_1, b_2)
θ
(a_1, a_2)
0
以原点为中心旋转 θ 角度后得到坐标（b_1, b_2）
这个（b_1, b_2）就可以具体写作（$a_1\cos\theta - a_2\sin\theta$, $a_1\sin\theta + a_2\cos\theta$）
嗯……
也就是说……
$\begin{cases} b_1 = a_1\cos\theta - a_2\sin\theta \\ b_1 = a_1\sin\theta + a_2\cos\theta \end{cases}$
$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_1\cos\theta - a_2\sin\theta \\ a_1\sin\theta + a_2\cos\theta \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$
也可以写成这种形式。
这样啊！


做一道例题试试看吧！
求一下坐标（$\sqrt{3}$，1）旋转90°后得到的坐标（b_1, b_2）
1
($\sqrt{3}$, 1)
0
$\sqrt{3}$
一看就是（−1，$\sqrt{3}$），对不对啊？
还是算一下吧！
$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \cos 90° & -\sin 9 \\ \sin 90° & \cos 90 \end{bmatrix}$
嗯……
就是这样。
$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} \cos 90° & -\sin 90° \\ \sin 90° & \cos 90° \end{bmatrix} \begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{3} \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \times \sqrt{3} + (-1) \times 1 \\ 1 \times \sqrt{3} + 0 \times 1 \end{bmatrix} = \begin{bmatrix} -1 \\ \sqrt{3} \end{bmatrix}$
对，完全正确。

点 (a_1, a_2) 旋转 θ 角度后得到点 (b_1, b_2)，具体应写作 $(a_1\cos\theta - a_2\sin\theta, a_1\sin\theta + a_2\cos\theta)$。那么，点 (b_1, b_2) 如果旋转 $-\theta$ 角度的话会得到怎样的坐标呢？当然是点 (a_1, a_2) 了。

用矩阵的形式应写作：$\begin{bmatrix}\cos(-\theta) & -\sin(-\theta)\\ \sin(-\theta) & \cos(-\theta)\end{bmatrix}\begin{bmatrix}b_1\\ b_2\end{bmatrix} = \begin{bmatrix}a_1\\ a_2\end{bmatrix}$

$\begin{bmatrix}\cos(-\theta) & -\sin(-\theta)\\ \sin(-\theta) & \cos(-\theta)\end{bmatrix}\begin{bmatrix}b_1\\ b_2\end{bmatrix} = \begin{bmatrix}a_1\\ a_1\end{bmatrix}$的详细计算过程，如下所示：

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

点（b_1，b_2）的具体形式。

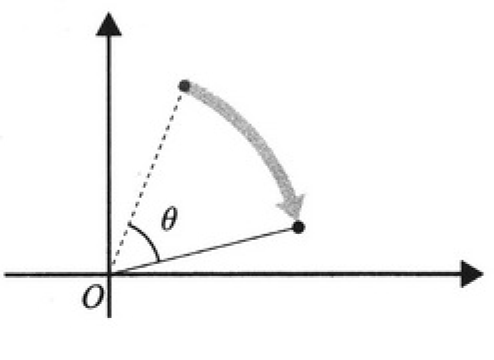
请仔细阅读第 70 页讲解的内容。

请仔细观察上式，

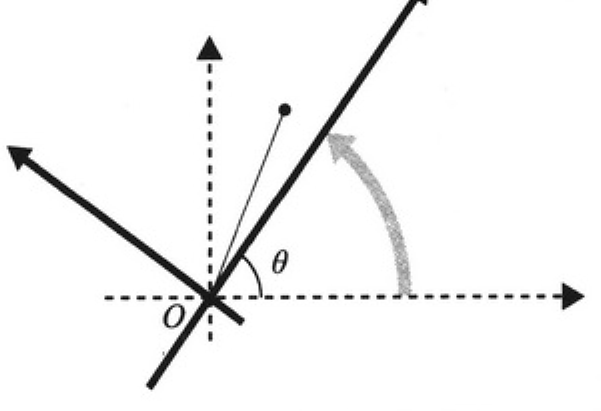
$\begin{bmatrix}\cos(-\theta) & -\sin(-\theta)\\ \sin(-\theta) & \cos(-\theta)\end{bmatrix}\begin{bmatrix}\cos\theta & -\sin\theta\\ \sin\theta & \cos\theta\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$ **是不是也成立啊？**

顺便讲一下，下面两幅图的意思是相同的。

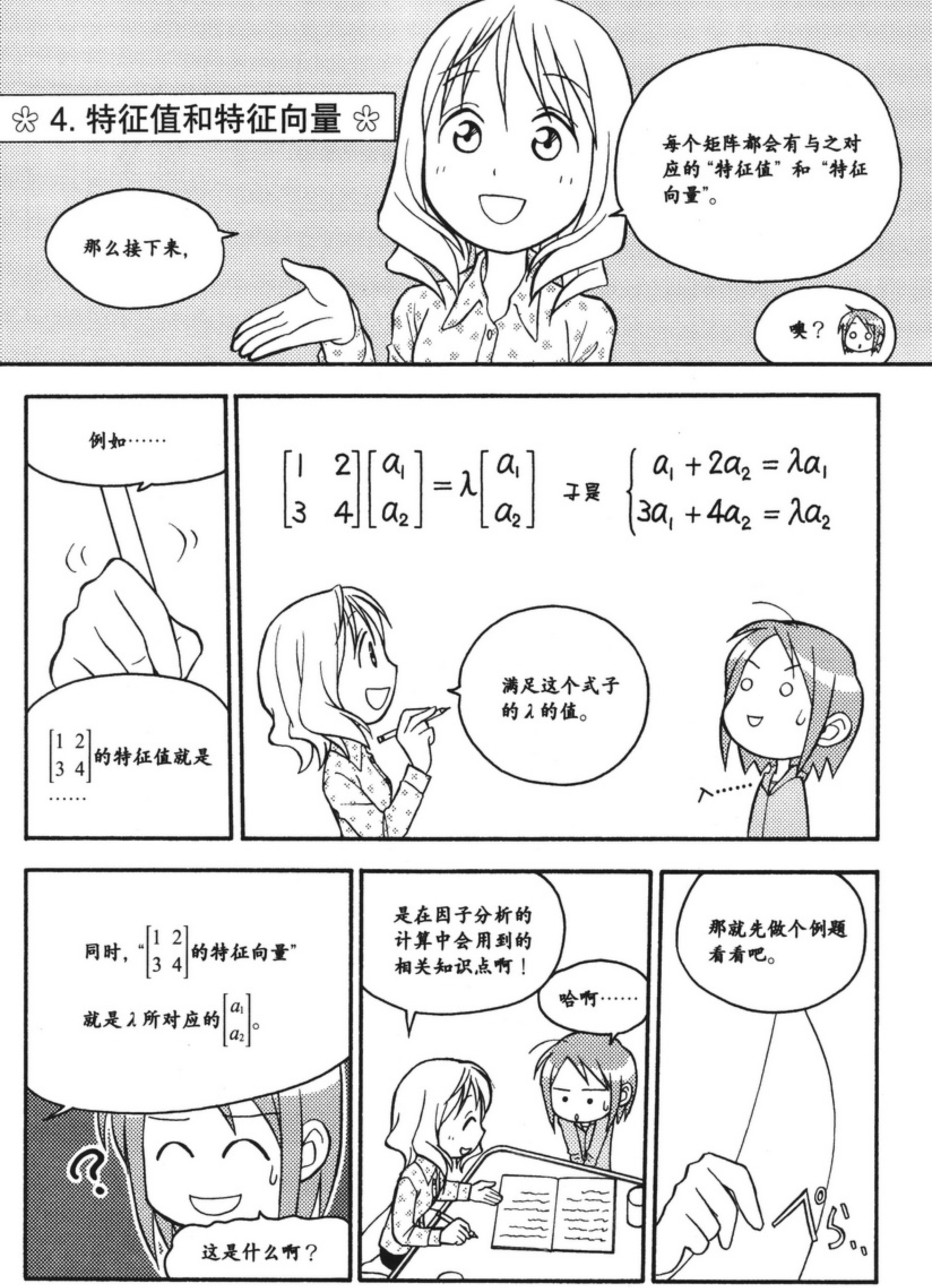
点（b_1,b_2）旋转 $-\theta$ 角度



纵轴和横轴旋转 $+\theta$ 角度



4. 特征值和特征向量



那么接下来，
每个矩阵都会有与之对应的“特征值”和“特征向量”。
噢？
例如……
$\begin{bmatrix}1&2\\3&4\end{bmatrix}$的特征值就是……
$\begin{bmatrix}1&2\\3&4\end{bmatrix}\begin{bmatrix}a_1\\a_2\end{bmatrix}=\lambda\begin{bmatrix}a_1\\a_2\end{bmatrix}$ 于是 $\begin{cases}a_1+2a_2=\lambda a_1\\3a_1+4a_2=\lambda a_2\end{cases}$
满足这个式子的λ的值。
λ……
同时，“$\begin{bmatrix}1&2\\3&4\end{bmatrix}$的特征向量”
就是λ所对应的$\begin{bmatrix}a_1\\a_2\end{bmatrix}$。
这是什么啊？
是在因子分析的计算中会用到的相关知识点啊！
哈啊……
那就先做个例题看看吧。

举个例子吧。如果要讲解详细计算过程的话，会非常繁琐，所以我们只需记住结论。

例 1

$$\cdot\begin{bmatrix}-10 & 6\\-18 & 11\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix}=\begin{bmatrix}-10\times1+\ \ 6\times2\\-18\times1+11\times2\end{bmatrix}=\begin{bmatrix}2\\4\end{bmatrix}=2\begin{bmatrix}1\\2\end{bmatrix}$$

$$\cdot\begin{bmatrix}-10 & 6\\-18 & 11\end{bmatrix}\begin{bmatrix}2\\3\end{bmatrix}=\begin{bmatrix}-10\times2+\ \ 6\times3\\-18\times2+11\times3\end{bmatrix}=\begin{bmatrix}-2\\-3\end{bmatrix}=-\begin{bmatrix}2\\3\end{bmatrix}$$

所以，由上可知 2 和 –1 就是$\begin{bmatrix}-10 & 6\\-18 & 11\end{bmatrix}$的特征值，2 对应的特征向量是$\begin{bmatrix}1\\2\end{bmatrix}$，–1 对应的特征向量是$\begin{bmatrix}2\\3\end{bmatrix}$。

例 2

$$\cdot\begin{bmatrix}2&0&0\\0&4&0\\0&0&6\end{bmatrix}\begin{bmatrix}1\\0\\0\end{bmatrix}=\begin{bmatrix}2\times1+0\times0+0\times0\\0\times1+4\times0+0\times0\\0\times1+0\times0+6\times0\end{bmatrix}=\begin{bmatrix}2\\0\\0\end{bmatrix}=2\begin{bmatrix}1\\0\\0\end{bmatrix}$$

$$\cdot\begin{bmatrix}2&0&0\\0&4&0\\0&0&6\end{bmatrix}\begin{bmatrix}0\\1\\0\end{bmatrix}=\begin{bmatrix}2\times0+0\times1+0\times0\\0\times0+4\times1+0\times0\\0\times0+0\times1+6\times0\end{bmatrix}=\begin{bmatrix}0\\4\\0\end{bmatrix}=4\begin{bmatrix}0\\1\\0\end{bmatrix}$$

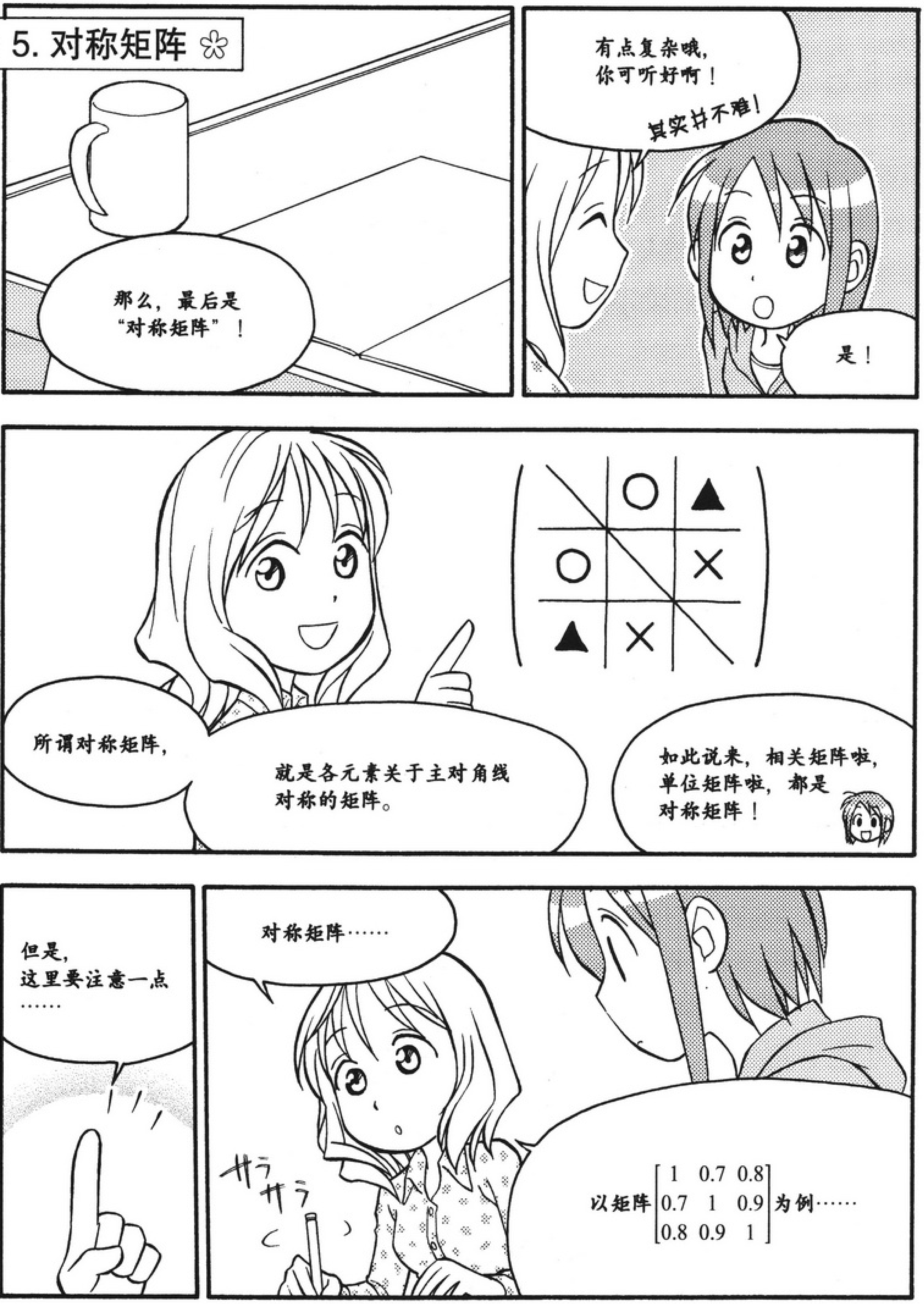
$$\cdot\begin{bmatrix}2&0&0\\0&4&0\\0&0&6\end{bmatrix}\begin{bmatrix}0\\0\\1\end{bmatrix}=\begin{bmatrix}2\times0+0\times0+0\times1\\0\times0+4\times0+0\times1\\0\times0+0\times0+6\times1\end{bmatrix}=\begin{bmatrix}0\\0\\6\end{bmatrix}=6\begin{bmatrix}0\\0\\1\end{bmatrix}$$

所以，由上可知，2、4、6 是$\begin{bmatrix}2&0&0\\0&4&0\\0&0&6\end{bmatrix}$的特征值，2 对应的特征向量是$\begin{bmatrix}1\\0\\0\end{bmatrix}$，4 对应的特征向量是$\begin{bmatrix}0\\1\\0\end{bmatrix}$，6 对应的特征向量是$\begin{bmatrix}0\\0\\1\end{bmatrix}$。

p 行 p 列矩阵的特征值和特征向量，原则上讲，存在 p 组。


但是这个要怎样计算呢？
真的要通过纸笔计算的话，过程就太复杂了。
啊！知道了！是不是用 Excel 进行计算？
遗憾的是，Excel 中也没有相应的函数。
啊？那要怎么办啊？
还是通过纸笔计算吧，
或者用数据分析软件计算。
是这样啊！

5. 对称矩阵



那么，最后是“对称矩阵”！
有点复杂哦，你可听好啊！
其实并不难！
是！
所谓对称矩阵，
就是各元素关于主对角线对称的矩阵。
如此说来，相关矩阵啦，单位矩阵啦，都是对称矩阵！
但是，这里要注意一点……
对称矩阵……
サラ サラ
以矩阵$\begin{bmatrix} 1 & 0.7 & 0.8 \\ 0.7 & 1 & 0.9 \\ 0.8 & 0.9 & 1 \end{bmatrix}$为例……

√最大特征值　　√第2大特征值　　√第3大特征值

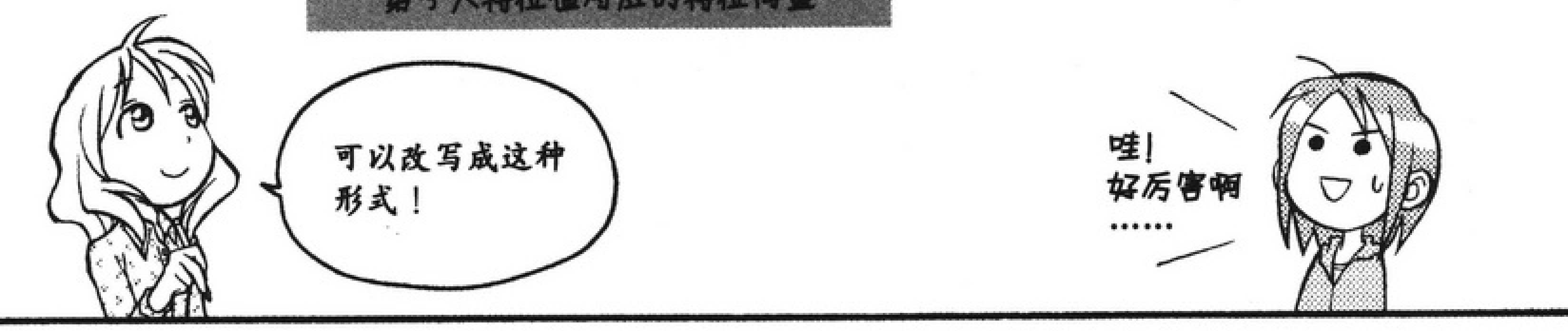
$$\begin{bmatrix}1&0.7&0.8\\0.7&1&0.9\\0.8&0.9&1\end{bmatrix}=\begin{bmatrix}\sqrt{2.60}\times\boxed{0.55}&\sqrt{0.31}\times\boxed{0.80}&\sqrt{0.08}\times\boxed{0.23}\\\sqrt{2.60}\times\boxed{0.58}&\sqrt{0.31}\times\boxed{(-0.57)}&\sqrt{0.08}\times\boxed{0.59}\\\sqrt{2.60}\times\boxed{0.60}&\sqrt{0.31}\times\boxed{(-0.19)}&\sqrt{0.08}\times\boxed{(-0.78)}\end{bmatrix}\begin{bmatrix}\sqrt{2.60}\times0.55&\sqrt{2.60}\times0.58&\sqrt{2.60}\times0.60\\\sqrt{0.31}\times0.80&\sqrt{0.31}\times(-0.57)&\sqrt{0.31}\times(-0.19)\\\sqrt{0.08}\times0.23&\sqrt{0.08}\times0.59&\sqrt{0.08}\times(-0.78)\end{bmatrix}$$

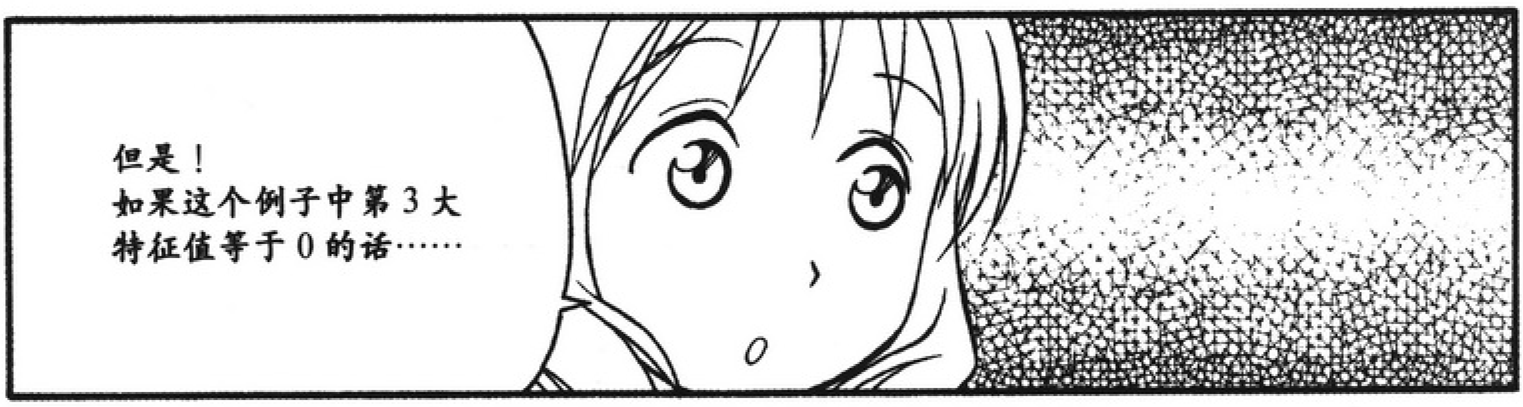
将左边矩阵的行、列互换后得到的矩阵

最大特征值对应的特征向量

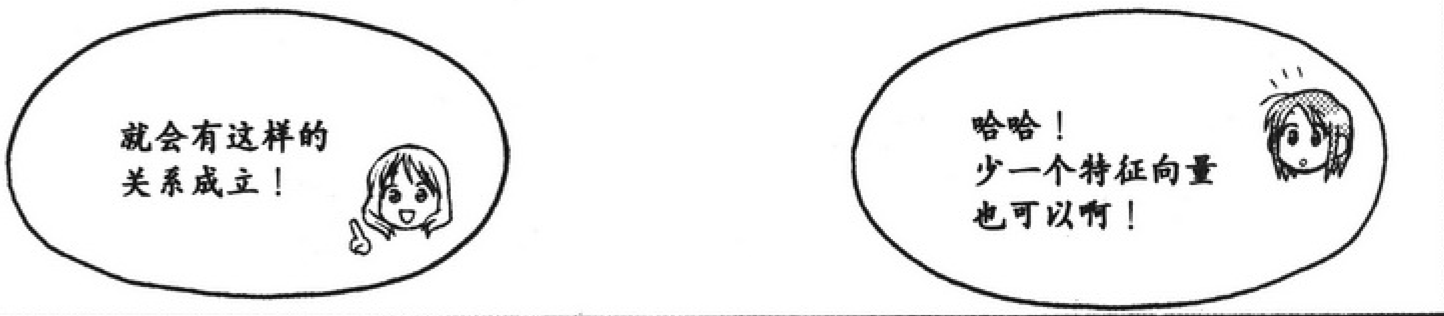
第2大特征值对应的特征向量

第3大特征值对应的特征向量





$$\begin{bmatrix}1&0.7&0.8\\0.7&1&0.9\\0.8&0.9&1\end{bmatrix}\approx\begin{bmatrix}\sqrt{2.60}\times0.55&\sqrt{0.31}\times0.80\\\sqrt{2.60}\times0.58&\sqrt{0.31}\times(-0.57)\\\sqrt{2.60}\times0.60&\sqrt{0.31}\times(-0.19)\end{bmatrix}\begin{bmatrix}\sqrt{2.60}\times0.55&\sqrt{2.60}\times0.58&\sqrt{2.60}\times0.60\\\sqrt{0.31}\times0.80&\sqrt{0.31}\times(-0.57)&\sqrt{0.31}\times(-0.19)\end{bmatrix}$$



那还不算什么，如果在这个例子中第 2 大和第 3 大特征值都等于零的话……
$$\begin{bmatrix} 1 & 0.7 & 0.8 \\ 0.7 & 1 & 0.9 \\ 0.8 & 0.9 & 1 \end{bmatrix} \approx \begin{bmatrix} \sqrt{2.60}\times 0.55 \\ \sqrt{2.60}\times 0.58 \\ \sqrt{2.60}\times 0.60 \end{bmatrix} \begin{pmatrix} \sqrt{2.60}\times 0.55 & \sqrt{2.60}\times 0.58 & \sqrt{2.60}\times 0.60 \end{pmatrix}$$
如上关系也成立！
哇……好有趣啊！
这个对理解因子分析的计算方法，非常重要，可一定要记好啊！
是，我会记住的！
那、今天我们就到此为止吧！
真是太谢谢您了！


啊、这样啊！
小露儿下次来我家吧？
我买了好茶哟。
啊？要走啊！
那……我们下周见！
请多关照！
パタン
太棒了！
离山本又近了一步啊！
再来复习一遍！
露儿……
我不是在做梦吧?!

✿ 6. 矩阵的补充 ✿

6.1 矩阵的书写规则

例如，$\begin{cases} x_1 + 2x_2 = -1 \\ 3x_1 + 4x_2 = 5 \end{cases}$ 可以写作 $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} -1 \\ 5 \end{bmatrix}$，$\begin{cases} x_1 + 2x_2 \\ 3x_1 + 4x_2 \end{cases}$ 可以写作 $\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$。

小 结

· $\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1q}x_q = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2q}x_q = b_2 \\ \cdots\cdots \\ a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pq}x_q = b_p \end{cases}$ 可以写作 $\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix}$。

· $\begin{cases} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1q}x_q \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2q}x_q \\ \cdots\cdots \\ a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pq}x_q \end{cases}$ 可以写作 $\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{bmatrix}$。

6.2 矩阵的加法

例如，$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}$和$\begin{bmatrix}4 & 5\\-2 & 4\end{bmatrix}$相加，

$$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}+\begin{bmatrix}4 & 5\\-2 & 4\end{bmatrix}$$

可以按照如此方式$\begin{bmatrix}1+\ \ 4 & 2+5\\3+(-2) & 4+4\end{bmatrix}$进行计算。

小　结

$\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1q}\\a_{21} & a_{22} & \cdots & a_{2q}\\\vdots & \vdots & \ddots & \vdots\\a_{p1} & a_{p2} & \cdots & a_{pq}\end{bmatrix}$和$\begin{bmatrix}b_{11} & b_{12} & \cdots & b_{1q}\\b_{21} & b_{22} & \cdots & b_{2q}\\\vdots & \vdots & \ddots & \vdots\\b_{p1} & b_{p2} & \cdots & b_{pq}\end{bmatrix}$相加，

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

计算结果为：$\begin{bmatrix}a_{11}+b_{11} & a_{12}+b_{12} & \cdots & a_{1q}+b_{1q}\\a_{21}+b_{21} & a_{22}+b_{22} & \cdots & a_{2q}+b_{2q}\\\vdots & \vdots & \ddots & \vdots\\a_{p1}+b_{p1} & a_{p2}+b_{p2} & \cdots & a_{pq}+b_{pq}\end{bmatrix}$。

6.3 矩阵的乘法

例如，$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}$和$\begin{bmatrix}x_1 & y_1\\x_2 & y_2\end{bmatrix}$相乘，

$$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}\begin{bmatrix}x_1 & y_1\\x_2 & y_2\end{bmatrix}$$

虽然称为“乘法”，但其实相当于将$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}\begin{bmatrix}x_1\\x_2\end{bmatrix}$和$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}\begin{bmatrix}y_1\\y_2\end{bmatrix}$，也就是$\begin{cases}x_1 + 2x_2\\3x_1 + 4x_2\end{cases}$和$\begin{cases}y_1 + 2y_2\\3y_1 + 4y_2\end{cases}$同时表示了出来。

例

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

就是分别进行

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

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

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

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

然后再将以上结果合并表示出来

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

小　结

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

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

虽然也称为“乘法”，但是要先将

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{bmatrix} \begin{bmatrix} x_{11} \\ x_{21} \\ \vdots \\ x_{q1} \end{bmatrix} 、 \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{bmatrix} \begin{bmatrix} x_{12} \\ x_{22} \\ \vdots \\ x_{q2} \end{bmatrix} \cdots\cdots$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1q} \\ a_{21} & a_{22} & \cdots & a_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pq} \end{bmatrix} \begin{bmatrix} x_{1r} \\ x_{2r} \\ \vdots \\ x_{qr} \end{bmatrix} \text{分别进行运算，}$$

$$\text{也就是将} \begin{cases} a_{11}x_{11} + a_{12}x_{12} + \cdots + a_{1q}x_{q1} \\ a_{21}x_{11} + a_{22}x_{21} + \cdots + a_{2q}x_{q1} \\ \cdots\cdots \\ a_{p1}x_{11} + a_{p2}x_{21} + \cdots + a_{pq}x_{q1} \end{cases} 、 \begin{cases} a_{11}x_{12} + a_{12}x_{22} + \cdots + a_{1q}x_{q2} \\ a_{21}x_{12} + a_{22}x_{22} + \cdots + a_{2q}x_{q2} \\ \cdots\cdots \\ a_{p1}x_{12} + a_{p2}x_{22} + \cdots + a_{pq}x_{q2} \end{cases} \cdots\cdots$$

$$\begin{cases} a_{11}x_{1r} + a_{12}x_{2r} + \cdots + a_{1q}x_{qr} \\ a_{21}x_{1r} + a_{22}x_{2r} + \cdots + a_{2q}x_{qr} \\ \cdots\cdots \\ a_{p1}x_{1r} + a_{p2}x_{2r} + \cdots + a_{pq}x_{qr} \end{cases} \text{同时表示出来即可。}$$

6.4 逆矩阵

例如，$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}$的逆矩阵就是与$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}$相乘以后得到$\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$的那个2行2列矩阵。

$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}$的逆矩阵，通常记作$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}^{-1}$。

例

由于$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}\begin{bmatrix}-2 & 1\\1.5 & -0.5\end{bmatrix}=\begin{bmatrix}1\times(-2)+2\times1.5 & 1\times1+2\times(-0.5)\\3\times(-2)+4\times1.5 & 3\times1+4\times(-0.5)\end{bmatrix}=\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$

因此$\begin{bmatrix}-2 & 1\\1.5 & -0.5\end{bmatrix}=\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix}^{-1}$。

小 结

$\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1p}\\a_{21} & a_{22} & \cdots & a_{2p}\\\vdots & \vdots & \ddots & \vdots\\a_{p1} & a_{p2} & \cdots & a_{pp}\end{bmatrix}$ 的逆矩阵 $\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1p}\\a_{21} & a_{22} & \cdots & a_{2p}\\\vdots & \vdots & \ddots & \vdots\\a_{p1} & a_{p2} & \cdots & a_{pp}\end{bmatrix}^{-1}$ 就是

与 $\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1p}\\a_{21} & a_{22} & \cdots & a_{2p}\\\vdots & \vdots & \ddots & \vdots\\a_{p1} & a_{p2} & \cdots & a_{pp}\end{bmatrix}$ 相乘后得到 $\begin{bmatrix}1 & 0 & \cdots & 0\\0 & 1 & \cdots & 0\\\vdots & \vdots & \ddots & \vdots\\0 & 0 & \cdots & 1\end{bmatrix}$ 的p行p列矩阵。

6.5 转置矩阵

例如，$\begin{bmatrix}1&2\\3&4\end{bmatrix}$的转置矩阵就是$\begin{bmatrix}1&3\\2&4\end{bmatrix}$。也就是说将$\begin{bmatrix}1&2\\3&4\end{bmatrix}$的行和列互换后得到的矩阵。

$\begin{bmatrix}1&2\\3&4\end{bmatrix}$的转置矩阵通常记作$\begin{bmatrix}1&2\\3&4\end{bmatrix}^{\mathrm{T}}$或$\begin{bmatrix}1&2\\3&4\end{bmatrix}'$。

例

$\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\\10&11&12\\13&14&15\end{bmatrix}$的转置矩阵$\begin{bmatrix}1&2&3\\4&5&6\\7&8&9\\10&11&12\\13&14&15\end{bmatrix}^{\mathrm{T}}$，也就是$\begin{bmatrix}1&4&7&10&13\\2&5&8&11&14\\3&6&9&12&15\end{bmatrix}$。

$(-3\ \ 0\ \ 8\ \ -7)$的转置矩阵$(-3\ \ 0\ \ 8\ \ -7)^{\mathrm{T}}$就是$\begin{bmatrix}-3\\0\\8\\-7\end{bmatrix}$。

小　结

$\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1q}\\a_{21}&a_{22}&\cdots&a_{2q}\\\vdots&\vdots&\ddots&\vdots\\a_{p1}&a_{p2}&\cdots&a_{pq}\end{bmatrix}$的转置矩阵$\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1q}\\a_{21}&a_{22}&\cdots&a_{2q}\\\vdots&\vdots&\ddots&\vdots\\a_{p1}&a_{p2}&\cdots&a_{pq}\end{bmatrix}^{\mathrm{T}}$就是

将$\begin{bmatrix}a_{11}&a_{12}&\cdots&a_{1q}\\a_{21}&a_{22}&\cdots&a_{2q}\\\vdots&\vdots&\ddots&\vdots\\a_{p1}&a_{p2}&\cdots&a_{pq}\end{bmatrix}$的行和列互换后，得到的$\begin{bmatrix}a_{11}&a_{21}&\cdots&a_{p1}\\a_{12}&a_{22}&\cdots&a_{p2}\\\vdots&\vdots&\ddots&\vdots\\a_{1q}&a_{2q}&\cdots&a_{pq}\end{bmatrix}$。

✿ 7. 离差平方和、方差、标准差 ✿

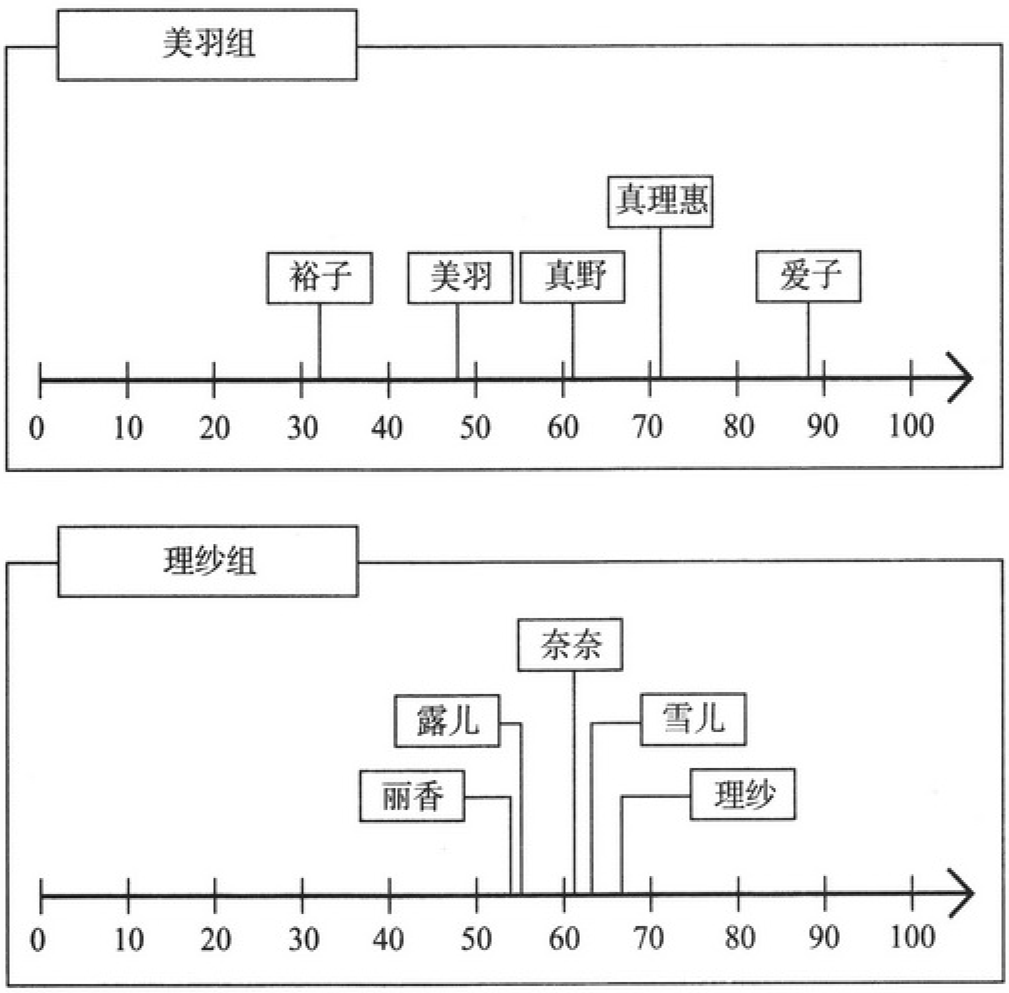
美羽、理纱与打工的同伴们一起去唱卡拉 OK。她们每 5 人一组，分成两组根据演唱得分进行比赛。比赛结果如下表。

◆表3.1　卡拉OK的PK结果

	美羽组（得分）
美羽	48
裕子	32
爱子	88
真野	61
真理惠	71
平均分	60

	理纱组（得分）
理纱	67
露儿	55
奈奈	61
雪儿	63
丽香	54
平均分	60

将上表作成图便可得到下图。



◆图3.1　卡拉OK的PK结果

虽然美羽组和理纱组的平均得分都是60分，但具体的情况却相去甚远。美羽组这方，每个人的得分是不是分布得很不平均呢？也就是说数据的“分散程度”比较大。

人们通常采用离差平方和、(总体)方差和(总体)标准差作为表征数据“离散程度”的指标。这些指标都具有如下性质：

· 最小值为0。
· 数据的“离散程度”越大，它们的值也就越大。

离差平方和，常常会出现在以回归分析为代表的多种分析方法的计算过程中。

$$\text{离差平方和} = (\text{每个数据} - \text{平均值})^2\ \text{相加之和}$$

通过上述计算便可求解出离差平方和的值。然而数据的个数越多，它的值也就会变得越大，这也成为它的一个致命缺点，所以在实际操作中，我们很少使用它作为表征“离散程度”的指标。

(总体)方差，解决了离差平方和的缺点。可以通过如下计算求得它的值[1]。

$$(\text{总体})\text{方差} = \frac{\text{离差平方和}}{\text{数据的个数}}$$

(总体)标准差，从本质上讲与(总体)方差是相同的。可以通过如下计算求得它的值。

$$(\text{总体})\text{标准差} = \sqrt{(\text{总体})\text{方差}}$$

让我们来求一下美羽组和理纱组的离差平方和、(总体)方差和(总体)标准差吧！

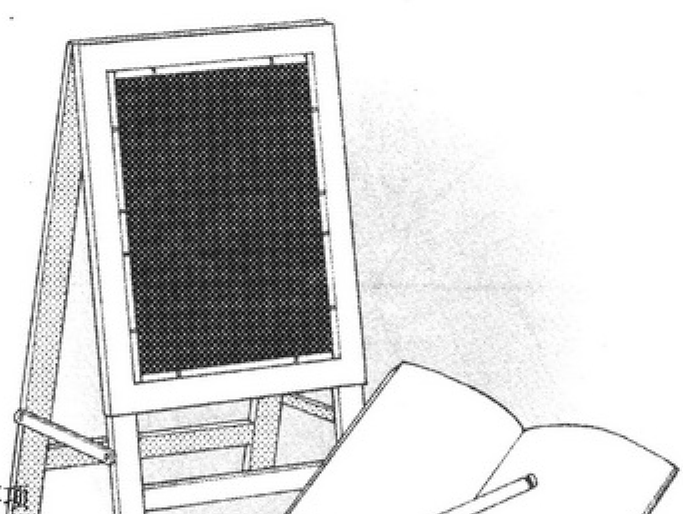
◆**表3.2 美羽组和理纱组的离差平方和、(总体)方差和(总体)标准差**

	美羽组	理纱组
离差平方和	$(48-60)^2+(32-60)^2+(88-60)^2+(61-60)^2+(71-60)^2$ $=(-12)^2+(-28)^2+28^2+1^2+11^2$ $=1834$	$(67-60)^2+(55-60)^2+(61-60)^2+(63-60)^2+(54-60)^2$ $=7^2+(-5)^2+1^2+3^2+(-6)^2$ $=120$
方差	$\frac{1834}{5}=366.8$	$\frac{120}{5}=24$
标准差	$\sqrt{366.8}=19.2$	$\sqrt{24}=4.9$

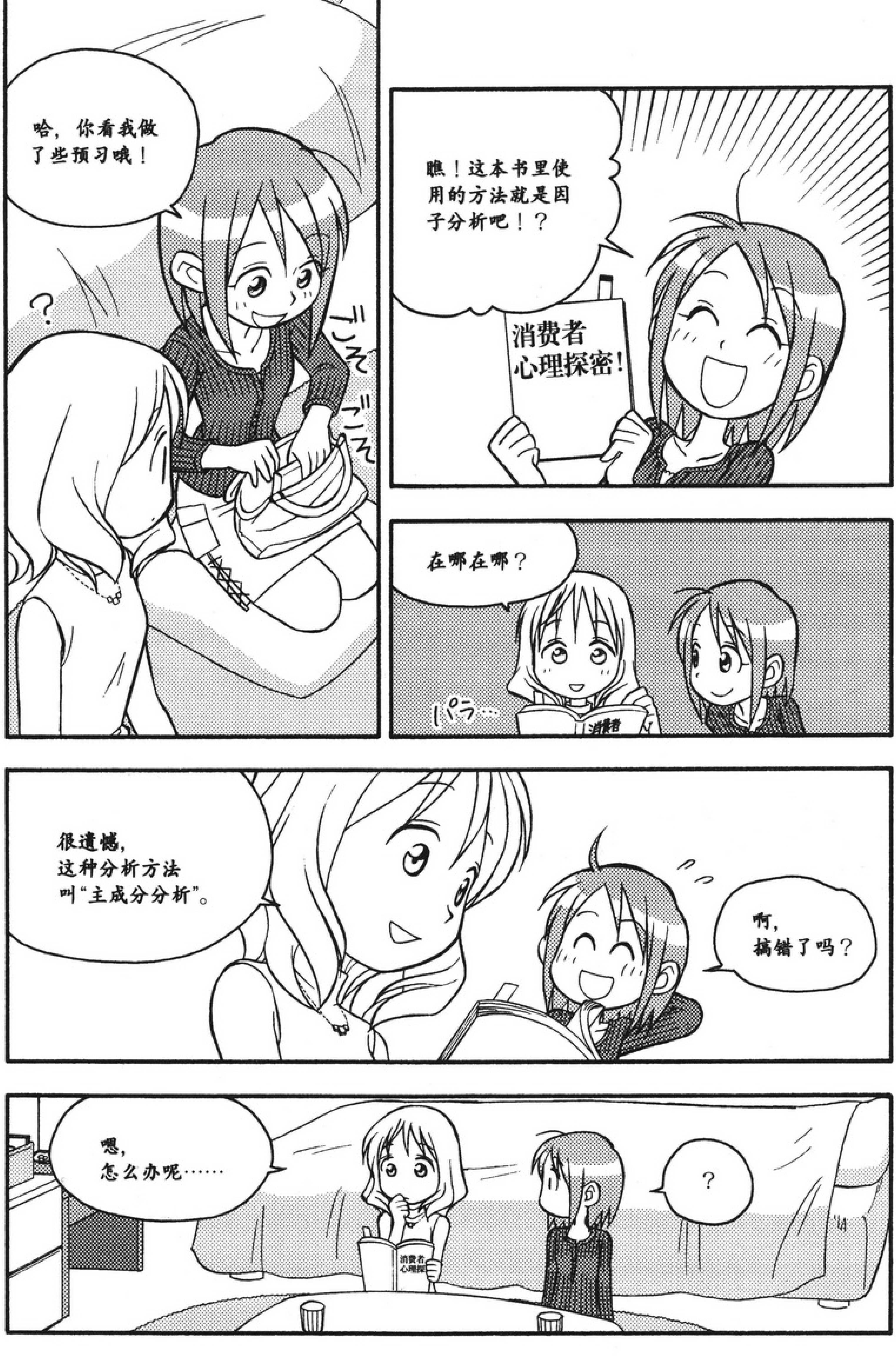
1. 在方差中，也有不采用“数据的个数”而采用“数据的个数 -1”作为分母的情况，我们将其称为样本方差。由于篇幅所限，这两种方差的区别在本书中就不做讨论了。同时，在本书后面的章节中出现的方差基本上指的都是样本方差，而标准差也是指$\sqrt{\text{样本方差}}$。

第 4 章

主成分分析



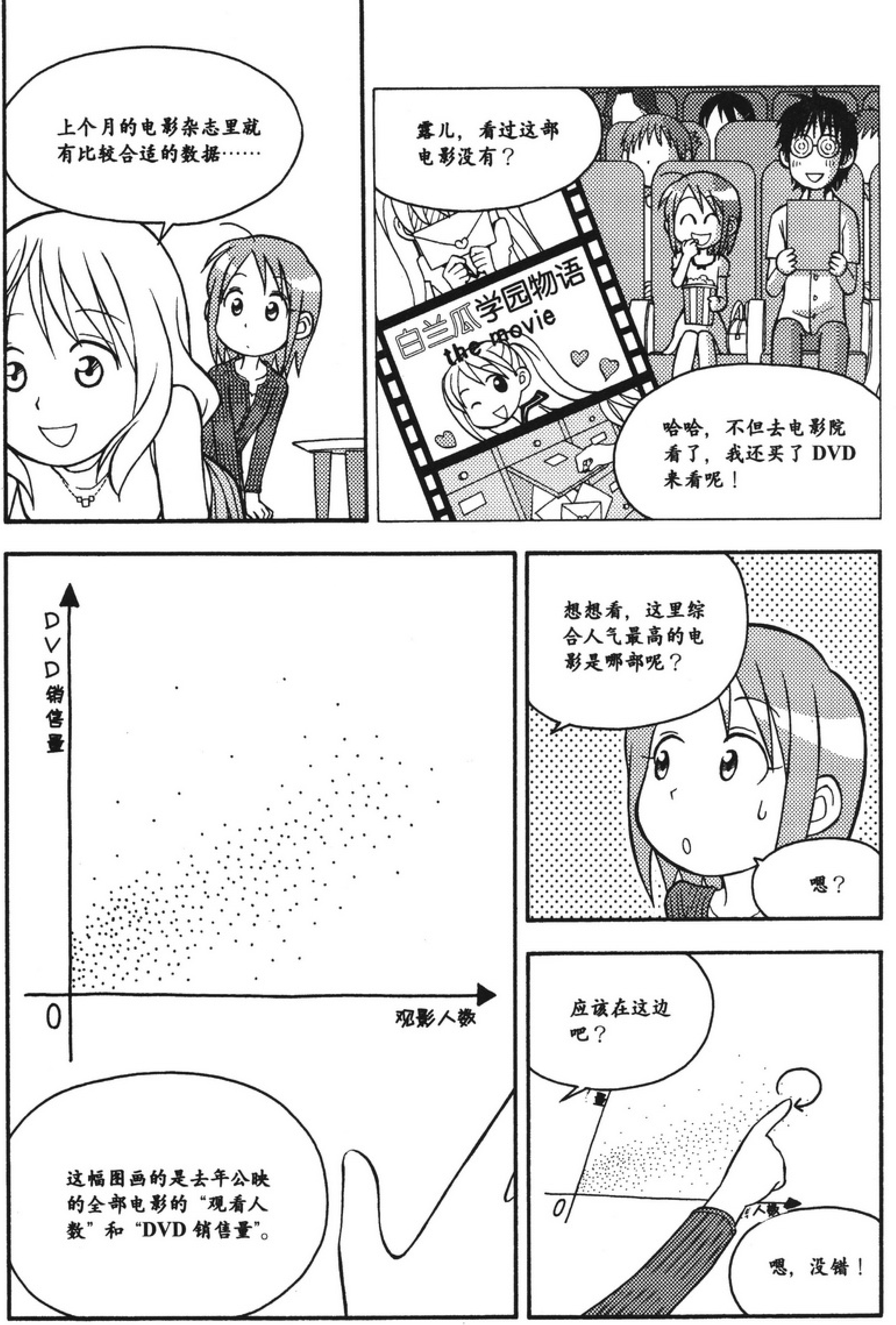
啊哈……
美羽师姐的房间
好可爱啊！♡♡♡
是……
是吗？
哇……
钢、钢琴……
哇……
可爱的小装饰!
咦……
啊!
ゴホン
那，那么，
我们要开始
学习因子分析了。
是！

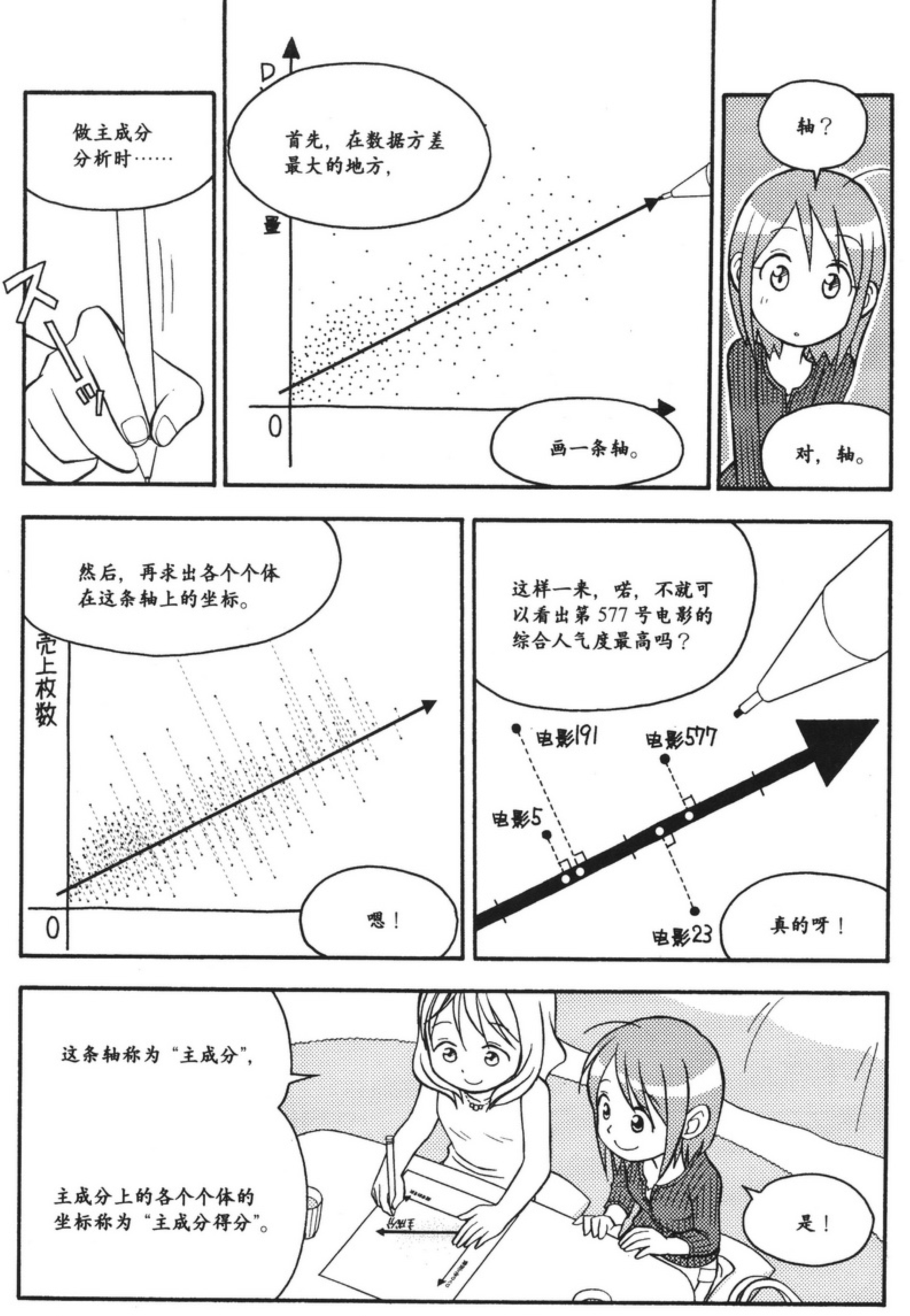


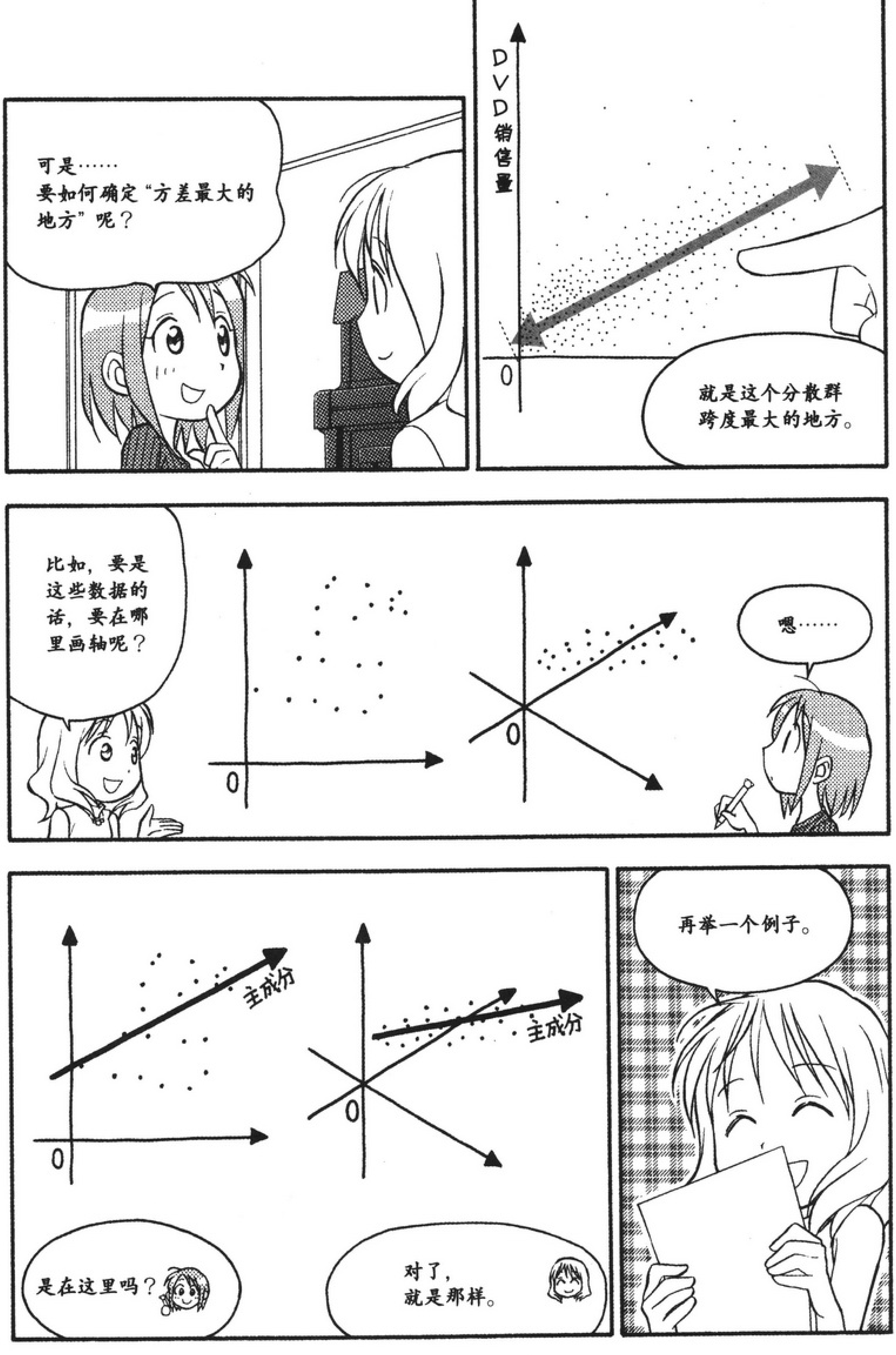
哈，你看我做了些预习哦！
？
ごそ
ごそ
瞧！这本书里使用的方法就是因子分析吧！？
消费者心理探密！
在哪在哪？
パラ…
很遗憾，这种分析方法叫“主成分分析”。
啊，搞错了吗？
嗯，怎么办呢……
？

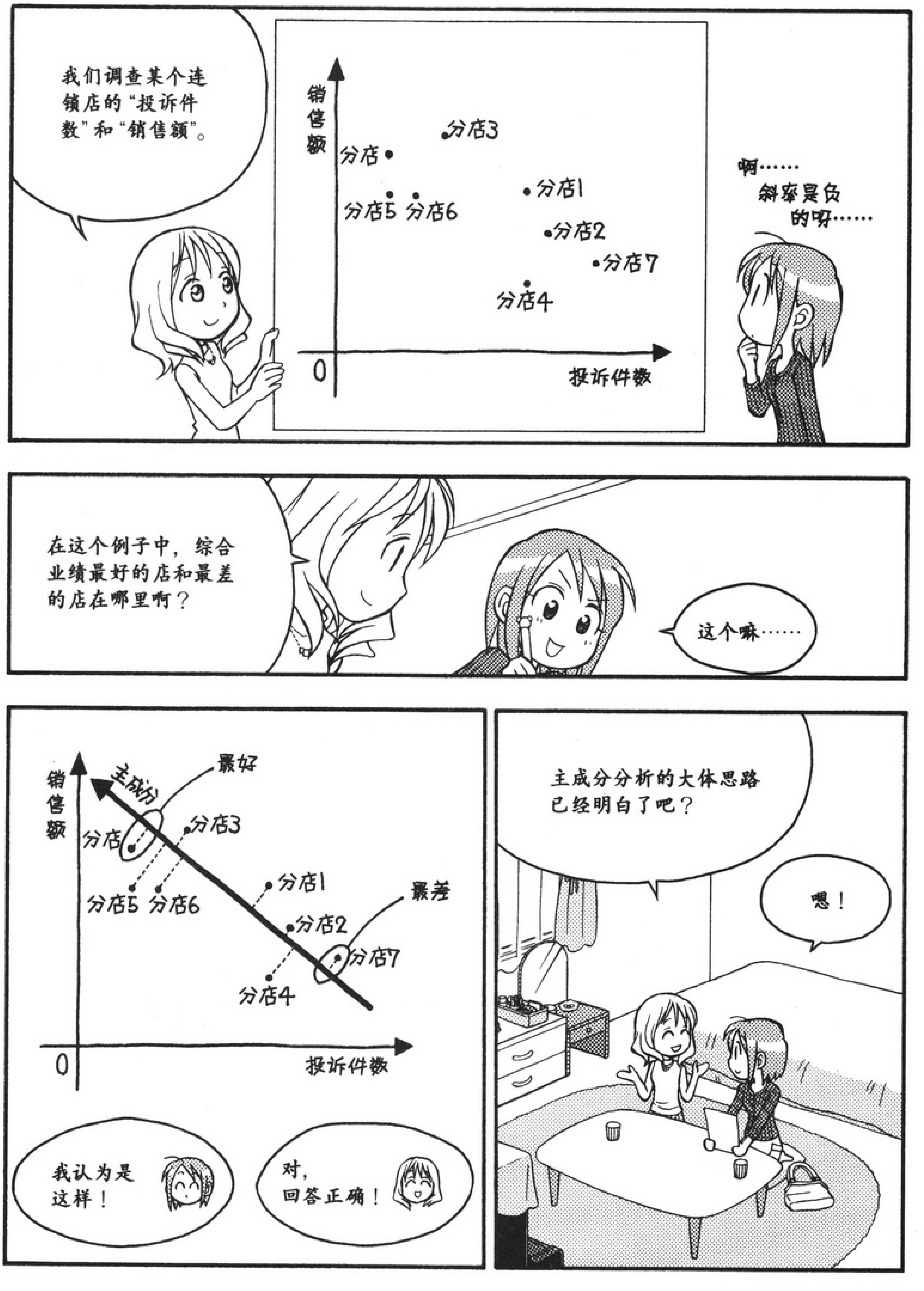

露儿，要不我们先学一下主成分分析吧？
啊？
主成分分析同因子分析看起来很相似，很多人都会混淆这两种方法。
为了避免今后露儿也犯同样的错误，不如就借这个机会先学习一下吧！
是这样啊……
这个，还是你拿主意吧！
唉……如何是好呢……
真想赶快学会因子分析啊，可是……
啫……连两者的差别都知道！
那当然！
要是这样的话，也不错啊！
那、那就学吧……
决定了？
フッ フッフッ

1. 主成分分析
那我们就更改原定计划，先学习主成分分析吧！
好的！请多多指教！
顺便说一下，如果给主成分分析增加一些约束条件，那么两者就完全相同了。
这么相似啊！
但是，反过来讲，如果不加约束条件，两者就是不一样的。
如此说来，那它是怎样的分析呢？
主成分分析……
是为了"选出综合实力最强"的一种分析方法。
太好了——
综合实力最强
唉，什么意思呢？


上个月的电影杂志里就
有比较合适的数据……
露儿，看过这部
电影没有？
白兰瓜学园物语
the movie
哈哈，不但去电影院
看了，我还买了 DVD
来看呢！
DVD销售量
0
观影人数
这幅图画的是去年公映
的全部电影的“观看人
数”和“DVD 销售量”。
想想看，这里综
合人气最高的电
影是哪部呢？
嗯？
应该在这边
吧？
嗯，没错！

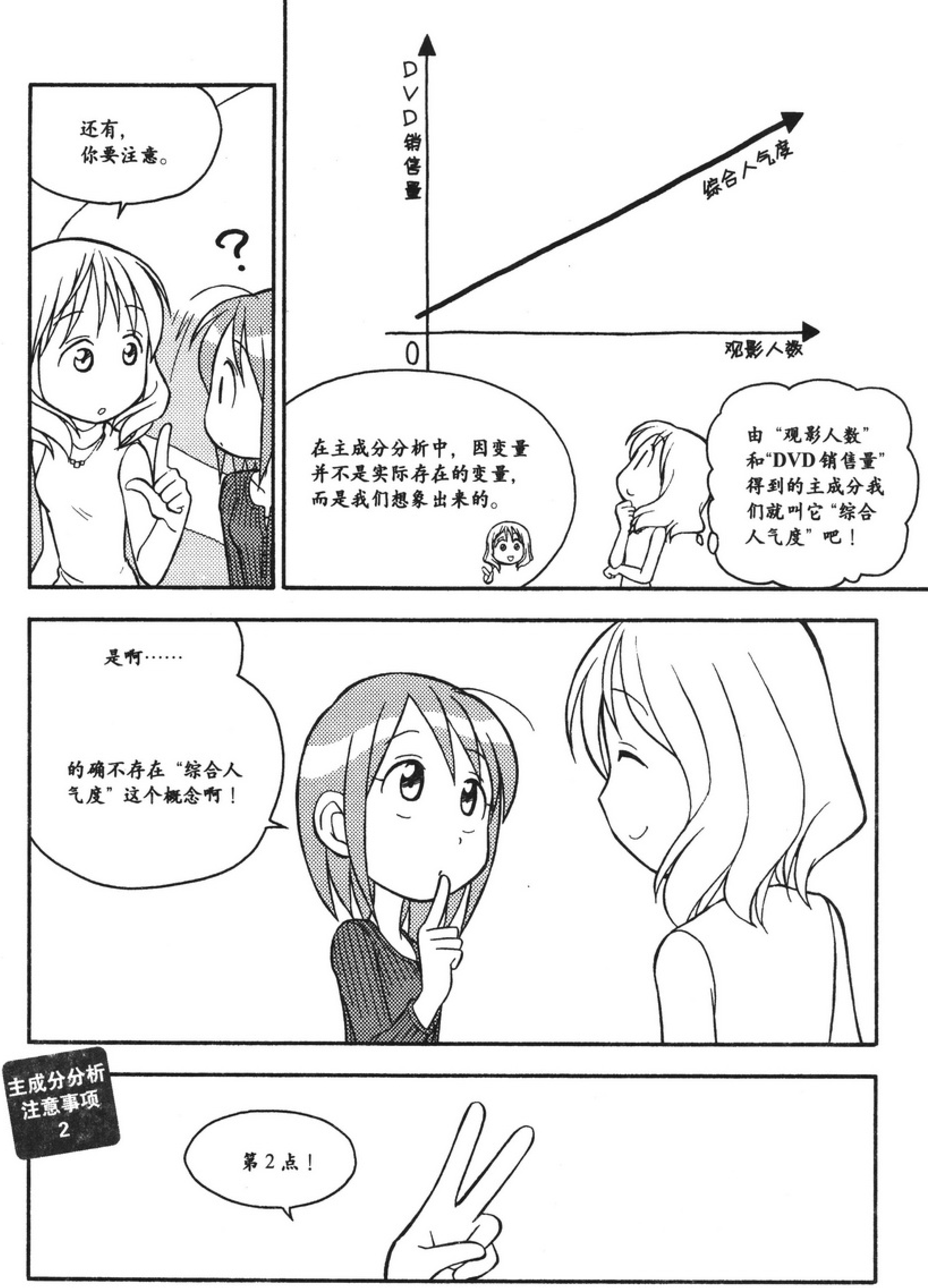

做主成分
分析时……
首先，在数据方差
最大的地方，
画一条轴。
0
轴？
对，轴。
然后，再求出各个个体
在这条轴上的坐标。
売上枚数
0
嗯！
这样一来，喏，不就可
以看出第577号电影的
综合人气度最高吗？
电影191
电影577
电影5
电影23
真的呀！
这条轴称为“主成分”，
主成分上的各个个体的
坐标称为“主成分得分”。
是！

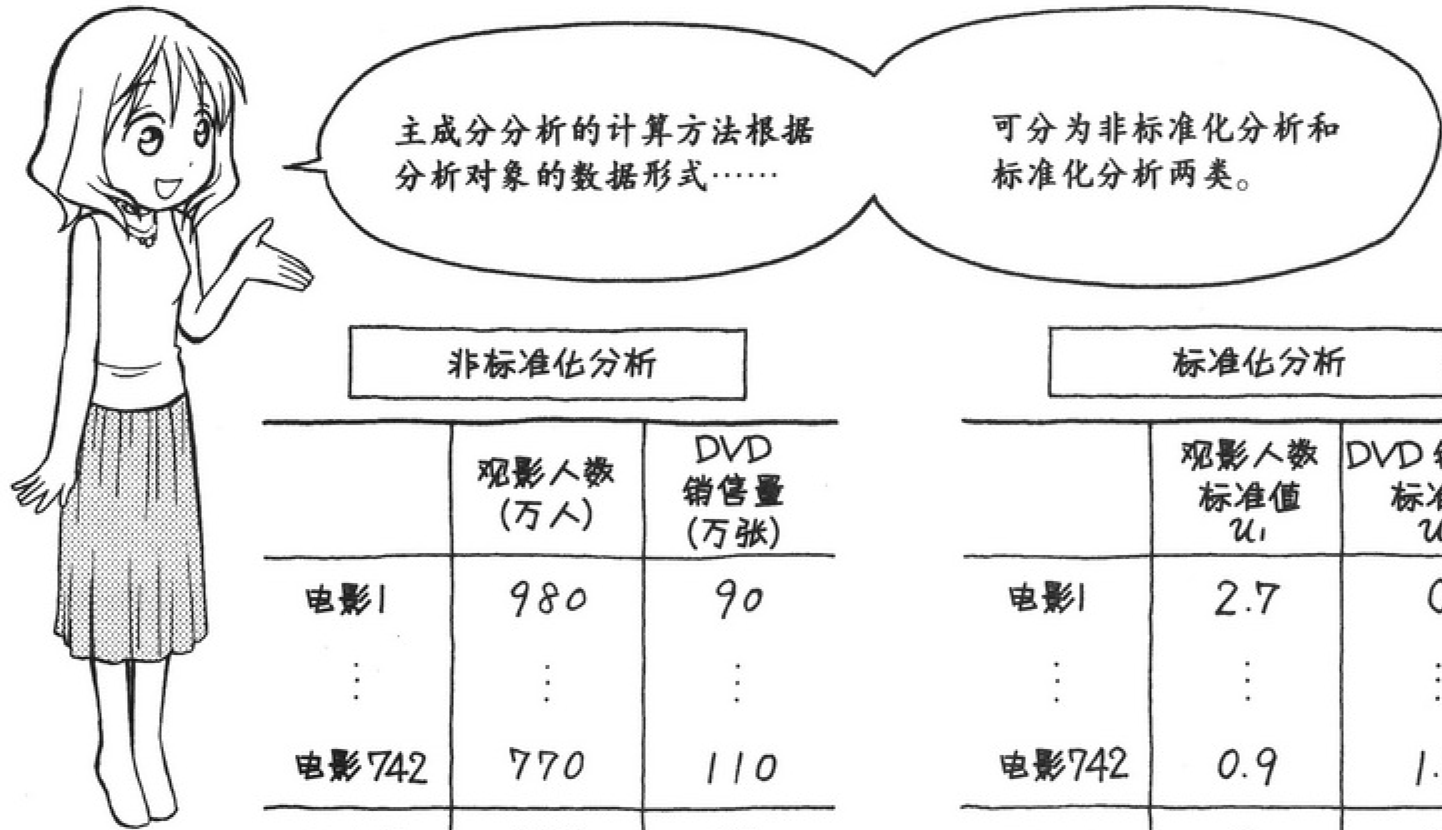

可是……
要如何确定“方差最大的地方”呢？
DVD销售量
0
就是这个分散群跨度最大的地方。
比如，要是这些数据的话，要在哪里画轴呢？
0
0
嗯……
再举一个例子。
主成分
0
主成分
0
是在这里吗？
对了，
就是那样。



我们调查某个连锁店的“投诉件数”和“销售额”。
销售额
分店3
分店
分店1
分店5
分店6
分店2
分店7
分店4
0
投诉件数
啊……
斜率是负的呀……
在这个例子中，综合业绩最好的店和最差的店在哪里啊？
这个嘛……
销售额
主成分
最好
分店3
分店
分店1
最差
分店5
分店6
分店2
分店7
分店4
0
投诉件数
我认为是这样！
对，回答正确！
主成分分析的大体思路已经明白了吧？
嗯！

✿ 2. 主成分分析的注意事项 ✿


还有，
你要注意。
?
DVD销售量
综合人气度
0
观影人数
在主成分分析中，因变量并不是实际存在的变量，而是我们想象出来的。
由“观影人数”和“DVD销售量”得到的主成分我们就叫它“综合人气度”吧！
是啊……
的确不存在“综合人气度”这个概念啊！
主成分分析注意事项2
第2点！

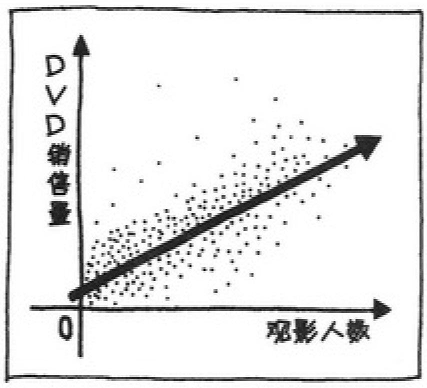


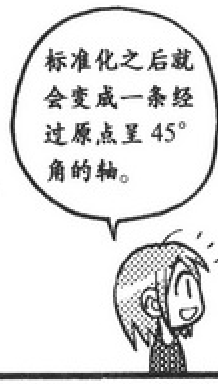
非标准化分析

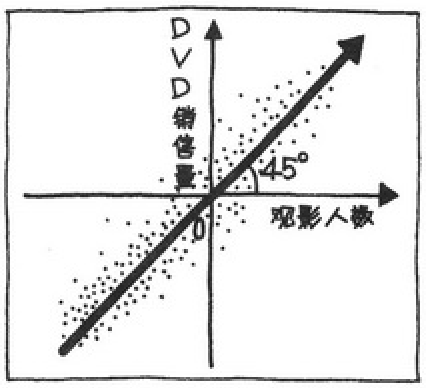
	观影人数（万人）	DVD 销售量（万张）
电影1	980	90
⋮	⋮	⋮
电影742	770	110
平均值	660	90
标准差	120	17

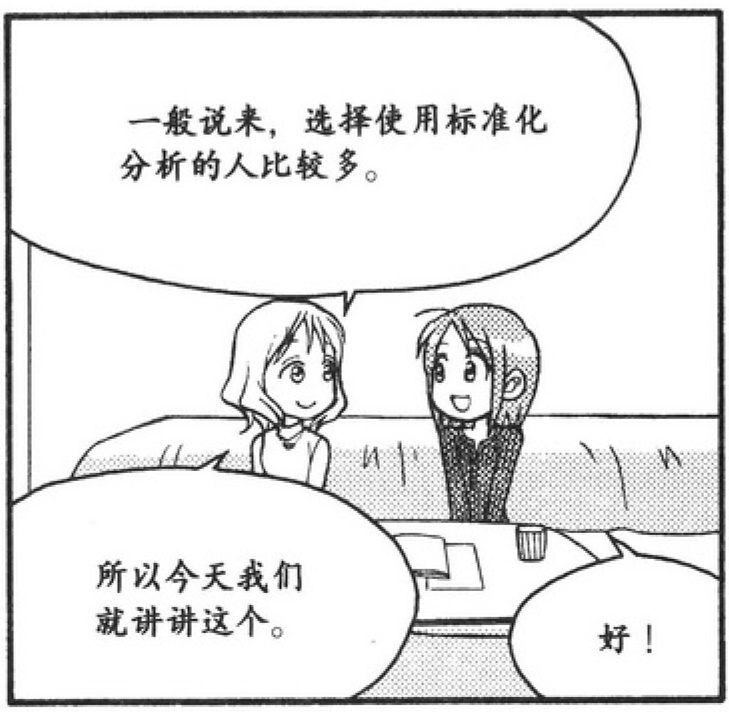
标准化分析

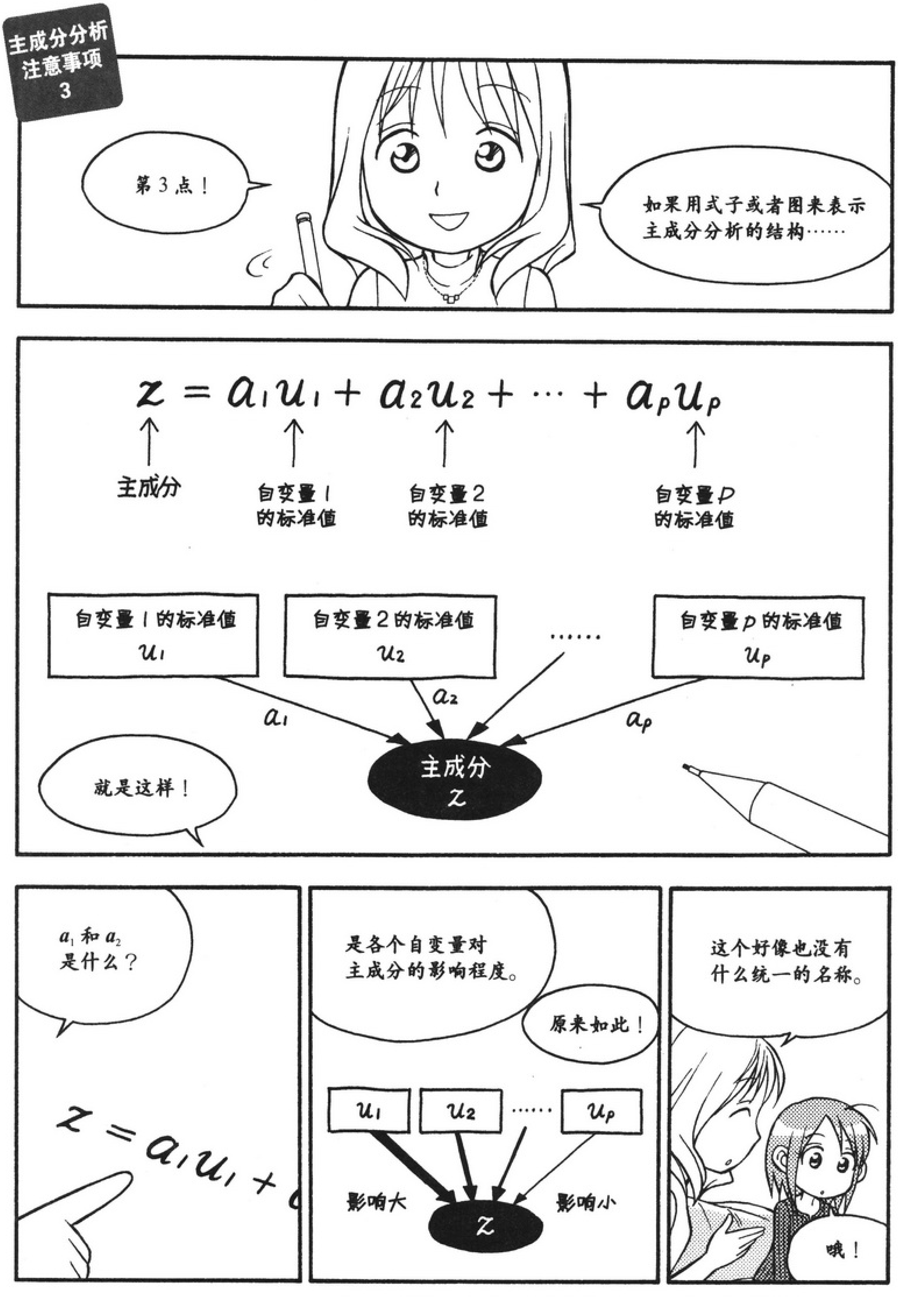
	观影人数标准值 u_1	DVD 销售量标准值 u_2
电影1	2.7	0
⋮	⋮	⋮
电影742	0.9	1.2
平均值	0	0
标准差	1	1

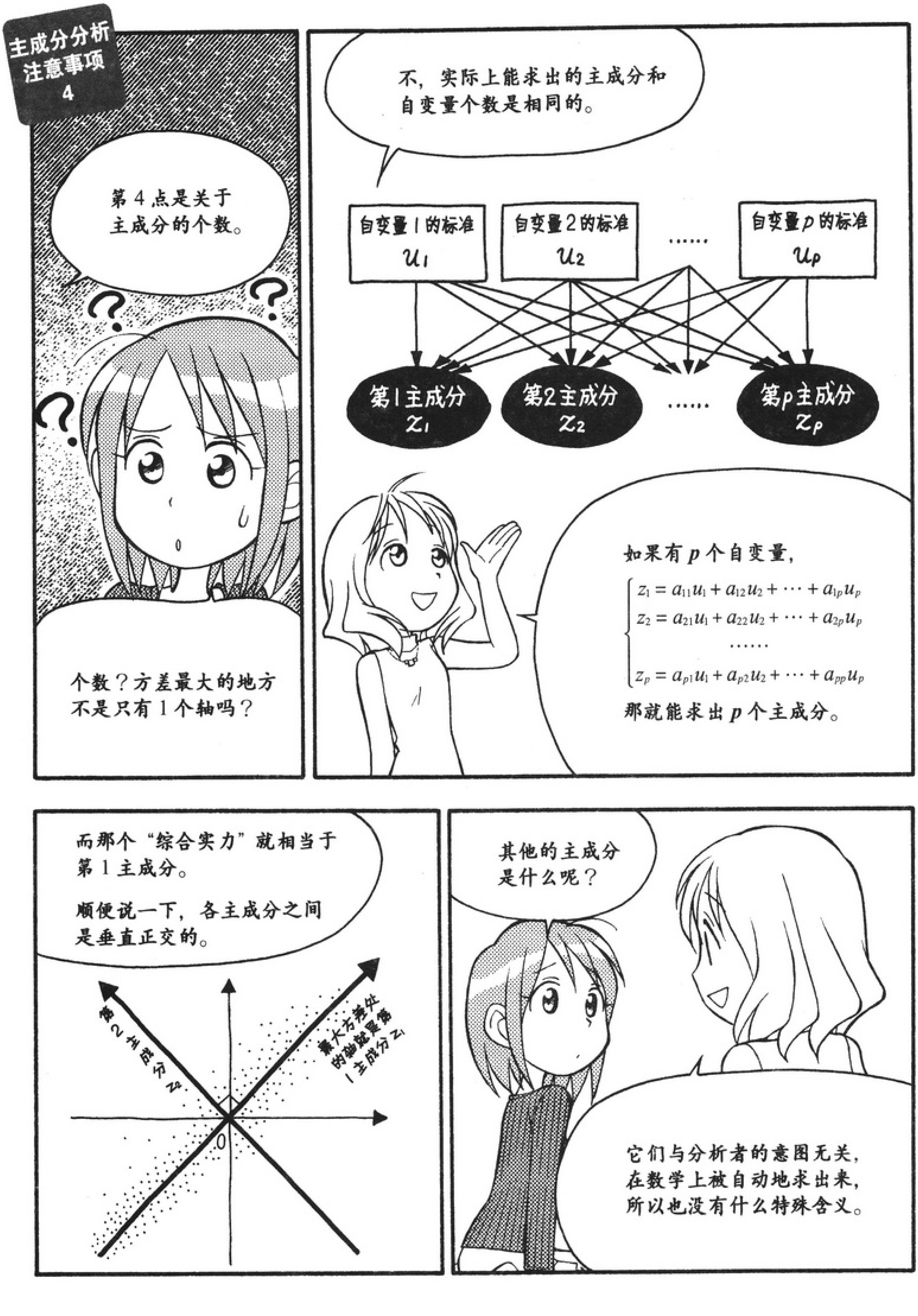


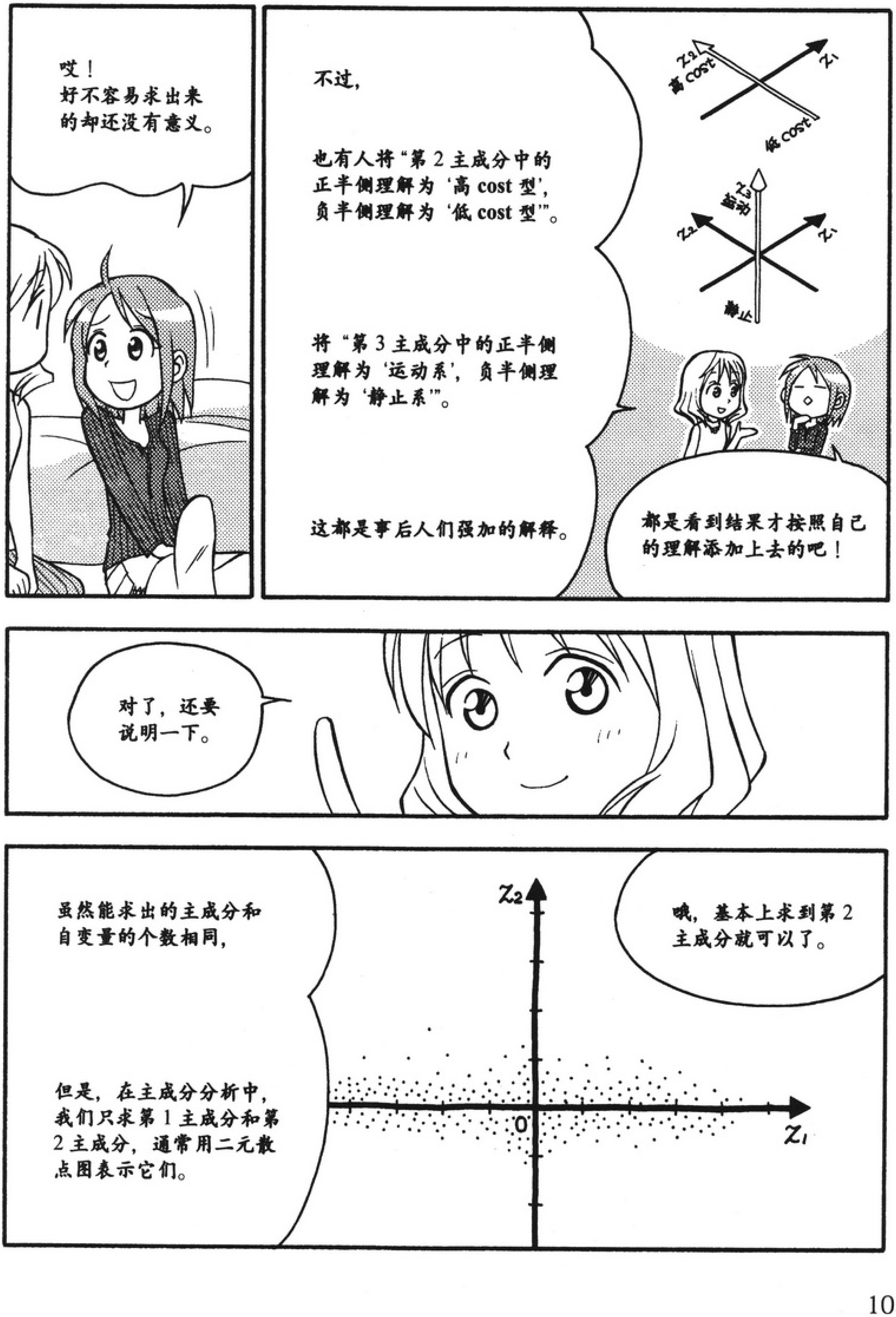






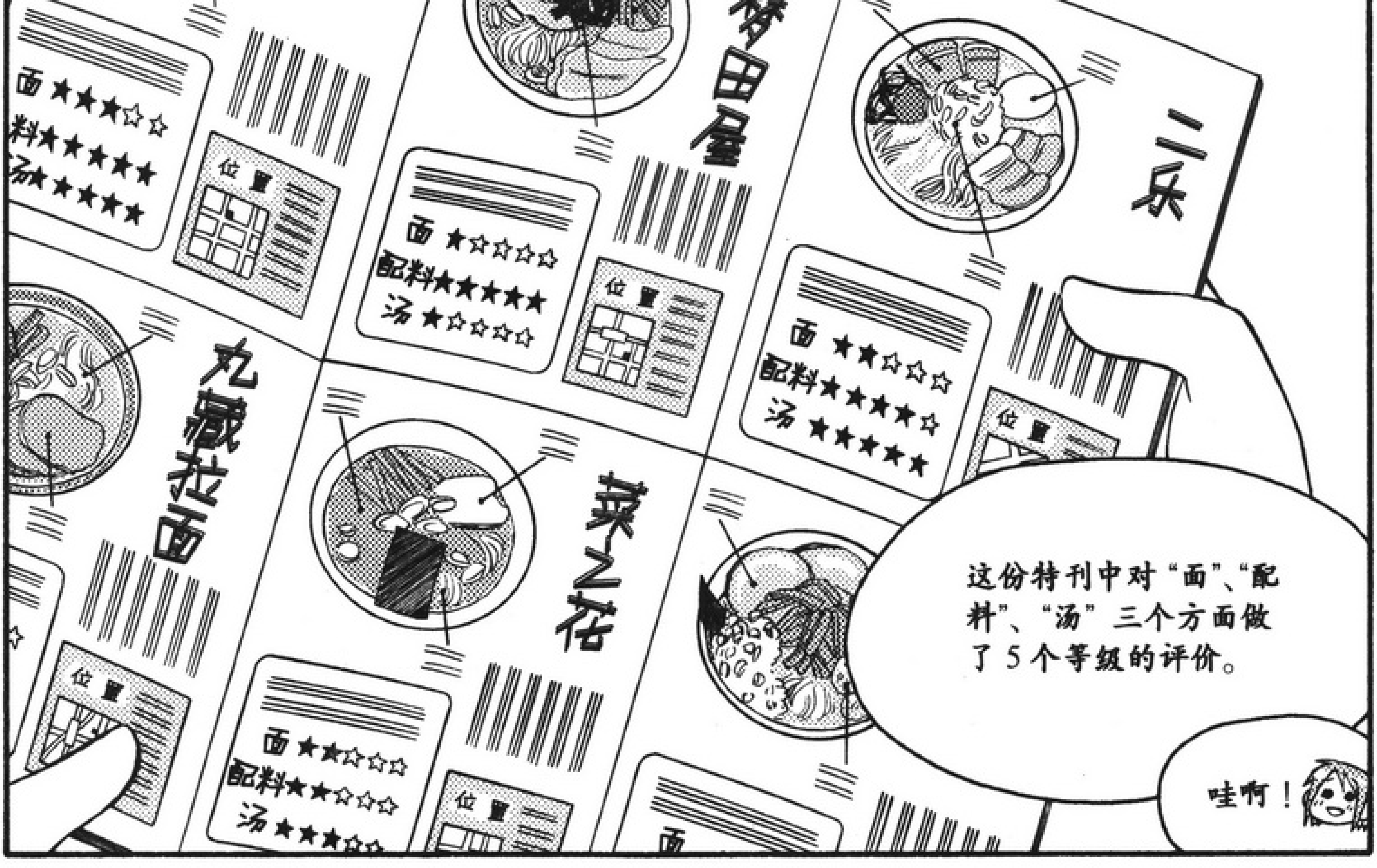

主成分分析注意事项 3
第 3 点！
如果用式子或者图来表示主成分分析的结构……
$z = a_1u_1 + a_2u_2 + \cdots + a_pu_p$
主成分
自变量 1 的标准值
自变量 2 的标准值
自变量 p 的标准值
自变量 1 的标准值 u_1
自变量 2 的标准值 u_2
……
自变量 p 的标准值 u_p
a_1
a_2
a_p
主成分 z
就是这样！
a_1 和 a_2 是什么？
$z = a_1u_1 +$
是各个自变量对主成分的影响程度。
原来如此！
u_1
u_2
……
u_p
影响大
影响小
z
这个好像也没有什么统一的名称。
哦！

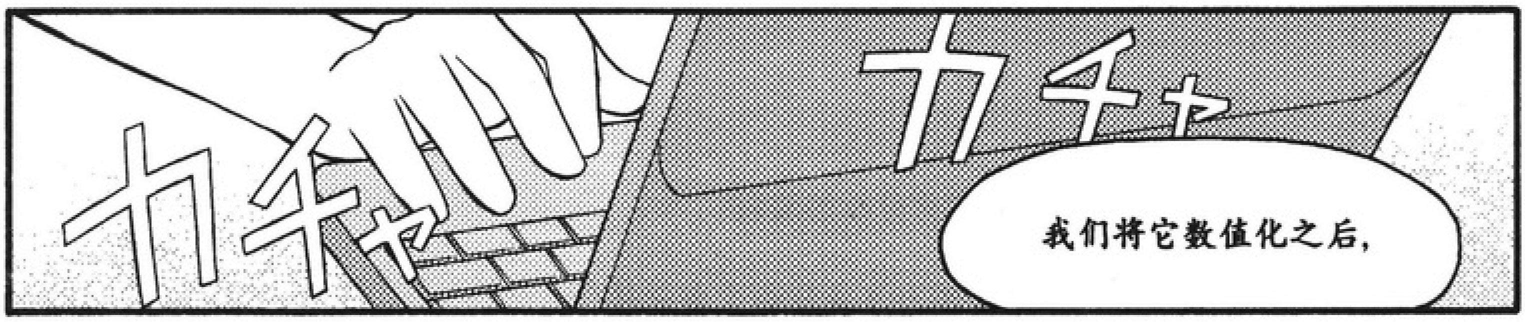

主成分分析注意事项 4
第 4 点是关于主成分的个数。
个数？方差最大的地方不是只有 1 个轴吗？
不，实际上能求出的主成分和自变量个数是相同的。
自变量 1 的标准 u_1
自变量 2 的标准 u_2
……
自变量 p 的标准 u_p
第 1 主成分 z_1
第 2 主成分 z_2
……
第 p 主成分 z_p
如果有 p 个自变量，
$$\begin{cases} z_1 = a_{11}u_1 + a_{12}u_2 + \cdots + a_{1p}u_p \\ z_2 = a_{21}u_1 + a_{22}u_2 + \cdots + a_{2p}u_p \\ \cdots\cdots \\ z_p = a_{p1}u_1 + a_{p2}u_2 + \cdots + a_{pp}u_p \end{cases}$$
那就能求出 p 个主成分。
而那个“综合实力”就相当于第 1 主成分。
顺便说一下，各主成分之间是垂直正交的。
第 2 主成分 z_2
最大方差处的轴就是第 1 主成分 z_1
0
其他的主成分是什么呢？
它们与分析者的意图无关，在数学上被自动地求出来，所以也没有什么特殊含义。



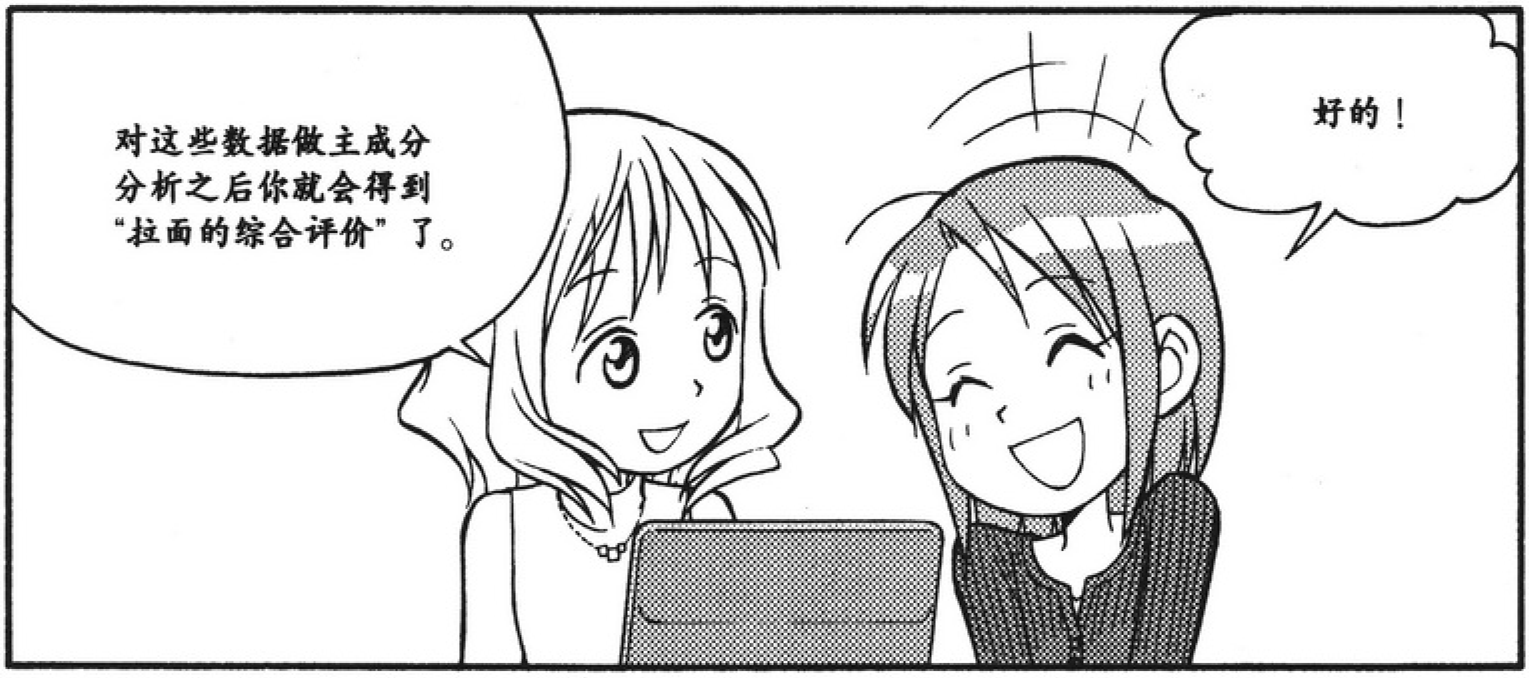
哎！好不容易求出来的却还没有意义。
不过，
也有人将“第2主成分中的正半侧理解为‘高cost型’，负半侧理解为‘低cost型’”。
将“第3主成分中的正半侧理解为‘运动系’，负半侧理解为‘静止系’”。
这都是事后人们强加的解释。
都是看到结果才按照自己的理解添加上去的吧！
Z2
高cost
低cost
Z1
Z3
运动
静止
对了，还要说明一下。
虽然能求出的主成分和自变量的个数相同，
但是，在主成分分析中，我们只求第1主成分和第2主成分，通常用二元散点图表示它们。
哦，基本上求到第2主成分就可以了。
Z2
0
Z1

✿ 3. 主成分分析的具体实例 ✿





	面	配料	汤
二乐	2	4	5
梦田屋	1	5	1
地回	5	3	4
菜之花	2	2	3
花之节	3	5	5
升辰轩	4	3	2
丸藏拉面	4	4	3
海乐亭	1	2	1
鸣海家	3	3	2
奏月	5	5	3


对这些数据做主成分分析之后你就会得到“拉面的综合评价”了。
好的！


主成分分析的流程
① 求出主成分和主成分得分。
② 确认分析结果的精度。
③ 讨论分析结果。
主成分分析的大致流程就是这样了。
噢，这样啊！

① 求出主成分和主成分得分

主成分和主成分得分点，可以按照步聚 1 到步骤 7 的计算过程进行求解。
在计算的过程中，我们使用了“拉格朗日乘数法”，由于从数学上解释起来比较困难，这里就不详细介绍了。读者只需记住这个名字即可。

步骤 1　变量标准化

	面	配料	汤
二乐	2	4	5
梦田屋	1	5	1
地回	5	3	4
菜之花	2	2	3
花之节	3	5	5
升辰轩	4	3	2
丸藏拉面	4	4	3
海乐亭	1	2	1
鸣海家	3	3	2
奏月	5	5	3
平均值	3.0	3.6	2.9
标准差	1.5	1.2	1.4

$$\sqrt{\frac{(2-3.0)^2+\cdots+(5-3.0)^2}{10-1}}=1.5$$

	“面”的标准值 u_1	“配料”的标准值 u_2	“汤”的标准值 u_3
二乐	−0.7	0.3	1.4
梦田屋	−1.3	1.2	−1.3
地回	1.3	−0.5	0.8
菜之花	−0.7	−1.4	0.1
花之节	0.0	1.2	1.4
升辰轩	0.7	−0.5	−0.6
丸藏拉面	0.7	0.3	0.1
海乐亭	−1.3	−1.4	−1.3
鸣海家	0.0	−0.5	−0.6
奏月	1.3	1.2	0.1
平均值	0	0	0
标准差	1	1	1

$$\frac{3-2.9}{1.4}=0.1$$

在主成分分析的标准化过程中，我们所用的标准差的分母，通常都是“数据个数－1”。

步骤 2 求相关矩阵

$$
\begin{array}{c|ccc}
 & \text{面} & \text{配料} & \text{汤} \\
\hline
\text{面} & \left[\begin{array}{c} 1 \\ 0.19 \\ 0.36 \end{array}\right. & \begin{array}{c} 0.19 \\ 1 \\ 0.30 \end{array} & \left.\begin{array}{c} 1.36 \\ 0.30 \\ 1 \end{array}\right] \\
\text{配料} & & & \\
\text{汤} & & &
\end{array}
$$

	面	配料	汤
面	1	0.19	1.36
配料	0.19	1	0.30
汤	0.36	0.30	1

步骤 3 求满足 $\begin{bmatrix} 1 & 0.19 & 0.36 \\ 0.19 & 1 & 0.30 \\ 0.36 & 0.30 & 1 \end{bmatrix}\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \lambda \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$ 的特征值 λ 和特征向量 $\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$。

再对特征向量进行单位化处理，即令其满足 $a_1^2+a_2^2+a_3^2=1$

通过数据分析软件，就可得到以下结果。

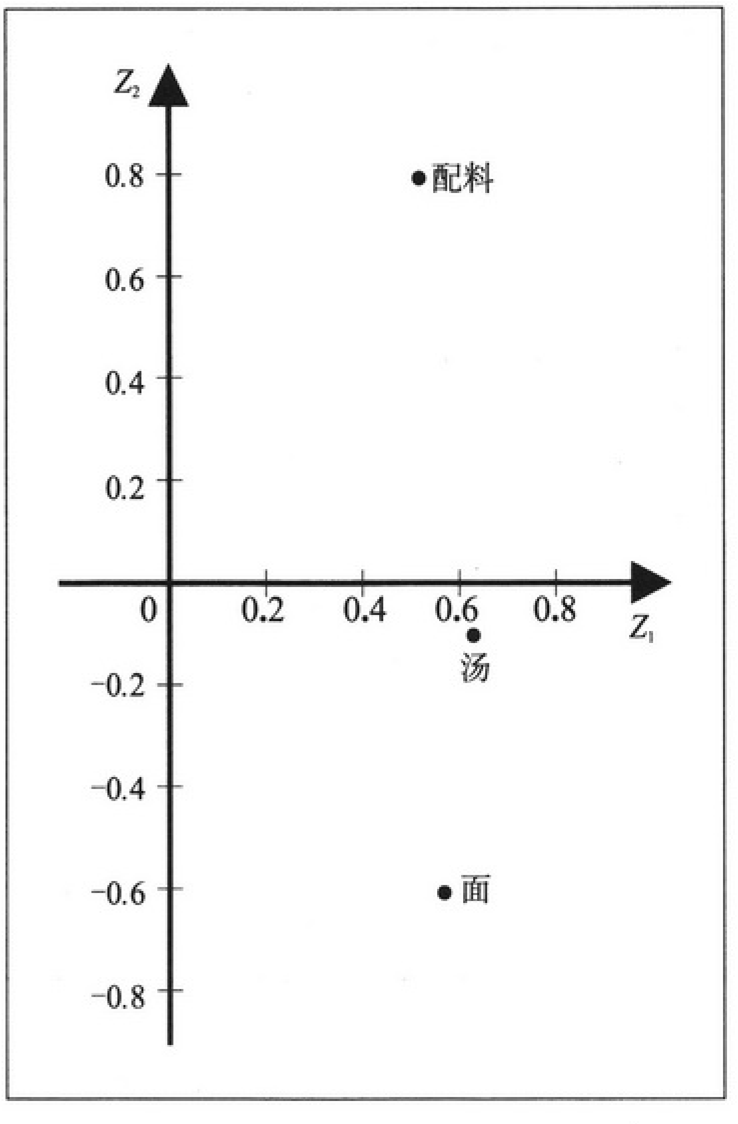
特征值 λ	特征向量 $\begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix}$
1.6	$\begin{bmatrix} 0.57 \\ 0.52 \\ 0.63 \end{bmatrix}$
0.8	$\begin{bmatrix} -0.60 \\ 0.79 \\ -0.11 \end{bmatrix}$
0.6	$\begin{bmatrix} -0.55 \\ -0.32 \\ 0.77 \end{bmatrix}$

步骤 4　根据步骤 3 中的

· 最大的特征值对应的特征向量
· 第 2 大特征值对应的特征向量

画出散点图。

	坐标
面	（0.57, −0.60）
配料	（0.52,　0.79）
汤	（0.63　−0.11）



步骤 5 根据步骤 4，我们就可以得出第 1 主成分和第 2 主成分，如下所示。

$$
\begin{array}{ccccccc}
z_1 & = & 0.57u_1 & + & 0.52u_2 & + & 0.63u_3 \\
z_2 & = & -0.60u_1 & + & 0.79u_2 & - & 0.11u_3 \\
 & & \uparrow & & \uparrow & & \uparrow \\
 & & \text{“面”的标准值} & & \text{“配料”的标准值} & & \text{“汤”的标准值}
\end{array}
$$

最大的特征值对应的特征向量就是第 1 主成分的系数。同样，第 n 大特征值对应的特征向量就是第 n 大主成分的系数。

步骤 6 求出各样本在第 1 主成分和第 2 主成分中的坐标，也就是说，求出它们的第 1 主成分得分和第 2 主成分得分。

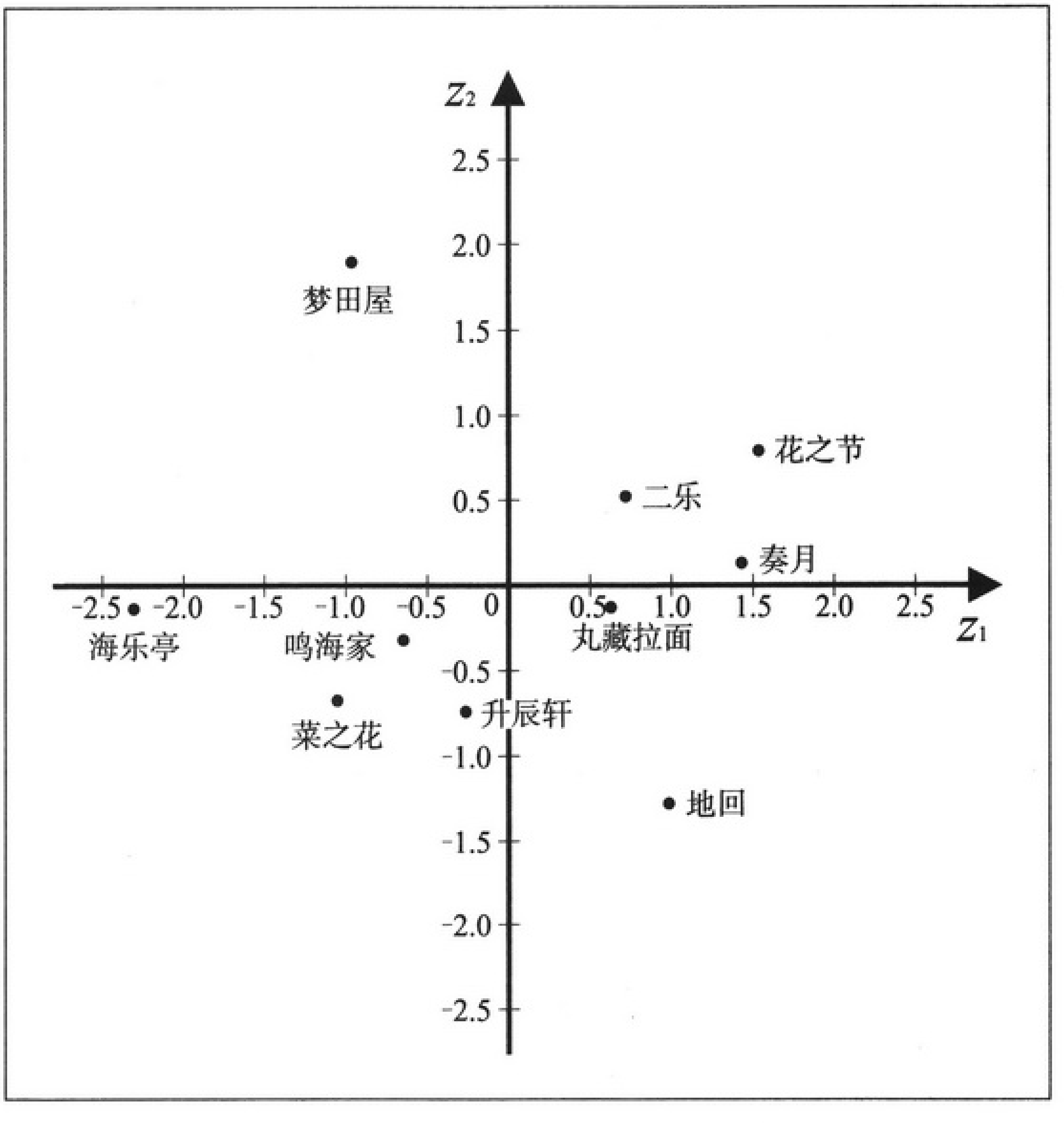
	“面”的标准值 u_1	“配料”的标准值 u_2	“汤”的标准值 u_3		第 1 主成分 z_1	第 2 主成分 z_2
二乐	−0.7	0.3	1.4	→	0.7	0.5
梦田屋	−1.3	1.2	−1.3	→	−1.0	1.9
地回	1.3	−0.5	0.8	→	1.0	−1.3
菜之花	−0.7	−1.4	0.1	→	−1.1	−0.7
花之节	0.0	1.2	1.4	→	1.5	0.8
升辰轩	0.7	−0.5	−0.6	→	−0.3	−0.7
丸藏拉面	0.7	0.3	0.1	→	0.6	−0.1
海乐亭	−1.3	−1.4	−1.3	→	−2.3	−0.1
鸣海家	0.0	−0.5	−0.6	→	−0.7	−0.3
秦月	1.3	1.2	0.1	→	1.4	0.1
平均值	0	0	0	→	0	0
标准差	1	1	1	→	$\sqrt{1.6}$	$\sqrt{0.8}$

$0.57 \times 1.3 + 0.52 \times 1.2 + 0.63 \times 0.1 = 1.4$

$\sqrt{\text{特征值}}$

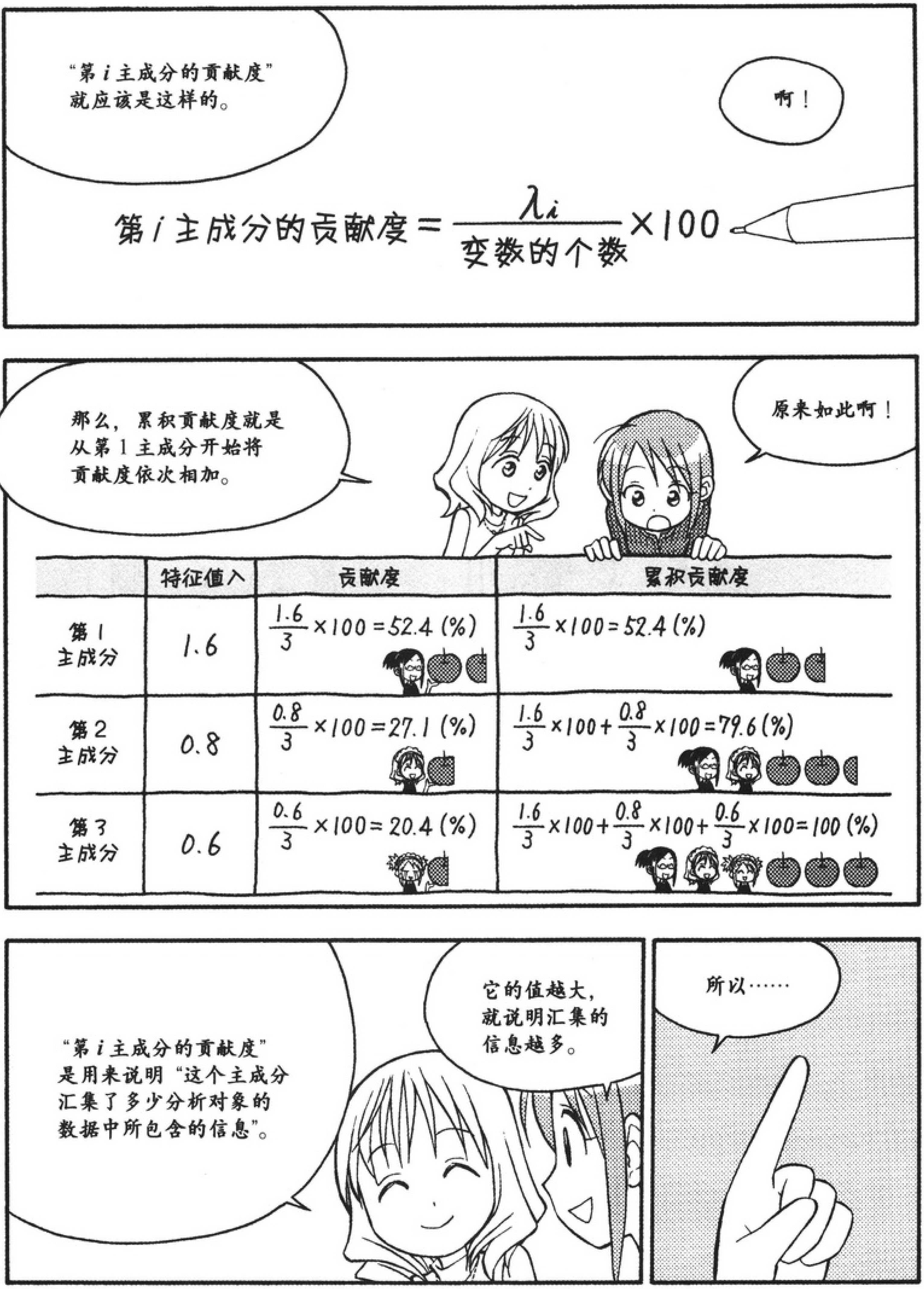
步骤 7　根据步骤 6 中的第 1 主成分得分和第 2 主成分得分画出散点图。

	坐标
二乐	(0.7, 0.5)
梦田屋	(−1.0, 1.9)
地回	(1.0, −1.3)
菜之花	(−1.1, −0.7)
花之节	(1.5, 0.8)
升辰轩	(−0.3, −0.7)
丸藏拉面	(0.6, −0.1)
海乐亭	(−2.3, −0.1)
鸣海家	(−0.7, −0.3)
秦月	(1.4, 0.1)



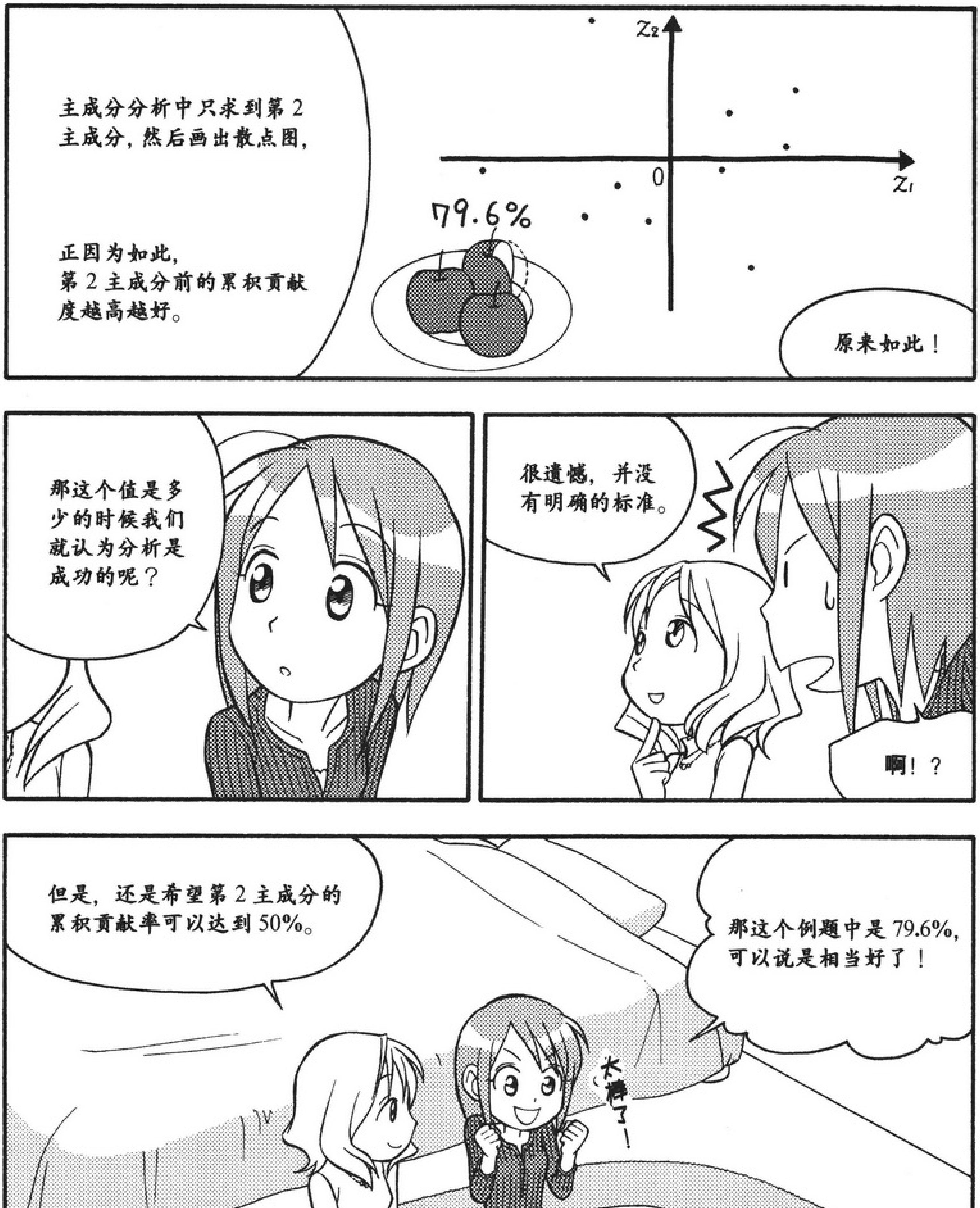
② 确认分析结果的精度

那么，来确认一下分析结果的精度吧！
好的！
主成分分析的成功与否，是通过“累积贡献度”的大小进行判断的。
那是什么？
在讲“累积贡献度”之前，我们先讲一下“贡献度”。
特征值λ
特征值向量
1.6
0.8
-0.60
0.79
-0.11
0.6
-0.55
-0.32
刚才在步骤3中是不是求出特征值了？
嗯！
在p个变量所组成的相关矩阵中，可以求出p个特征值。
p行p列的相关矩
$\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix}$
·特征值的个数是p
·所有特征值的和是p
所求出的特征值加在一起，和就是p。
哦。
那这个是3行3列，所以和是3。
没错！
也就是说，每个变量应该分得的特征值是1。
3人分
是这样啊！

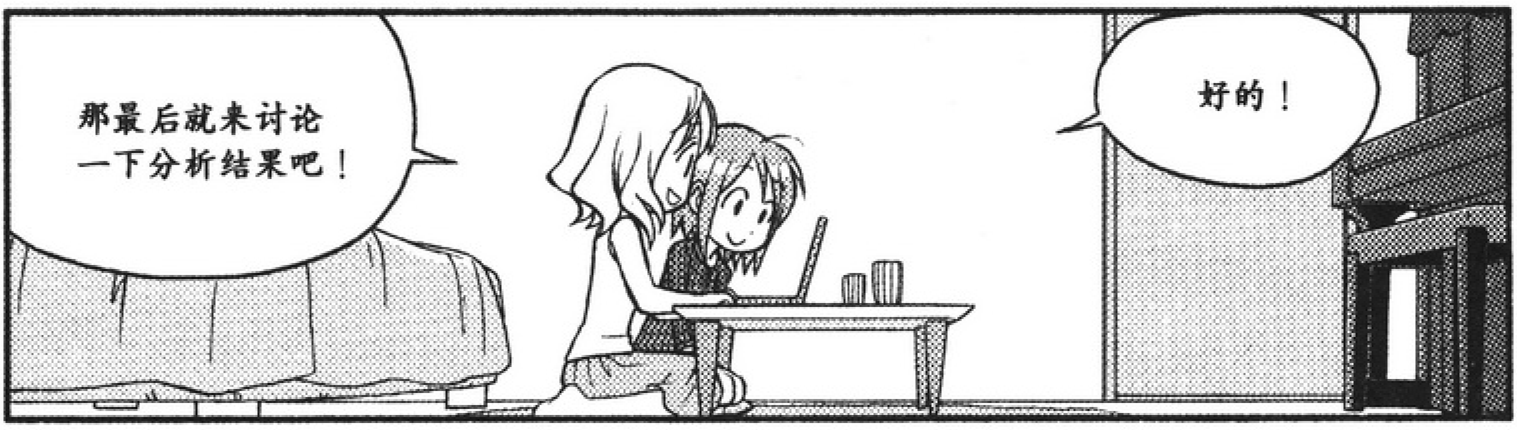

"第 i 主成分的贡献度"就应该是这样的。
啊！
$$第i主成分的贡献度=\frac{\lambda_i}{变数的个数}\times 100$$
那么，累积贡献度就是从第 1 主成分开始将贡献度依次相加。
原来如此啊！

	特征值λ	贡献度	累积贡献度
第1主成分	1.6	$\frac{1.6}{3}\times 100=52.4$ (%)	$\frac{1.6}{3}\times 100=52.4$ (%)
第2主成分	0.8	$\frac{0.8}{3}\times 100=27.1$ (%)	$\frac{1.6}{3}\times 100+\frac{0.8}{3}\times 100=79.6$ (%)
第3主成分	0.6	$\frac{0.6}{3}\times 100=20.4$ (%)	$\frac{1.6}{3}\times 100+\frac{0.8}{3}\times 100+\frac{0.6}{3}\times 100=100$ (%)

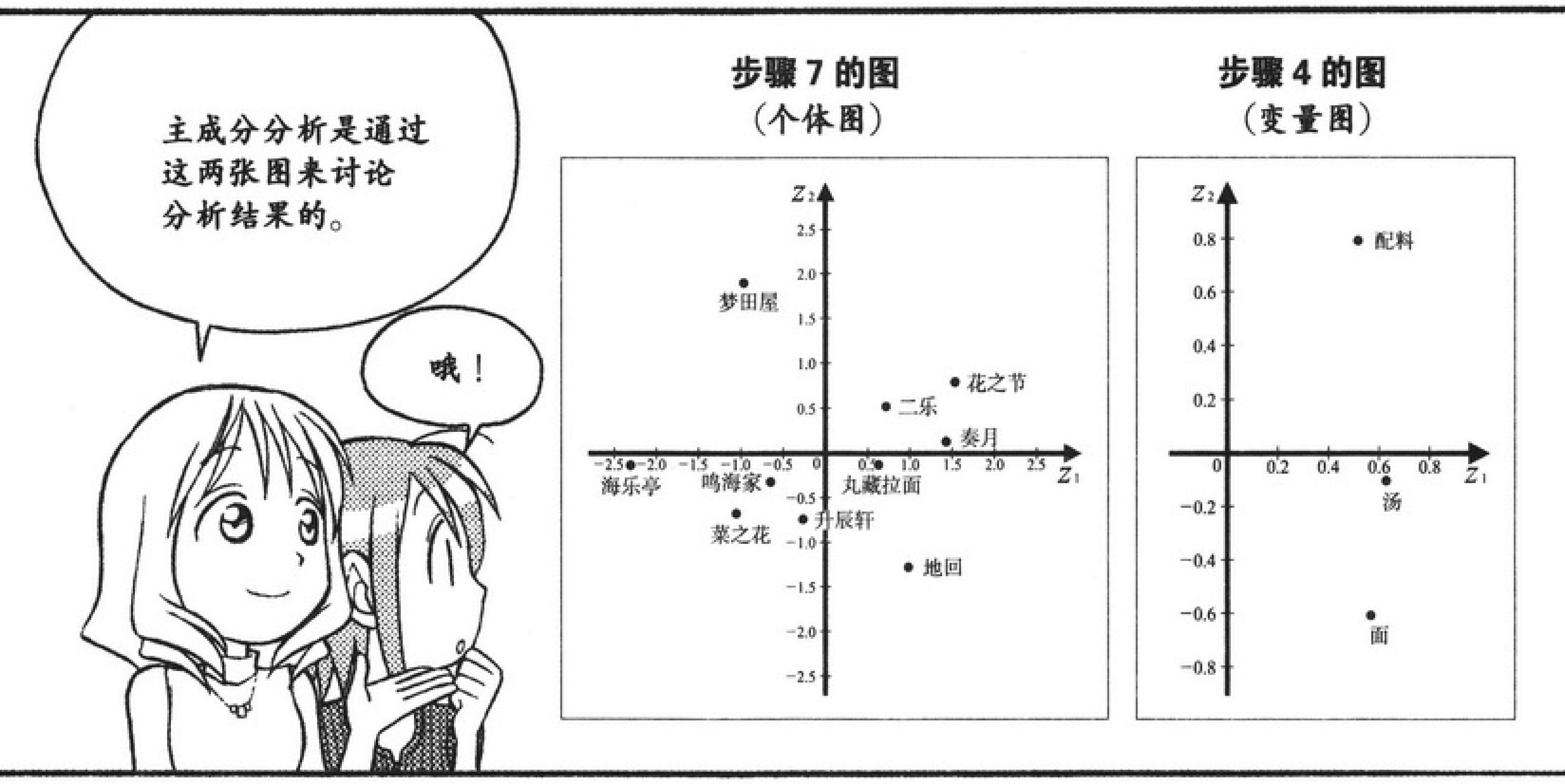
"第 i 主成分的贡献度"是用来说明"这个主成分汇集了多少分析对象的数据中所包含的信息"。
它的值越大，就说明汇集的信息越多。
所以……


主成分分析中只求到第2主成分，然后画出散点图，
正因为如此，第2主成分前的累积贡献度越高越好。
z_2
0
z_1
79.6%
原来如此！
那这个值是多少的时候我们就认为分析是成功的呢？
很遗憾，并没有明确的标准。
啊！？
但是，还是希望第2主成分的累积贡献率可以达到50%。
那这个例题中是79.6%，可以说是相当好了！
太棒了！

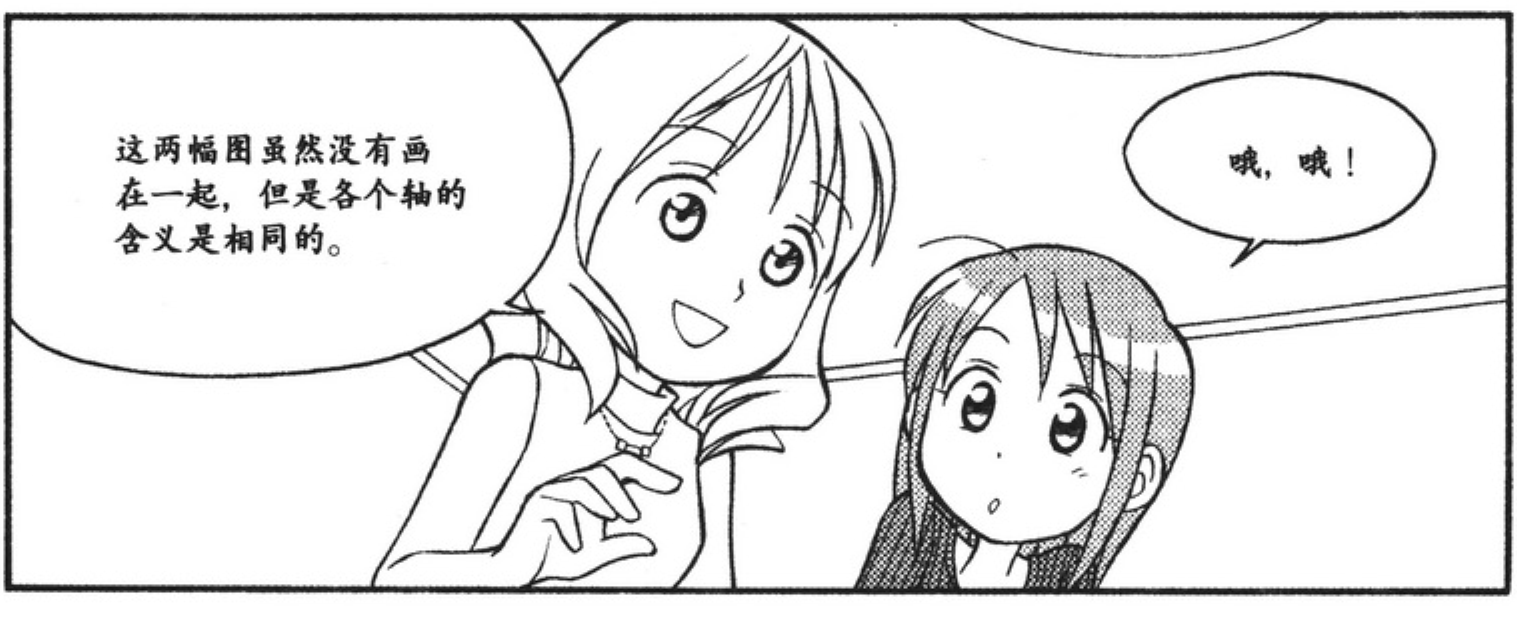
③ 讨论分析结果



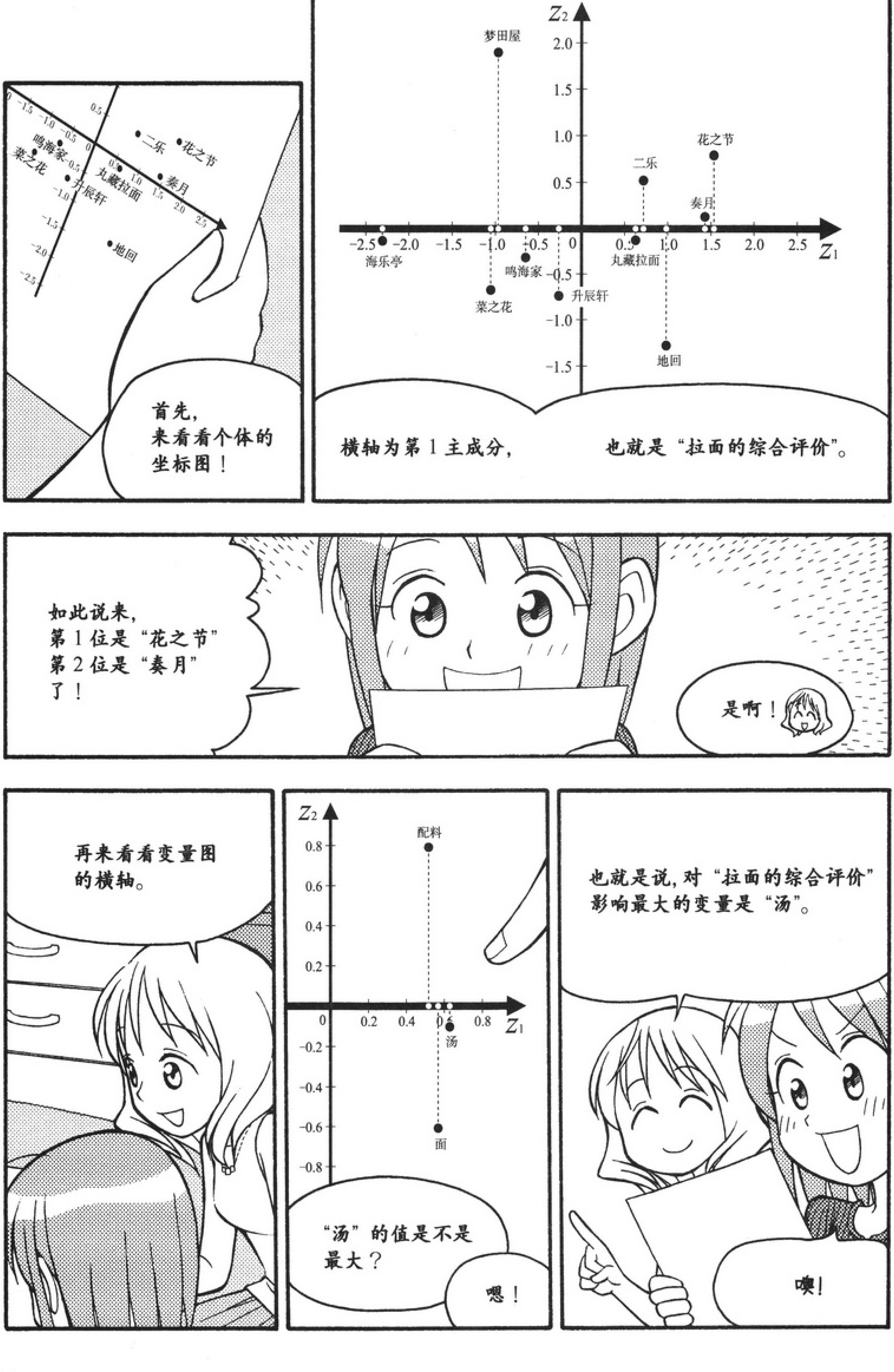
那最后就来讨论一下分析结果吧！
好的！



主成分分析是通过这两张图来讨论分析结果的。
哦！
步骤 7 的图
（个体图）
Z_2
Z_1
梦田屋
花之节
二乐
奏月
海乐亭
鸣海家
丸藏拉面
菜之花
升辰轩
地回
2.5
2.0
1.5
1.0
0.5
0
−0.5
−1.0
−1.5
−2.0
−2.5
步骤 4 的图
（变量图）
Z_2
Z_1
配料
汤
面
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8

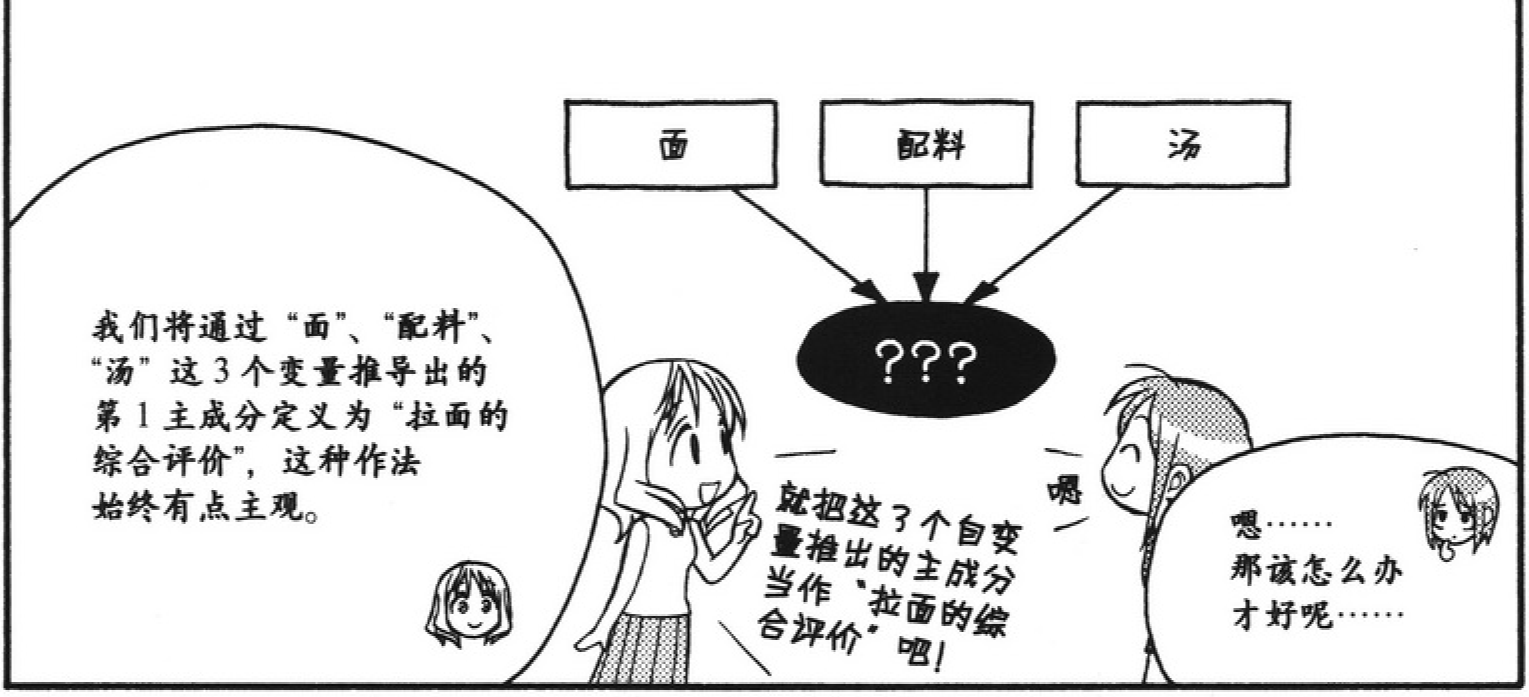


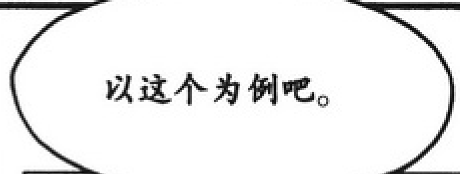
这两幅图虽然没有画在一起，但是各个轴的含义是相同的。
哦，哦！


首先，
来看看个体的
坐标图！
Z_2
梦田屋
花之节
二乐
奏月
海乐亭
鸣海家
丸藏拉面
菜之花
升辰轩
地回
Z_1
横轴为第1主成分，
也就是“拉面的综合评价”。
如此说来，
第1位是“花之节”
第2位是“奏月”
了！
是啊！
再来看看变量图
的横轴。
Z_2
配料
汤
面
Z_1
“汤”的值是不是
最大？
嗯！
也就是说，对“拉面的综合评价”
影响最大的变量是“汤”。
噢！

最后，再比较一下
这两张图。
上方是不是
有个“梦田屋”啊？
嗯，“拉面的综合评价”不是很好
……
Z_2
Z_1
梦田屋
花之节
二乐
秦月
海乐亭
鸣海家
丸藏拉面
菜之花
升辰轩
地回
配料
汤
面
但是，它的“配料”得到了应有的评价。
唉？
下方的“地回”，
它的面不也得
到相应的评价
了吗？
真的呀！
第 2 主成分的信息
也是可以利用的。
是那样的！
怎么样？
学会主成分分析了吗？
是的！
非常有用啊！
你这是
在写什么呢？
“花之节”的地
点和营业时间
……

✿ 4. 变量的选择和第 1 主成分 ✿

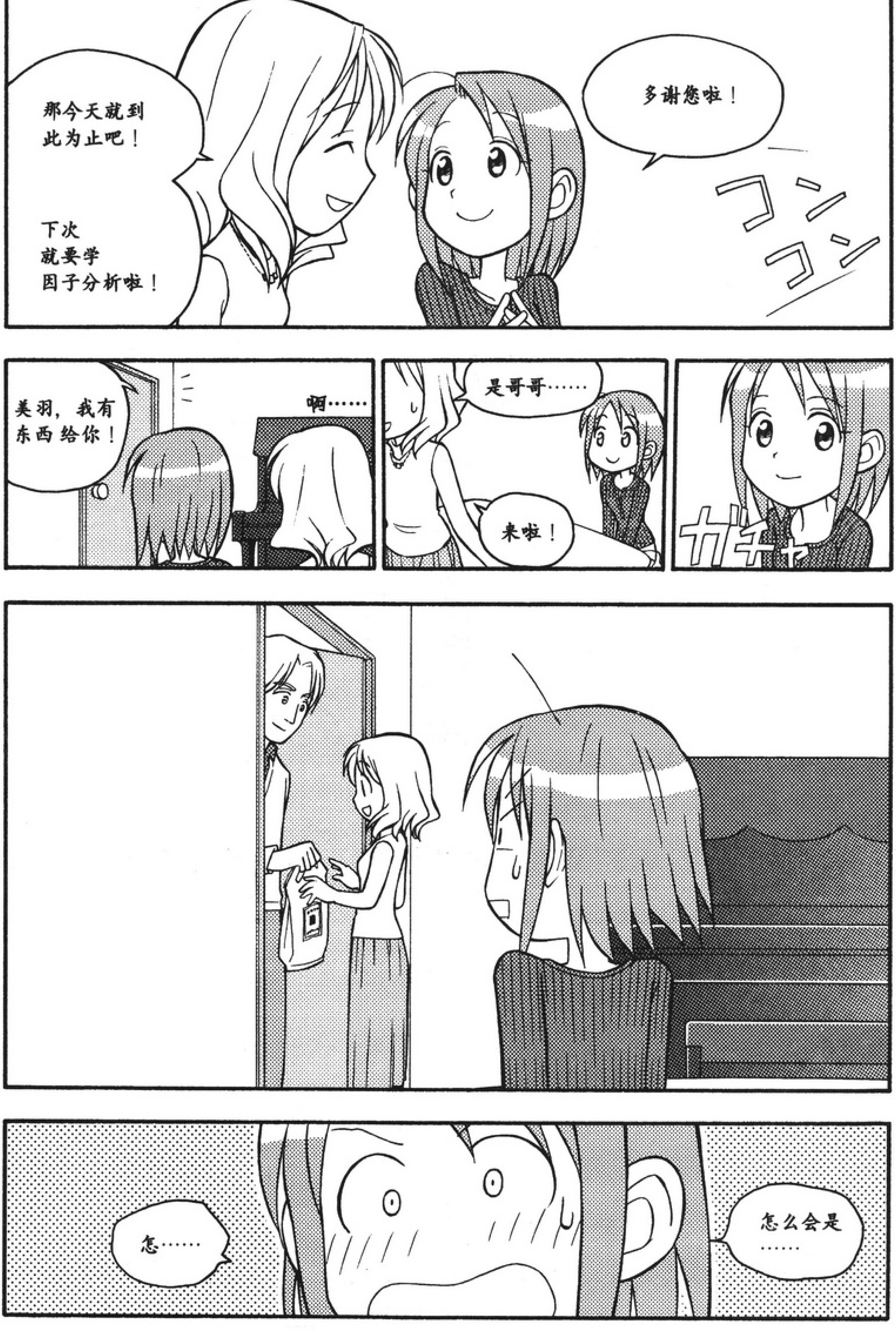




补习班离家近的孩子，下课后不用在路上耽搁时间，这样就会有较多的时间用于复习，学习成绩就应该好。我们在对“语文”、“社会”、“理科”、“英语”和“数学”进行主成分分析时，应该加上“从家到学校的距离的倒数”这一自变量，再将所得的第1主成分定义为“综合成绩”，才更为妥当。

	语文	社会	理科	英语	数学	从家到学校的距离的倒数
太郎	42	62	26	4	20	1/1200
花子	12	28	42	8	84	1/580
⋮	⋮	⋮	⋮	⋮	⋮	⋮

啊……“综合成绩”里还要加入从家到补习班的距离？
如果分析者这样认为的话，也是可以的啊！
不会是什么都可以用来做主成分分析吧……
那可不是！
如果不详细说明为何要用这些变量进行分析，那结果能令周围的人信服吗？
那还有不上补习班的学生呢？
虽然远但是有车接送。
うっ
那从家到学校的距离不用考虑吗？
是这样啊……
也可以说，这是一种与分析者关联性很强的分析方法。
噢，是这样啊……


那今天就到
此为止吧！
下次
就要学
因子分析啦！
多谢您啦！
コンコン
美羽，我有
东西给你！
啊……
是哥哥……
来啦！
怎……
怎么会是
……


五十岚！
嗯？
啊？
你是……嗯……
你们认识!?
カチ
コチ
我，我是高津露儿。
ニコッ
啊！高津部长的千金啊！
以前去你家的时候，我们聊过的！
是……
啊……露儿竟然认识我哥哥！
美羽师姐居然和五十岚是兄妹，太让人吃惊了……
那你们慢慢聊！
那，那个！


您知道山本什么时候回来吗？
山本？
就是哥哥公司的……
戴眼镜的
啊，一同入职的山本啊！他去国外了啊！
什么时候……
事情好像很多，还真不太好说！
ずーん
这、这样啊……
具体的情况我也不是很清楚，不好意思……
啊，别这么说，我应该谢谢您才对。
再见啦！
再见……
别灰心啊！露儿！
看哥哥给我送好吃的来了！
嗯……
这个冰激凌很好吃的。

✿ 5. 第 1 主成分和综合实力 ✿

在此前的讲解中，我们认为第 1 主成分一定能够表示“综合实力”。但事实并非如此。

请思考下面的例子。下表中记录的数据，是对某中学 3 年级学生的调查结果。

◆表4.1　调查结果的数据

	理科（分）	数学（分）	一天中添饭的次数（餐 / 日）
A	77	82	3
B	68	66	1
C	93	81	2
D	100	92	5
D	75	70	0

对这组数据进行主成分分析，可以求出它的第 1 主成分是：

$$z_1 = \underbrace{0.56u_1}_{\text{“理科”的标准值}} + \underbrace{0.60u_2}_{\text{“数学”的标准值}} + \underbrace{0.57u_3}_{\text{“一天中添饭的次数”的标准值}}$$

那么，这个第 1 主成分究竟表示哪种综合实力呢？只要用常识性思维去想一下，就知道它不能表示任何综合实力。

再举个例子。如果要用主成分分析来判断您的“综合运动能力”。那么就有，

① 无论是“握力”还是“最喜爱的电视节目”，无论什么变量都不假思索地进行收集。

② 对这些变量进行主成分分析。

按照这样的顺序求出的第 1 主成分，怎么能表示您的“综合运动能力”呢？

① 收集计算“综合运动能力”的相关变量。

② 对这些变量进行主成分分析。

要是按照这样的顺序求出第 1 主成分，不就能表示您的“综合运动能力”了吗？

总之，并不是说只要做了主成分分析，所求出的第 1 主成分就自然表示综合实力。分析者要依据所要探求的“综合 XX 实力”，收集相关变量，并对这些变量进行主成分分析，只有这样才能说所得到的第 1 主成分表示“综合 XX 实力”。也就是说，以烹饪炖菜为例，并不是说“只要将萝卜、青椒以及其他手边所有能找到的食材都放进锅里炖，就自然能做出炖菜”，而是“烹饪者依据所要做的炖菜，收集相应的食材，再进行烹饪，只有这样才能做出炖菜”。

附带说一下。

将足够的咖喱粉、泡菜和鱿鱼干放在一起炖，您就做出炖菜了。那个真的是炖菜吗？当然，您硬要说“这就是炖菜！”也没关系。但是，您周围的人坚持认为这和炖菜还差得远呢，非但如此，还会惊讶地看着您说：“你这家伙在说什么胡话呢？”到底如何炖才是真正的炖菜呢？不是至少要放多少鱿鱼干，更何况这也没有一个统一的答案，而是要靠厨师的聪明才智来判断，这样解释起来才较为妥当。以上所说的情况同样适用于第 4 节“变量的选择和第 1 主成分”。

✿ 6. 累积贡献度的标准 ✿

正如之前所说的那样，主成分分析的结果通常用二元散点图来进行表示。因此，只要第 2 主成分前的累积贡献度的值足够大，就可以比较确定地认为，这个散点图汇集了相当多的分析对象的数据中所包含的信息。我们就会自信地认为“这个分析很成功”。

遗憾的是，像“第 2 主成分前的累积贡献度达到 XX% 以上，就认为这个分析是成功的”这样的统计学标准并不存在。在 114 页出现的“50%”那个所谓的标准，其所基于的观点就是“如果没有汇集到一半分析对象的数据中所包含的信息，那这个散点图就很难说是有意义的”，这也只不过是笔者自己的想法。

下面再提出一个可能令读者困惑的问题。首先，请您以任意两个变量为对象进行主成分分析。您就会发现第 2 主成分前的累积贡献率一定是 100%。接下来，再请您以 200 个变量为对象进行主成分分析。只要不是一定量的数据，第 2 主成分前的累积贡献度的值是不会达到 50% 的。由此可见，第 2 主成分前的累积贡献度的值是否能达到 50%，很大程度上是依赖于分析对象的变量个数的。

虽然说在实际操作中并不存在累积贡献度的标准，但要知道它的标准并不是随意的。一般不会有人见到“第 2 主成分前的累积贡献度的值为 14%”这样的分析结果后，还对此分析持肯定态度。毕竟分析者也不好意思让周围的人看到这样的分析结果。

笔者建议您在按照本章的内容进行分析的同时，努力找到适合具体情况的标准。

✿ 7. 第 2 主成分及之后的主成分 ✿

如果读者并不是很关心数学原理的话，可以跳过本节内容，不做阅读。

我们之前讲过，第 2 主成分及之后的主成分和分析者的意图无关，是在数学上自动地被求出来的。其实，这么说也并不恰当。

实际上，第 2 主成分是分析者在下述前提的约束下求出来的轴[1]。

· 同第 1 主成分垂直正交。
· 通过数据方差第 2 大（仅次于第 1 主成分）处的轴。

实际上，第 3 主成分是分析者在下述前提的约束下求出来的轴[2]。

· 同第 1 主成分和第 2 主成分垂直正交。
· 通过数据方差第 3 大（仅次于第 1 主成分和第 2 主成分）处的轴。

也就是说，第 2 主成分及之后的主成分，并不是“和分析者的意图无关，是在数学上自动地被求出来的”，而应当是“在分析者加入了前提约束下求出来的”，分析者如果不加约束就求不出来了。

---

1. 换句话说，第 2 主成分是“先要同第 1 主成分垂直正交，并且通过数据方差第 2 大处的轴”。

2. 换句话说，第 3 主成分是“先要同第 1 主成分及第 2 主成分垂直正交，并且通过数据方差第 3 大处的轴”。同样地，第 i 主成分就是“另一个垂直正交的主成分，并且通过数据方差第 i 大处的轴”。

也许读者会认为，分析者不必刻意去提醒自己“要加约束条件”，使用数据分析软件自然就可以将第 2 主成分及之后的主成分求出。事实的确如此，数据分析软件在一开始便加入了之前所讲的那些约束条件，然后才进行计算的，这样做看上去多此一举，但却是一个非常符合实际情况的方法。

✿ 8. 方差和特征值 ✿

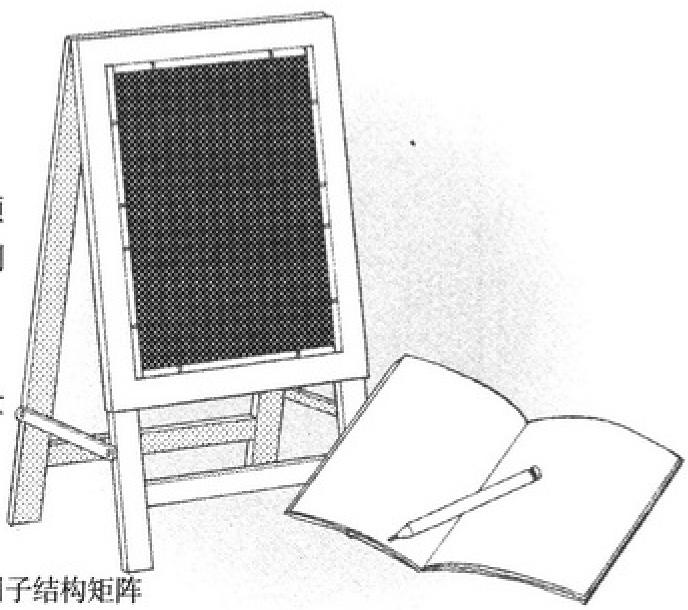
如果读者并不是很关心数学原理的话，可以跳过本节内容，不做阅读。

正如之前所讲，第 1 主成分是“通过数据方差最大处的轴”。请您根据这句话再观察一下第 107 页到第 111 页中主成分分析的计算过程。但在求解第 1 主成分时，有关数据方差的计算并没有完全给出，取而代之的却是如何计算特征值和特征向量的有关内容。

事实上，“求出通过数据方差最大处的轴”同“求出相关矩阵中最大的特征值所对应的特征向量”是一回事，这一点我们在数学上暂且不做详细介绍。同样地，“求出通过数据方差第 i 大处的轴”同“求出相关矩阵第 i 大特征值所对应的特征向量”也是一回事。

第 5 章

因子分析



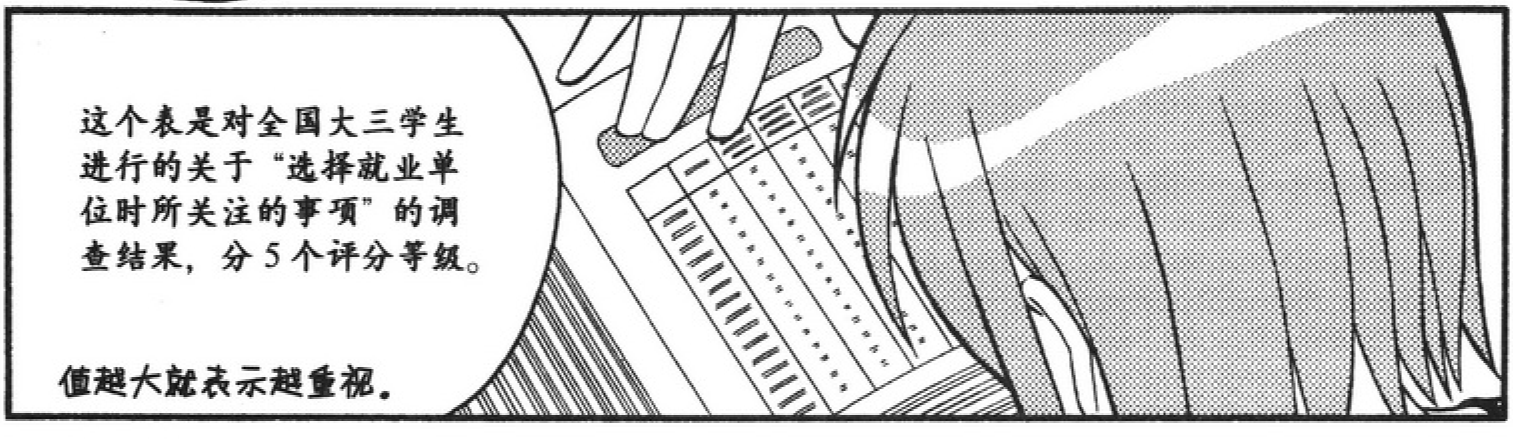
ぐにっ
山本……
山本
过得好吗？
忙吗？
有没有按时
吃饭啊？
真担心你啊……
呜……
露儿，
再去添乱
多不好
啊！
添乱……
ぐっ……
我一定会去看你的
……
ガタッ
啊……呀！
拼了命也要学会
因子分析……
书籍扫描：铜板+西瓜


第二天。
今天就要讲露儿你期待已久的因子分析啦！
请您多多指教！
ぺこ
今天我们要学习因子分析，不过，
我想把之前露儿花了一番工夫做出来的诺伦调查问卷再重新做一下！
这回就用那份调查问卷进行分析，
然后再动手做一下因子分析，
是！
这可是最后关头啦，小露儿你要加油啊！
嗯！
那就先来看看这本杂志。

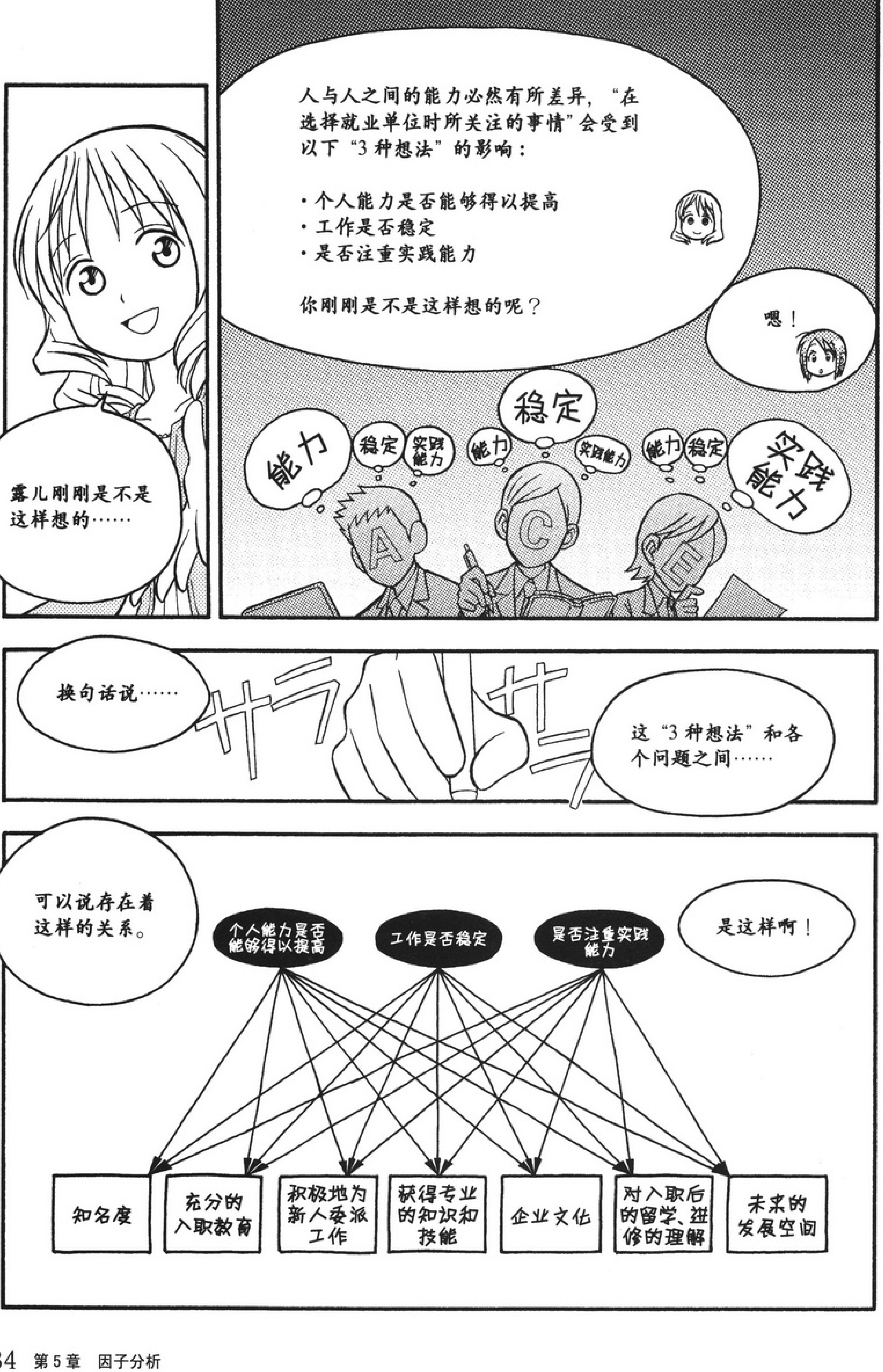
1. 因子分析

	知名度	充分的入职教育	积极地为新人委派工作	获得专业的知识和技能	企业文化	对入职后的留学、进修的理解	未来的发展空间
A	2	5	1	5	1	5	2
B	3	4	2	5	3	4	1
C	4	1	1	2	5	3	2
D	4	1	3	2	5	3	4
E	1	2	5	1	2	1	4
F	5	1	1	1	4	2	2
G	4	1	1	2	3	2	2
H	3	3	3	4	4	5	4
I	3	2	4	3	5	3	5
J	3	1	2	2	4	3	3
K	2	2	3	2	3	1	1
L	4	3	2	3	5	3	3
M	2	1	1	2	3	3	1
N	3	1	1	1	4	2	2
O	3	3	2	4	4	5	3

看看这个！



看一下 A 君的数据。
在"是否可以提高个人能力"的相关问题给出了较高的评分。
获得专业的知识和技能
对入职后的留学、进修的理解
充分的入职教育
这样说来，还真是……
B 的情况也类似。
那再看看 C 是怎么想的呢？
知名度
企业文化
嗯……
是有志于去大企业吧！
在"工作是否稳定"的相关问题中给出了较高的评分。
D 也是这样的。
那 E 呢？
未来的发展空间
嗯……
在"是否注重实践能力"的相关问题给出较高评分。
积极地为新人委派工作
你对数据相当敏锐啊！
是从山本那学来的吧？
也没有啦……
呵呵……

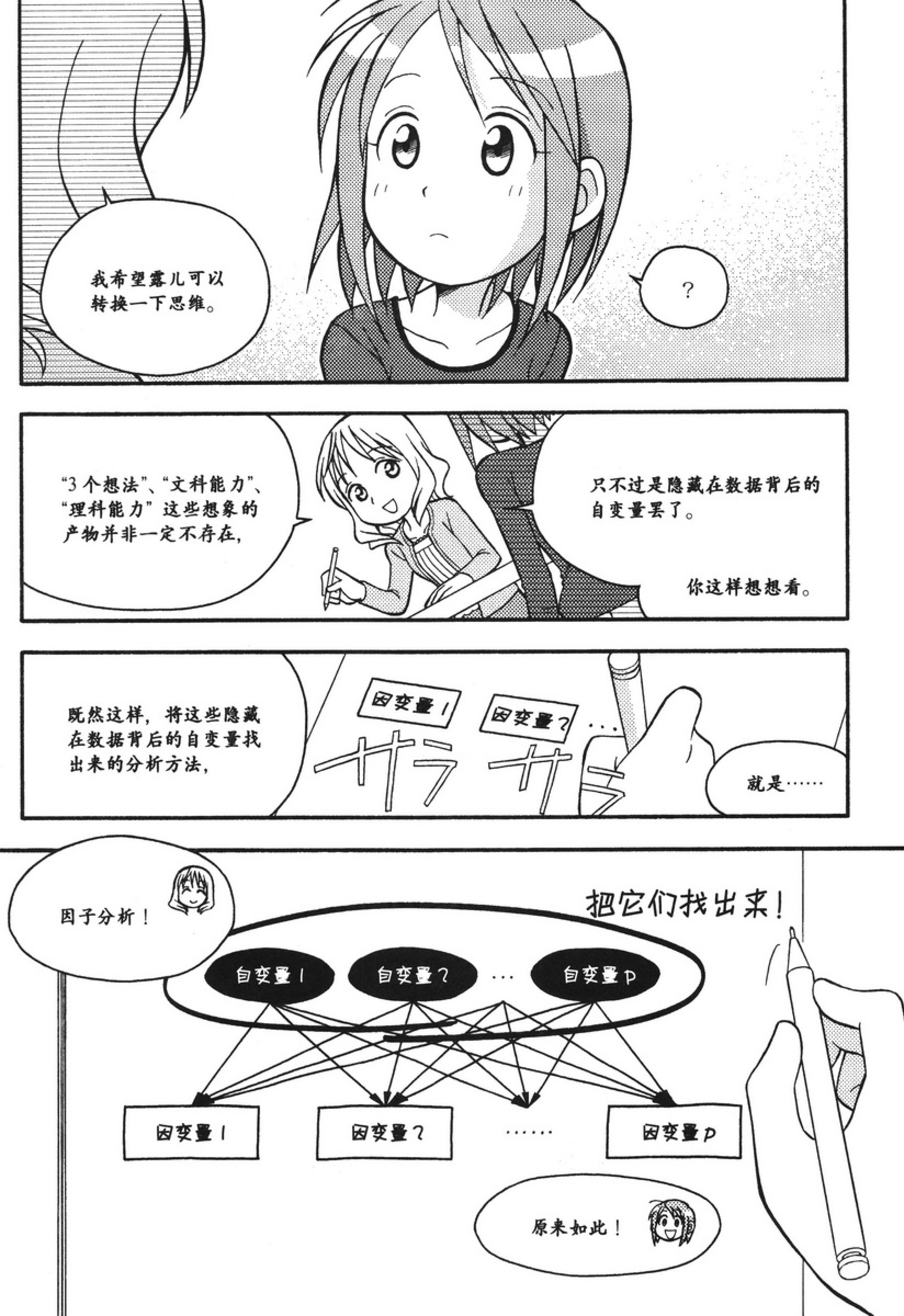

人与人之间的能力必然有所差异，“在选择就业单位时所关注的事情”会受到以下“3种想法”的影响：
・个人能力是否能够得以提高
・工作是否稳定
・是否注重实践能力
你刚刚是不是这样想的呢？
嗯！
能力
稳定
实践能力
能力
稳定
实践能力
能力
稳定
实践能力
露儿刚刚是不是这样想的……
换句话说……
サラ
サー
这“3种想法”和各个问题之间……
可以说存在着这样的关系。
是这样啊！
个人能力是否能够得以提高
工作是否稳定
是否注重实践能力
知名度
充分的入职教育
积极地为新人委派工作
获得专业的知识和技能
企业文化
对入职后的留学、进修的理解
未来的发展空间

这是某个补习班上高三学生的
5门课的测验结果。

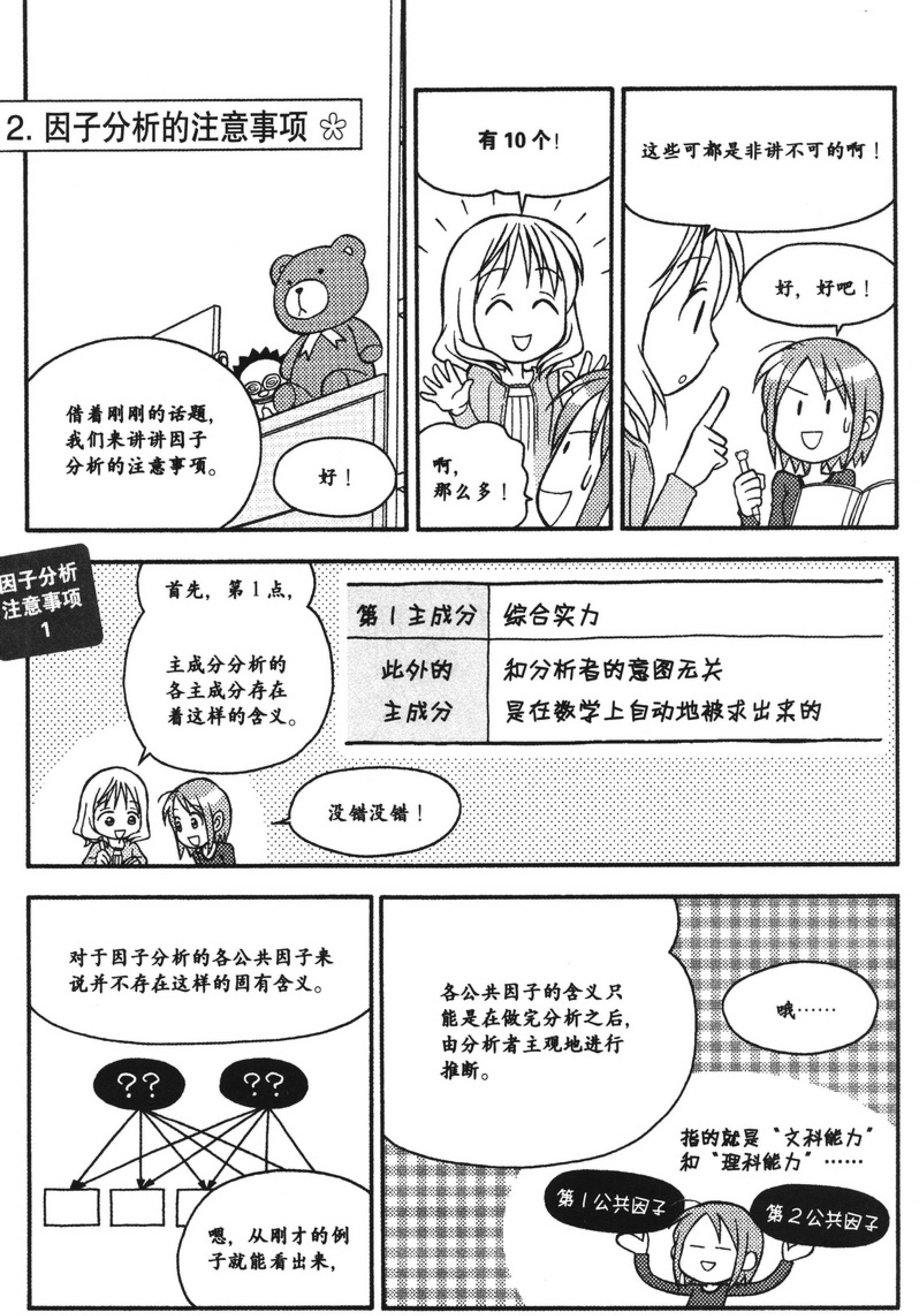
噢！

	语文	社会	理科	英语	数学
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

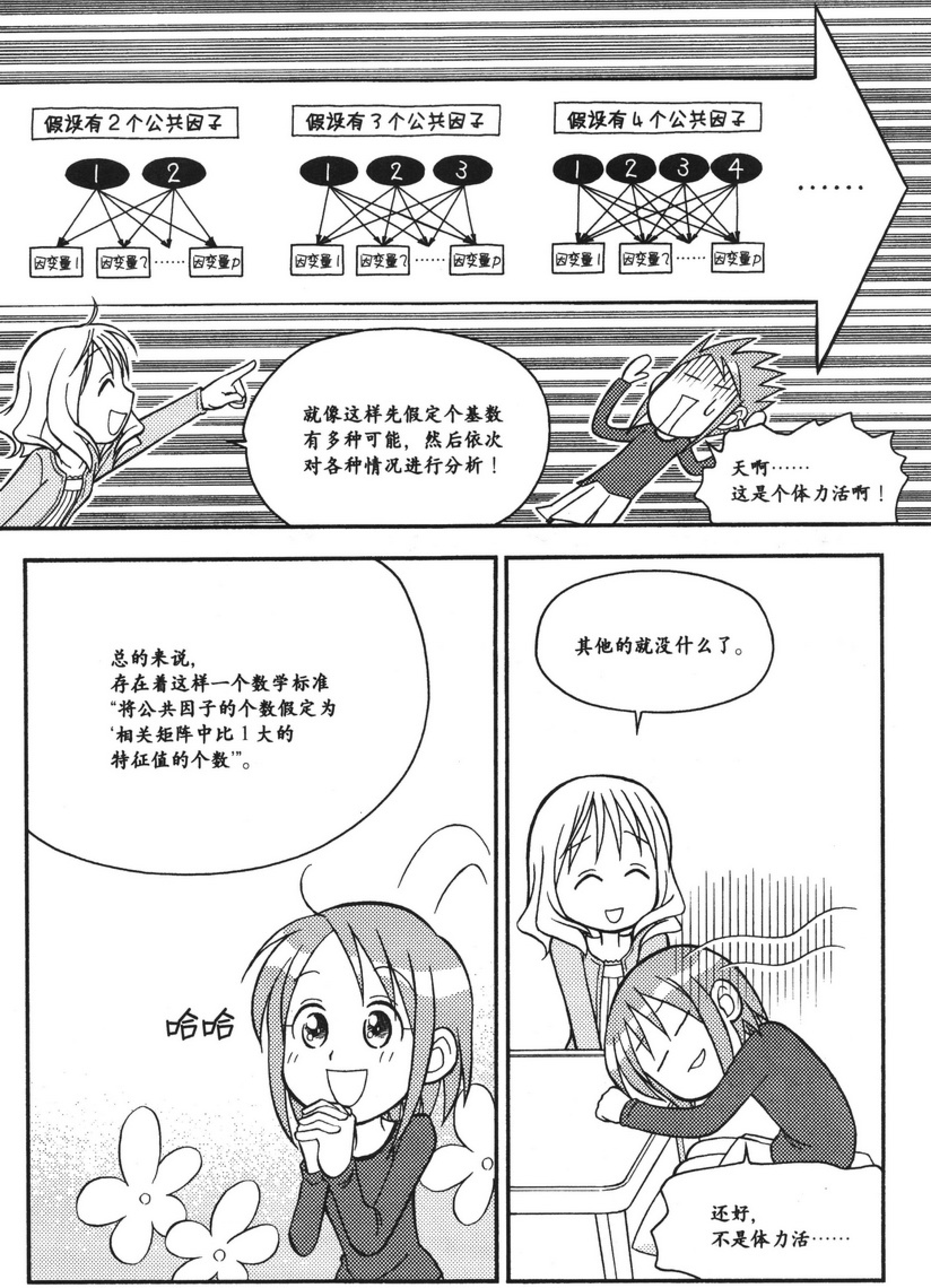
换句话说，它们之间是不是可以说存在着这样的关系？
文科能力
理科能力
语文
社会
理科
英语
数学
嗯……
但是，还是觉得有点不对劲……
为什么呢？
“3 个想法”、“文科能力”、“理科能力”这些……
是不是真的存在呢？这个就不得而知……
嗯，它们都是想像的产物。
用这些虚构的东西来解释数据也没关系吗？
但是，要是我们假定它们就隐藏在数据背后的话，不就能对数据的特征进行合理的解释了吗？
原来是这样啊，不过……


我希望露儿可以转换一下思维。
？
“3个想法”、“文科能力”、“理科能力”这些想象的产物并非一定不存在，
只不过是隐藏在数据背后的自变量罢了。
你这样想想看。
既然这样，将这些隐藏在数据背后的自变量找出来的分析方法，
因变量1
因变量2
……
サラ サラ
就是……
因子分析！
把它们找出来！
自变量1
自变量2
…
自变量p
因变量1
因变量2
……
因变量p
原来如此！

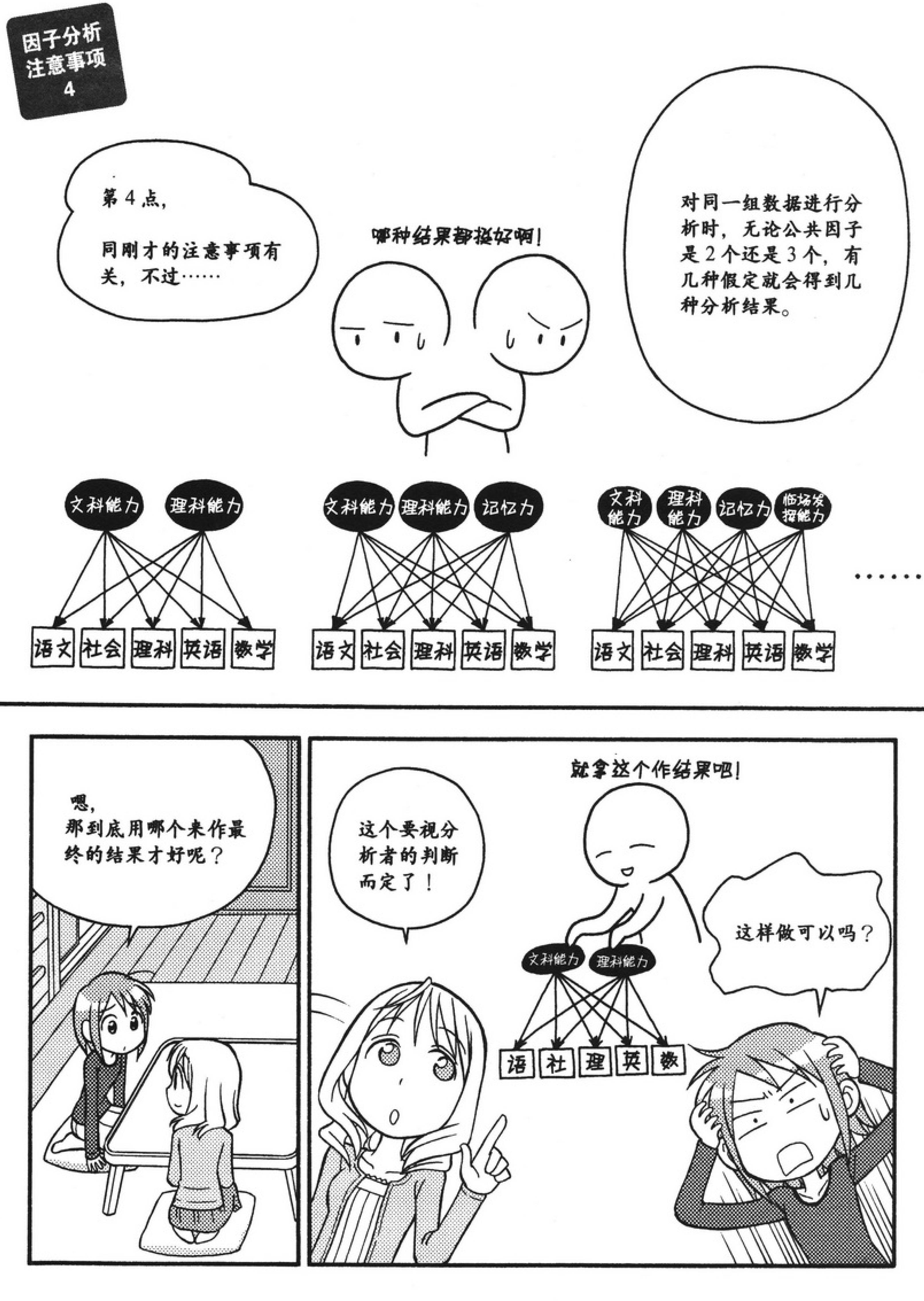
顺便讲一下，第i个自变量被称为“第i公共因子”或者“第i因子”或者“因子i”。
第1公共因子
第2公共因子
文科能力
理科能力
每个个体中的公共因子的具体值被称为“因子得分”。
是！
文科能力
理科能力
A
XX
XX
B
XX
XX
C
XX
XX
为了慎重起见，我们还是先来明确一下因子分析和主成分分析的差别。
嗯！
像这样，无论是思考方法还是结构图都完全不同！
主成分分析
自变量1
自变量2
……
自变量p
因变量1
因变量2
因子分析
因变量1
因变量2
…
因变量p
自变量1
自变量2
……
自变量p

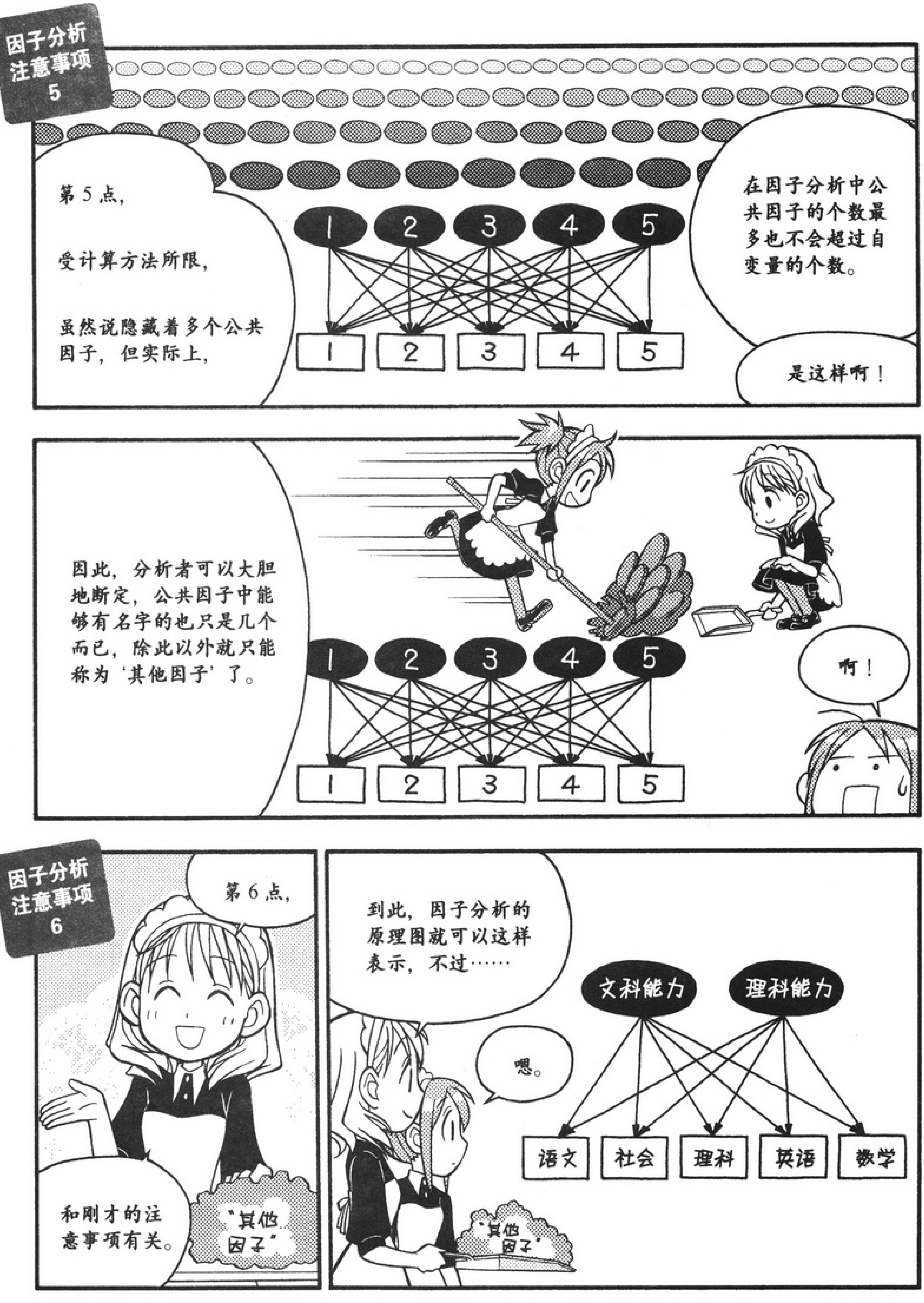

2. 因子分析的注意事项
借着刚刚的话题，我们来讲讲因子分析的注意事项。
好！
有 10 个！
啊，那么多！
这些可都是非讲不可的啊！
好，好吧！
因子分析注意事项 1
首先，第 1 点，
主成分分析的各主成分存在着这样的含义。
第 1 主成分　综合实力
此外的主成分　和分析者的意图无关是在数学上自动地被求出来的
没错没错！
对于因子分析的各公共因子来说并不存在这样的固有含义。
??
??
嗯，从刚才的例子就能看出来，
各公共因子的含义只能是在做完分析之后，由分析者主观地进行推断。
哦……
指的就是“文科能力”和“理科能力”……
第 1 公共因子
第 2 公共因子


因子分析注意事项2
第1主成分
第2点，
在主成分分析中，存在着这样的顺序：“第1主成分”、“第2主成分”……
第2主成分
3
4
5
但是，在因子分析中并没有这样的排序，各公共因子地位平等。
这个也能从刚刚的例子中看出来！
因子分析注意事项3
第3点，
因子分析的计算，难点在于分析者必须在分析前先假定出公共因子的个数。
啊!?
要在分析前想象出隐藏着几个公共因子，这简直是不可能的啊！
要怎么做呢!?

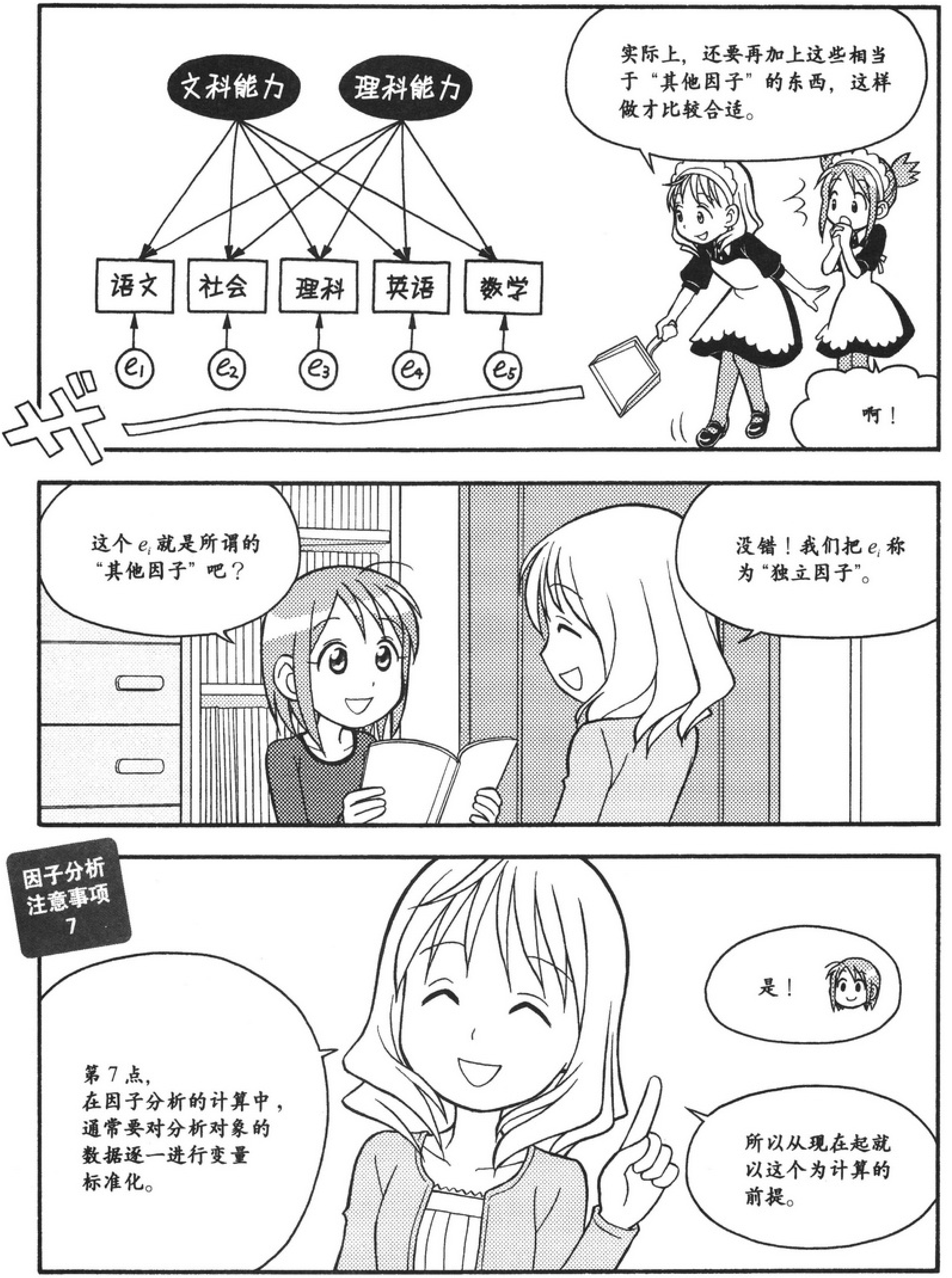


假设有2个公共因子
1 2
因变量1 因变量2 …… 因变量p
假设有3个公共因子
1 2 3
因变量1 因变量2 …… 因变量p
假设有4个公共因子
1 2 3 4
因变量1 因变量2 …… 因变量p
……
就像这样先假定个基数有多种可能，然后依次对各种情况进行分析！
天啊……这是个体力活啊！
总的来说，存在着这样一个数学标准“将公共因子的个数假定为‘相关矩阵中比1大的特征值的个数’”。
哈哈
其他的就没什么了。
还好，不是体力活……


因子分析
注意事项
4
第4点，
同刚才的注意事项有关，不过……
哪种结果都挺好啊！
对同一组数据进行分析时，无论公共因子是2个还是3个，有几种假定就会得到几种分析结果。
文科能力 理科能力
语文 社会 理科 英语 数学
文科能力 理科能力 记忆力
语文 社会 理科 英语 数学
文科能力 理科能力 记忆力 临场发挥能力
语文 社会 理科 英语 数学
……
嗯，
那到底用哪个来作最终的结果才好呢？
这个要视分析者的判断而定了！
就拿这个作结果吧！
文科能力 理科能力
语 社 理 英 数
这样做可以吗？

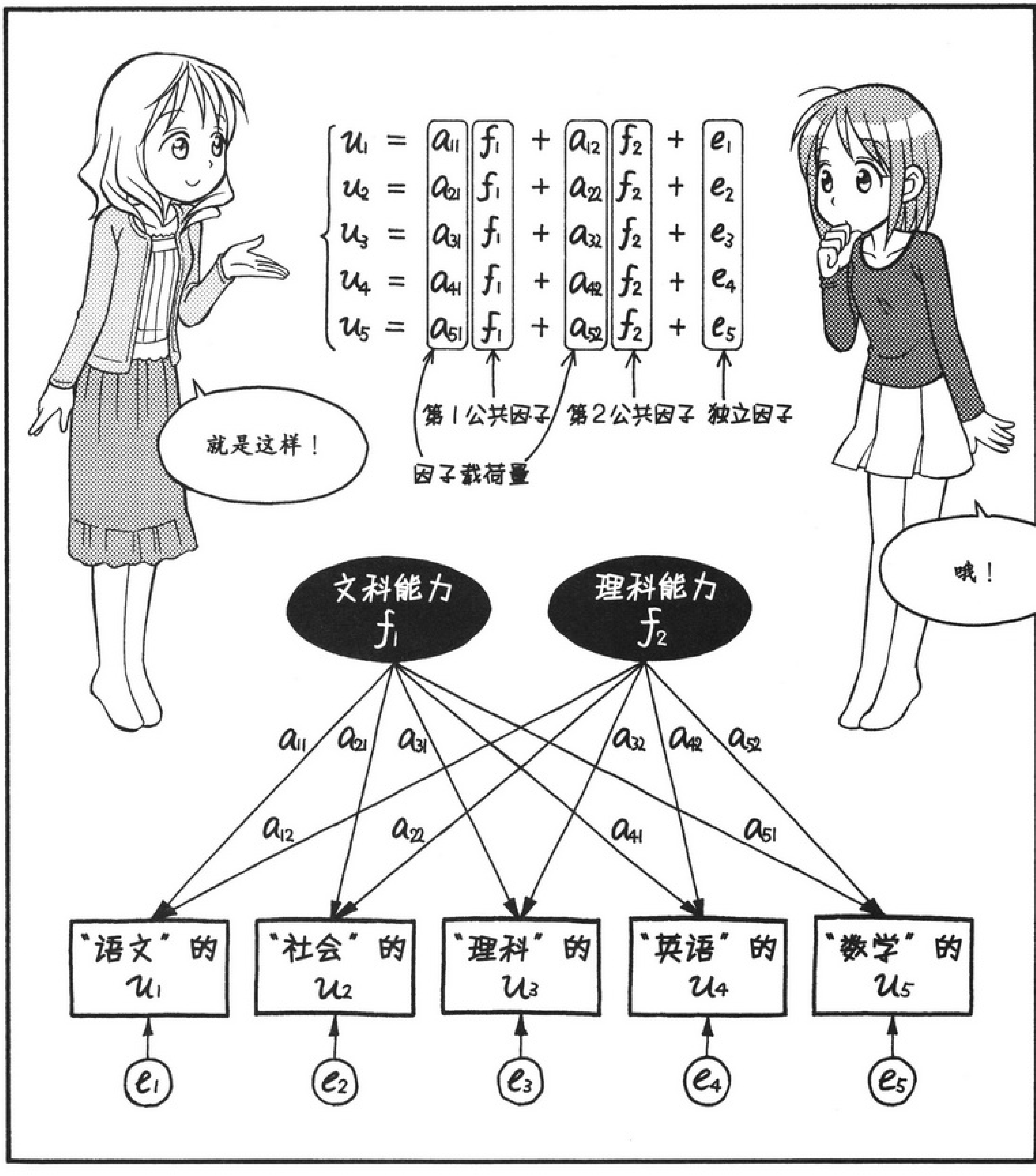


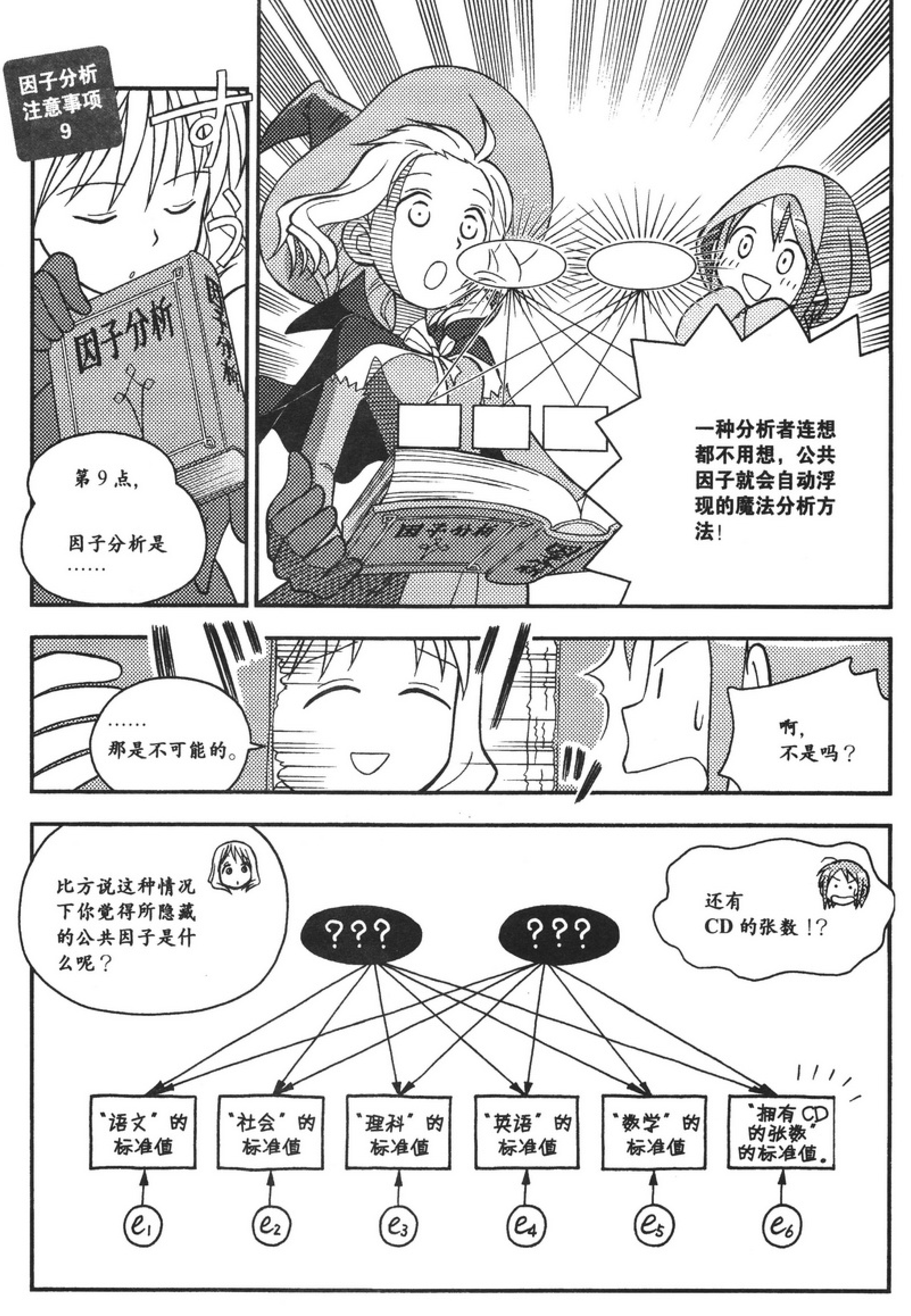
因子分析
注意事项
5
第 5 点，
受计算方法所限，
虽然说隐藏着多个公共因子，但实际上，
在因子分析中公共因子的个数最多也不会超过自变量的个数。
1 2 3 4 5
1 2 3 4 5
是这样啊！
因此，分析者可以大胆地断定，公共因子中能够有名字的也只是几个而已，除此以外就只能称为‘其他因子’了。
1 2 3 4 5
1 2 3 4 5
啊！
因子分析
注意事项
6
第 6 点，
和刚才的注意事项有关。
“其他因子”
到此，因子分析的原理图就可以这样表示，不过……
嗯。
文科能力
理科能力
语文
社会
理科
英语
数学
“其他因子”

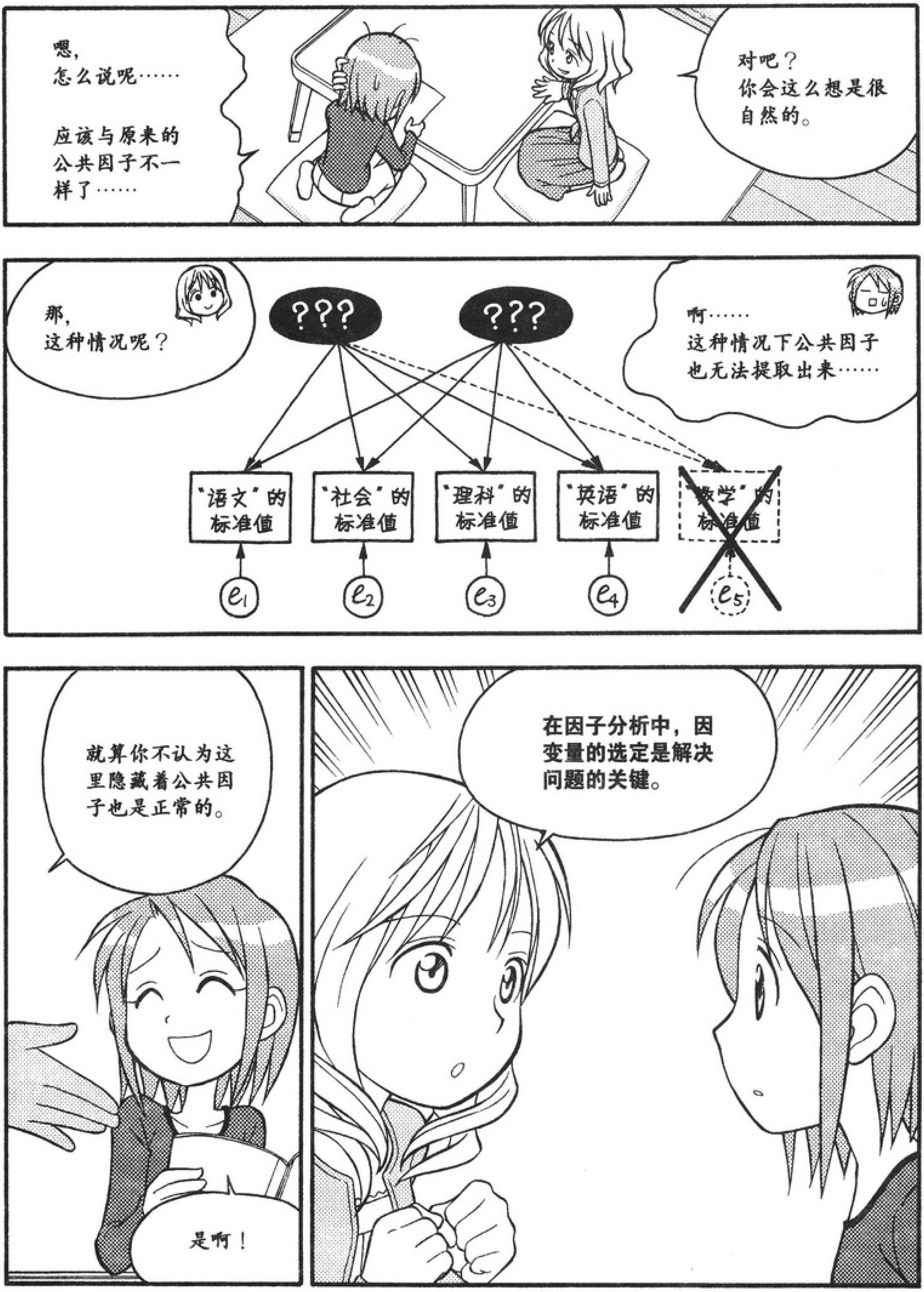


文科能力
理科能力
语文
社会
理科
英语
数学
e_1
e_2
e_3
e_4
e_5
实际上，还要再加上这些相当于“其他因子”的东西，这样做才比较合适。
啊！
这个 e_i 就是所谓的“其他因子”吧？
没错！我们把 e_i 称为“独立因子”。
因子分析注意事项7
第7点，在因子分析的计算中，通常要对分析对象的数据逐一进行变量标准化。
是！
所以从现在起就以这个为计算的前提。

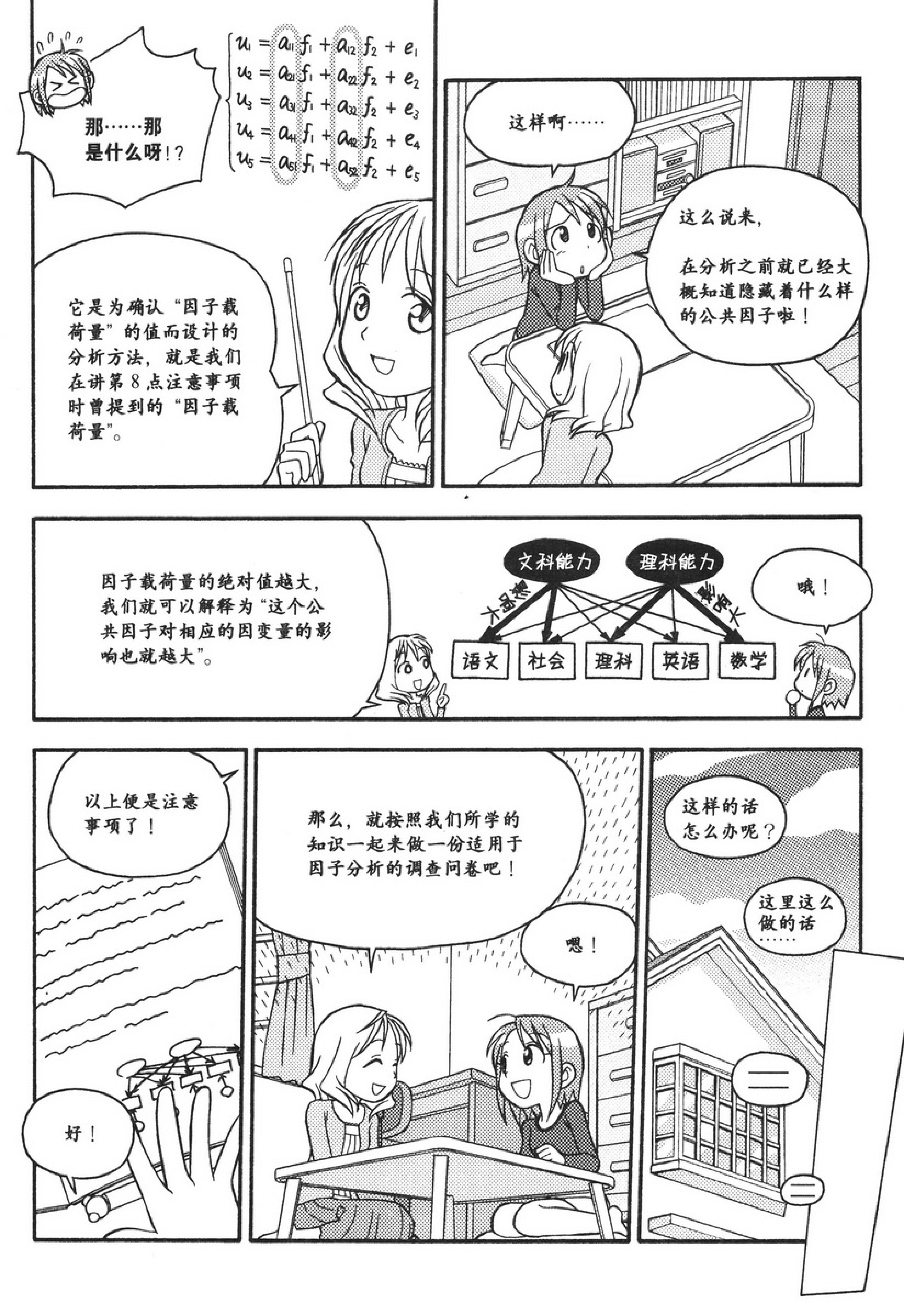

因子分析
注意事项
8
第 8 点，
用式子或者图形来表示
因子分析的结构……


u1 = a11 f1 + a12 f2 + e1
u2 = a21 f1 + a22 f2 + e2
u3 = a31 f1 + a32 f2 + e3
u4 = a41 f1 + a42 f2 + e4
u5 = a51 f1 + a52 f2 + e5
第 1 公共因子
第 2 公共因子
独立因子
因子载荷量
就是这样！
哦！
文科能力 f1
理科能力 f2
a11 a21 a31 a32 a42 a52
a12 a22 a41 a51
“语文”的 u1
“社会”的 u2
“理科”的 u3
“英语”的 u4
“数学”的 u5
e1
e2
e3
e4
e5


因子分析
注意事项
9
第9点，
因子分析是
……
一种分析者连想都不用想，公共因子就会自动浮现的魔法分析方法！
……
那是不可能的。
啊，
不是吗？
比方说这种情况下你觉得所隐藏的公共因子是什么呢？
还有
CD的张数！？
???
???
"语文"的标准值
"社会"的标准值
"理科"的标准值
"英语"的标准值
"数学"的标准值
"拥有CD的张数"的标准值。
e_1
e_2
e_3
e_4
e_5
e_6


嗯，
怎么说呢……
应该与原来的公共因子不一样了……
对吧？
你会这么想是很自然的。
那，
这种情况呢？
???
???
"语文"的标准值
"社会"的标准值
"理科"的标准值
"英语"的标准值
"数学"的标准值
e_1
e_2
e_3
e_4
e_5
啊……
这种情况下公共因子也无法提取出来……
就算你不认为这里隐藏着公共因子也是正常的。
是啊！
在因子分析中，因变量的选定是解决问题的关键。


也就是说，假设“这些因变量的背后隐藏着这样的公共因子”……
知名度
充分的入职教育
积极地为新人委派工作
获得专业的知识和技能
企业文化
对入职后的留学、进修的理解
未来的发展空间
只有当这个假设在一定程度上成立时，因子分析才能顺利进行。
这个……
不正是因为不知道隐藏着什么样的公共因子才要做因子分析的吗？
不能事先预测出一定程度的结果，就不能进行有效的分析……
还真是难办啊！
因子分析注意事项10
这正是我们最后要讲的注意事项了！
实际上，因子分析并不是为了找出背后隐藏的公共因子的分析方法。
えーー、えっ!?



那……那是什么呀!?
$u_1 = a_{11} f_1 + a_{12} f_2 + e_1$
$u_2 = a_{21} f_1 + a_{22} f_2 + e_2$
$u_3 = a_{31} f_1 + a_{32} f_2 + e_3$
$u_4 = a_{41} f_1 + a_{42} f_2 + e_4$
$u_5 = a_{51} f_1 + a_{52} f_2 + e_5$
它是为确认“因子载荷量”的值而设计的分析方法，就是我们在讲第8点注意事项时曾提到的“因子载荷量”。
这样啊……
这么说来，在分析之前就已经大概知道隐藏着什么样的公共因子啦！
因子载荷量的绝对值越大，我们就可以解释为“这个公共因子对相应的因变量的影响也就越大”。
文科能力
理科能力
影响大
影响大
语文
社会
理科
英语
数学
哦！
以上便是注意事项了！
好！
那么，就按照我们所学的知识一起来做一份适用于因子分析的调查问卷吧！
嗯！
这样的话怎么办呢？
这里这么做的话……


欢迎光临！
3. 因子分析的具体实例
您方便帮我们做一份调查问卷吗？
好啊！
这个会在您临走的时候收回。
好的！
嗨，给你！
十分感谢！


顾客调查问卷

问卷调查的目的是为了将诺伦经营得更好，请您务必协助。
参与此次调查的顾客将可获得我们赠送的优惠券，请在结账时索取！

■ 关于本店的问题

Q1. 请问您对本店的印象如何？（限选一项）

	很不好	不好	一般	好	很好
a. 店面设计	1	2	3	4	5
b. 店内环境	1	2	3	4	5
c. 女服务生的服务态度	1	2	3	4	5
d. 红茶的味道	1	2	3	4	5
e. 红茶的价格	1	2	3	4	5
f. 茶杯的美感	1	2	3	4	5

■ 关于顾客自身的问题

Q2. 请问您的性别是（限选一项）

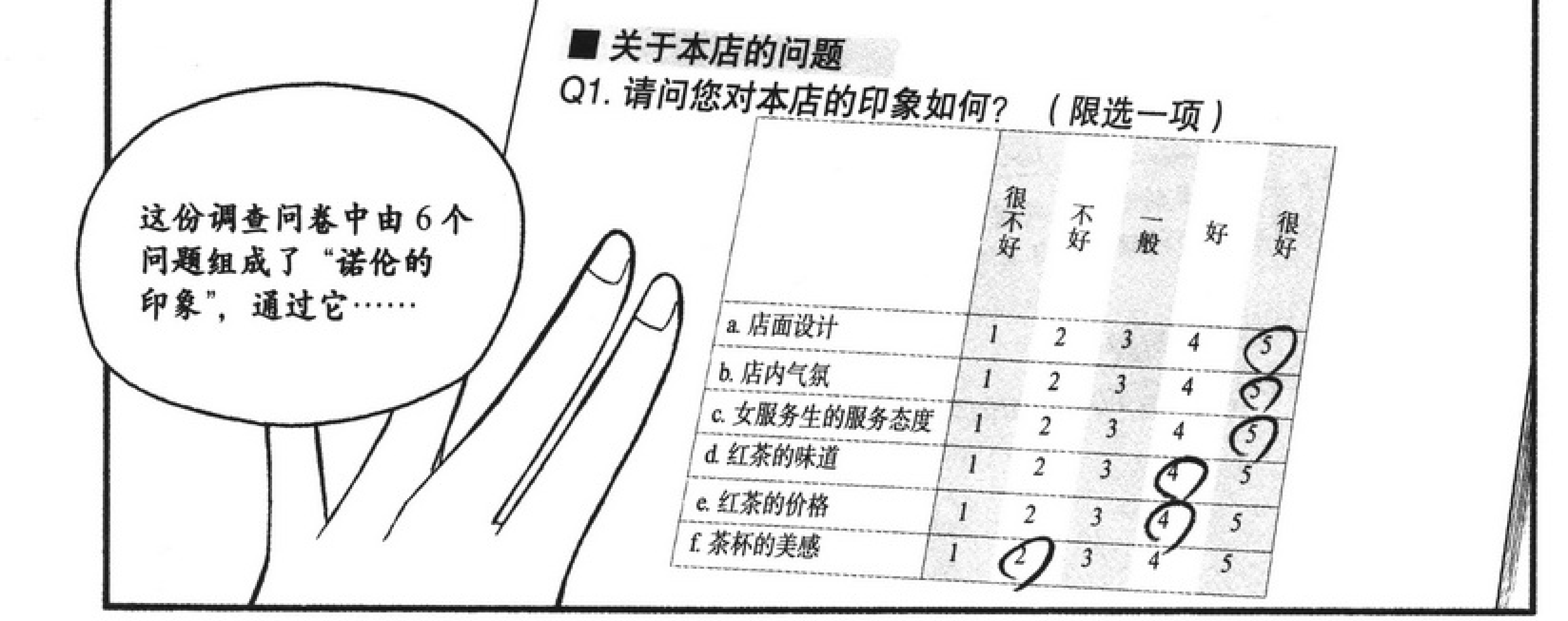
1. 男性　2. 女性

Q3. 请问您的年龄是

21 岁

Q4. 请问您的职业是

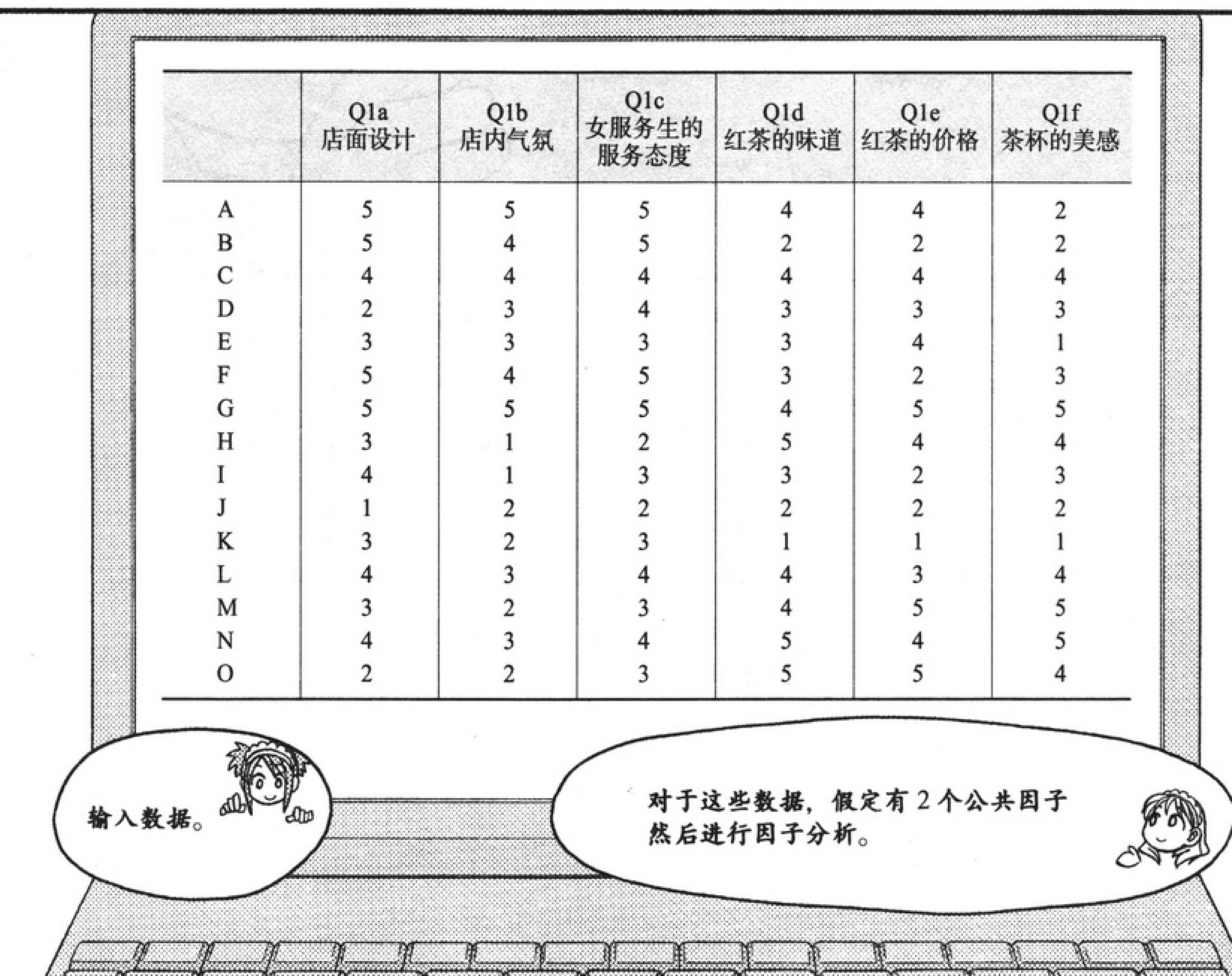
1. 职员　2. 个体经营



■关于本店的问题

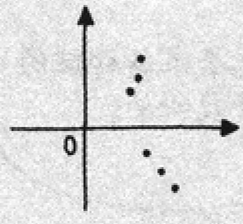
Q1. 请问您对本店的印象如何？（限选一项）

	很不好	不好	一般	好	很好
a. 店面设计	1	2	3	4	5
b. 店内气氛	1	2	3	4	5
c. 女服务生的服务态度	1	2	3	4	5
d. 红茶的味道	1	2	3	4	5
e. 红茶的价格	1	2	3	4	5
f. 茶杯的美感	1	2	3	4	5

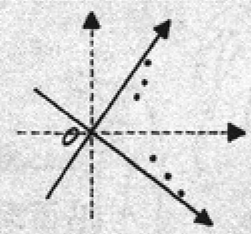


	Q1a 店面设计	Q1b 店内气氛	Q1c 女服务生的服务态度	Q1d 红茶的味道	Q1e 红茶的价格	Q1f 茶杯的美感
A	5	5	5	4	4	2
B	5	4	5	2	2	2
C	4	4	4	4	4	4
D	2	3	4	3	3	3
E	3	3	3	3	4	1
F	5	4	5	3	2	3
G	5	5	5	4	5	5
H	3	1	2	5	4	4
I	4	1	3	3	2	3
J	1	2	2	2	2	2
K	3	2	3	1	1	1
L	4	3	4	4	3	4
M	3	2	3	4	5	5
N	4	3	4	5	4	5
O	2	2	3	5	5	4

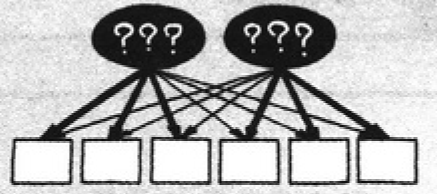
① 求解旋转前的因子载荷量。



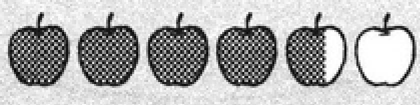
② 求解旋转后的因子载荷量。



③ 解释各公共因子的含义。

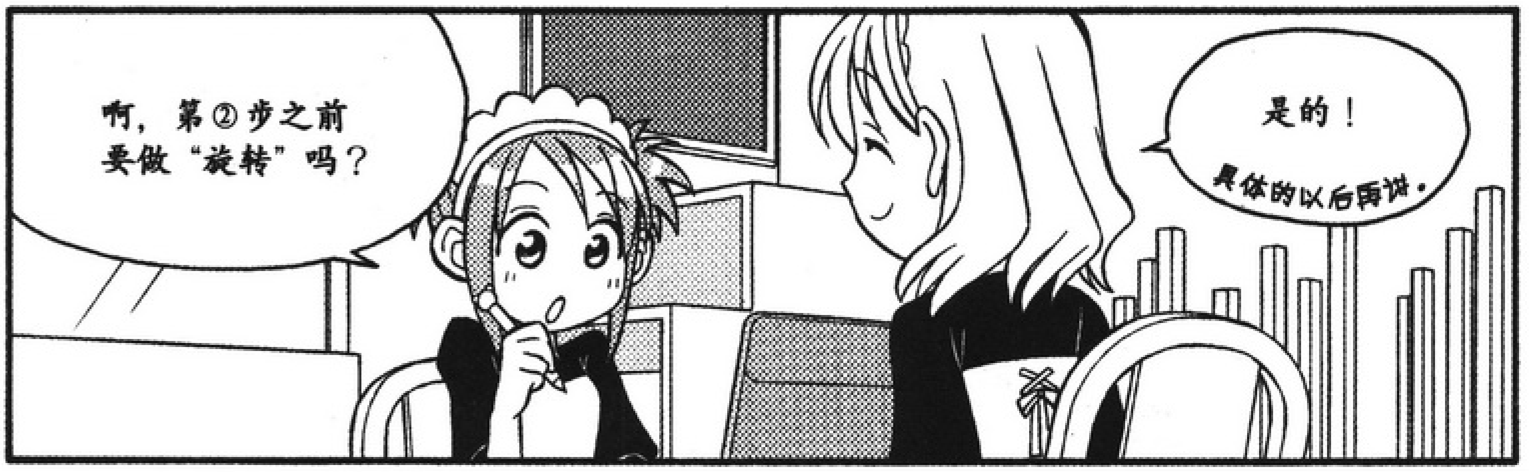


④ 确认分析结果的精度。

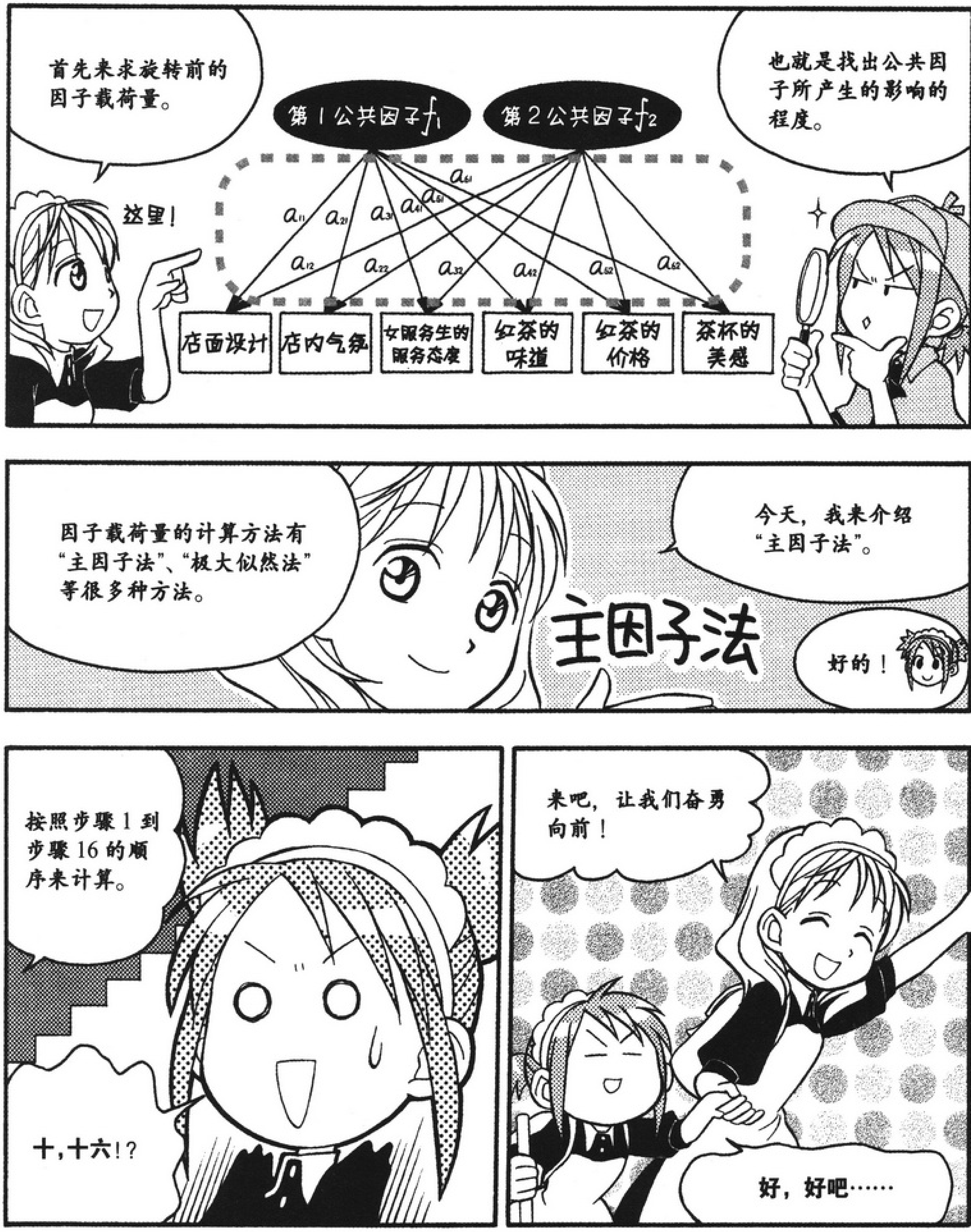


⑤ 求出因子得分，充分理解每个个体的特征。

	第1公共因子	第2公共因子
A	XX	XX
B	XX	XX
C	XX	XX
⋮	⋮	⋮



① 求解旋转前的因子载荷量。



步骤 1　逐一进行变量标准化。

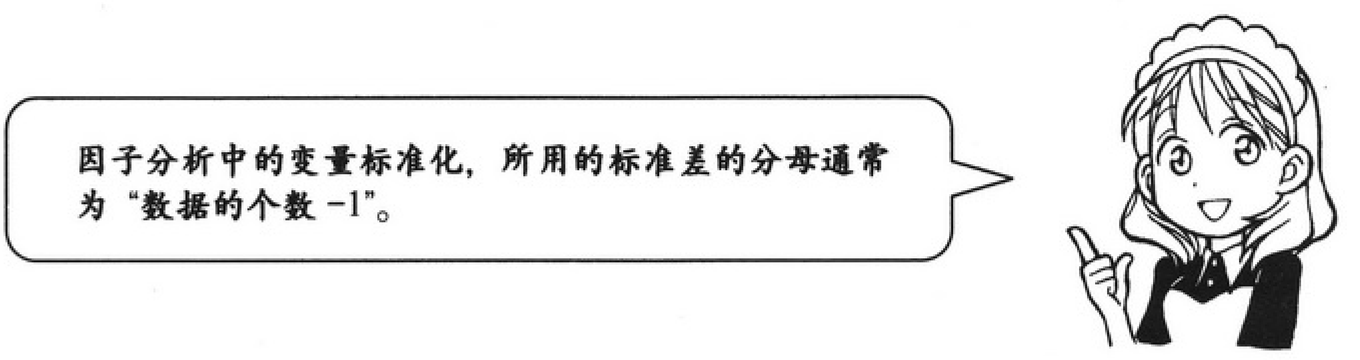
	Q1a 店面设计	…	Q1f 茶杯的美感
A	5	…	2
B	5	…	2
C	4	…	4
D	2	…	3
E	3	…	1
F	5	…	3
G	5	…	5
H	3	…	4
I	4	…	3
J	1	…	2
K	3	…	1
L	4	…	4
M	3	…	5
N	4	…	5
O	2	…	4
平均	3.5	…	3.2
标准差	1.2	…	1.4

$$\sqrt{\frac{(5-3.5)^2+\cdots+(2-3.5)^2}{15-1}}=1.2$$

	Q1a 的标准值 u_1	…	Q1f 的标准值 u_6
A	1.2	…	−0.9
B	1.2	…	−0.9
C	0.4	…	0.6
D	−1.2	…	−0.1
E	−0.4	…	−1.6
F	1.2	…	−0.1
G	1.2	…	1.3
H	−0.4	…	0.6
I	0.4	…	−0.1
J	−2.0	…	−0.9
K	−0.4	…	−1.6
L	0.4	…	0.6
M	−0.4	…	1.3
N	0.4	…	1.3
O	−1.2	…	0.6
平均	0	…	0
标准差	1	…	1

$$\frac{2-3.5}{1.2}=-1.2$$

$$\sqrt{\frac{(-0.9-0)^2+\cdots+(0.6-0)^2}{15-1}}=1$$



步骤 2　将标准化之后的数据假定为如下形式。

假定方框中的部分，平均值为 0、方差为 1。

	Q1a 的标准值 u_1	…	Q1f 的标准值 u_6
A	1.2	…	−0.9
:	:	:	:
O	−1.2	…	0.6
平均	0	…	0
标准差	1	…	1

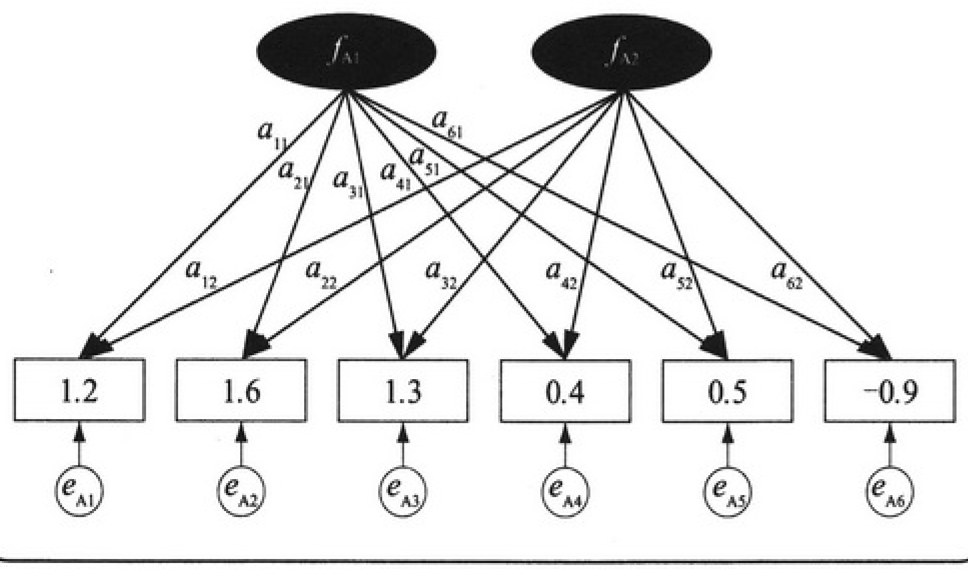
=

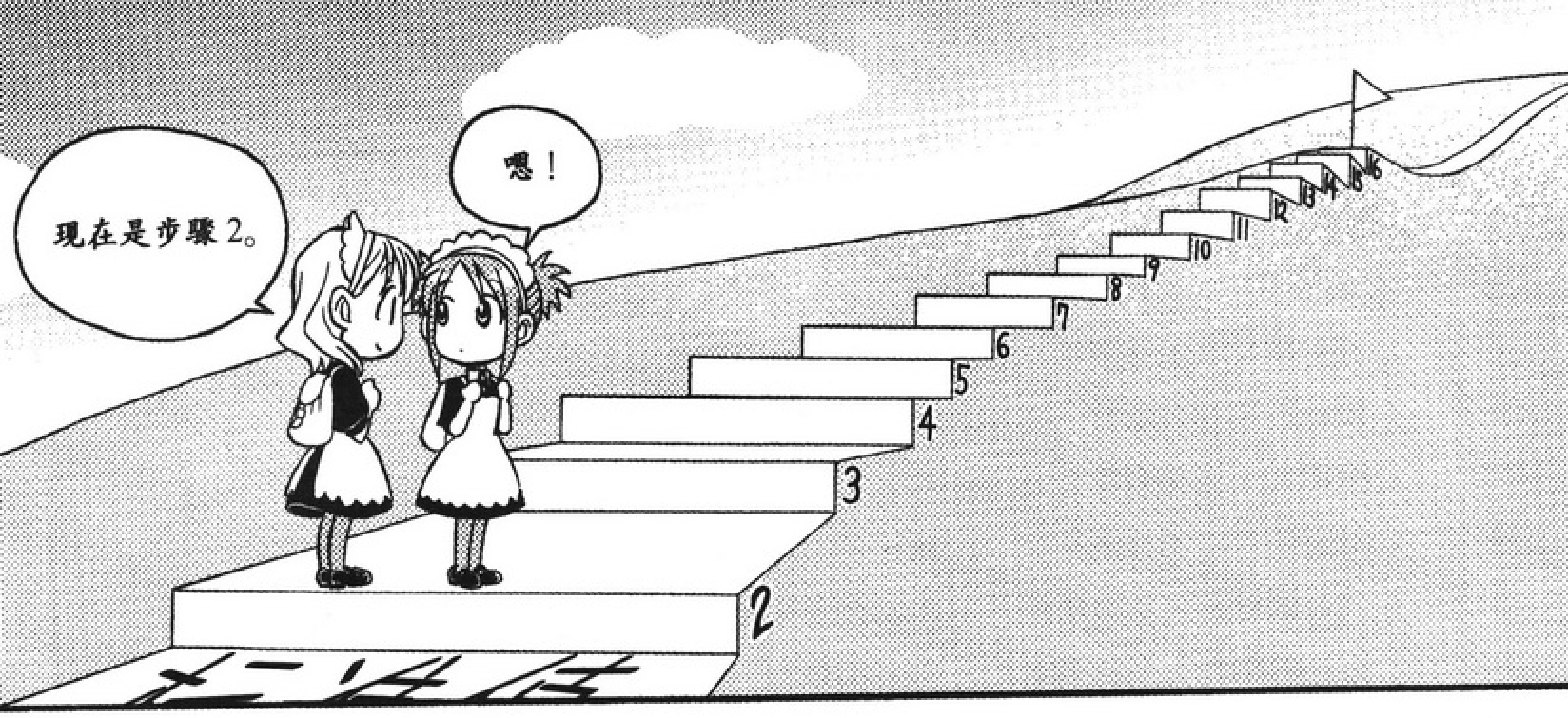
	Q1a 的标准值 u_1	…	Q1f 的标准值 u_6
A	$a_{11}\boxed{f_{A1}}+a_{12}\boxed{f_{A2}}+\boxed{e_{A1}}$	…	$a_{61}\boxed{f_{A1}}+a_{62}\boxed{f_{A2}}+\boxed{e_{A6}}$
:	:	:	:
O	$a_{11}\boxed{f_{O1}}+a_{12}\boxed{f_{O2}}+\boxed{e_{O1}}$	…	$a_{61}\boxed{f_{O1}}+a_{62}\boxed{f_{O2}}+\boxed{e_{O6}}$
平均	0	…	0
标准差	1	…	1

假定平均值为 0、方差为 d_1^2。

假定平均值为 0、方差为 d_6^2。

如果用图来表示 A 的数据，便会得到以下结果。

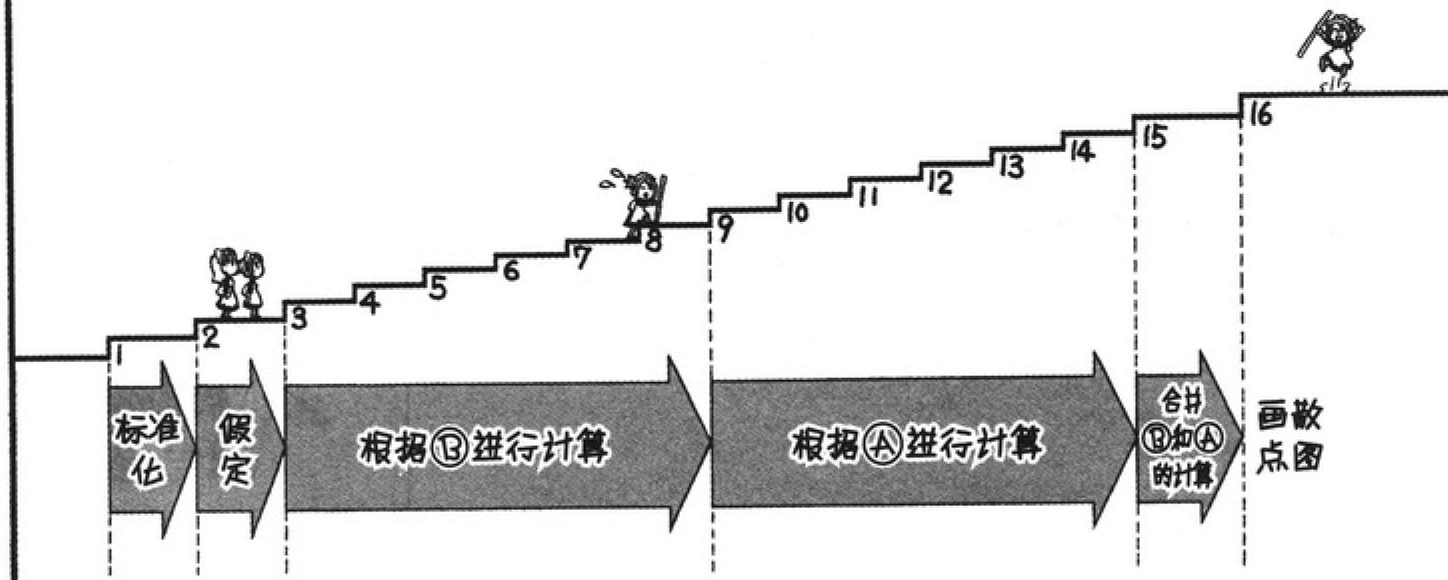


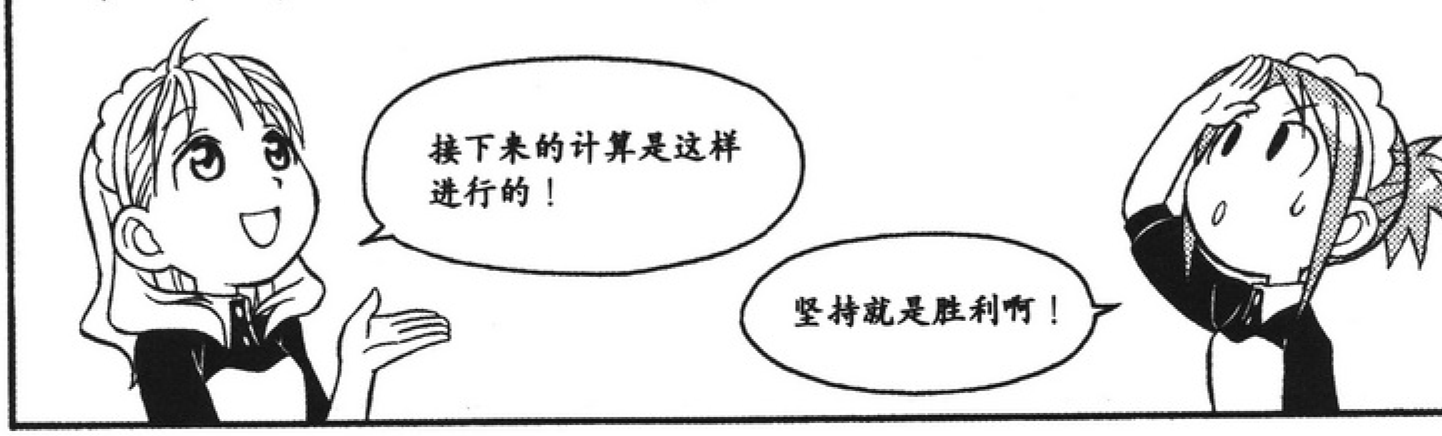


	Q1a 的标准值 u_1	…	Q1f 的标准值 u_6
A	1.2	…	−0.9
⋮	⋮	Ⓐ	⋮
O	−1.2	…	0.6
平均	0	…	0
标准差	1	…	1

=

	Q1a 的标准值 u_1	…	Q1f 的标准值 u_6
A	$a_{11}f_{A1}+a_{12}f_{A2}+e_{A1}$	…	$a_{61}f_{A1}+a_{62}f_{A2}+e_{A6}$
⋮	⋮	Ⓑ	⋮
O	$a_{11}f_{O1}+a_{12}f_{O2}+e_{O1}$	…	$a_{61}f_{O1}+a_{62}f_{O2}+e_{O6}$
平均	0	…	0
标准差	1	…	1





步骤 3

假定：

f_1 和 e_1 的单相关系数是 0
f_1 和 e_2 的单相关系数是 0
f_1 和 e_3 的单相关系数是 0
f_1 和 e_4 的单相关系数是 0
f_1 和 e_5 的单相关系数是 0
f_1 和 e_6 的单相关系数是 0
f_2 和 e_1 的单相关系数是 0
f_2 和 e_2 的单相关系数是 0
f_2 和 e_3 的单相关系数是 0
f_2 和 e_4 的单相关系数是 0
f_2 和 e_5 的单相关系数是 0
f_2 和 e_6 的单相关系数是 0

e_1 和 e_2 的单相关系数是 0
e_1 和 e_3 的单相关系数是 0
e_1 和 e_4 的单相关系数是 0
e_1 和 e_5 的单相关系数是 0
e_1 和 e_6 的单相关系数是 0
e_2 和 e_3 的单相关系数是 0
e_2 和 e_4 的单相关系数是 0
e_2 和 e_5 的单相关系数是 0
e_2 和 e_6 的单相关系数是 0
e_3 和 e_4 的单相关系数是 0
e_3 和 e_5 的单相关系数是 0
e_3 和 e_6 的单相关系数是 0
e_4 和 e_5 的单相关系数是 0
e_4 和 e_6 的单相关系数是 0
e_5 和 e_6 的单相关系数是 0

例如：假定 f_2 和 e_6 的单相关系数的值为 0，也就是说

$$\frac{f_2\text{ 和 }e_6\text{ 的离差积和}}{\sqrt{f_2\text{ 的离差平方和 }\times e_6\text{ 的离差平方和}}}=0$$

换句话说，就是

f_2 和 e_6 的离差积和

$$
\begin{aligned}
&=(f_{A2}-\overline{f}_2)(e_{A6}-\overline{e}_6)+\cdots+(f_{O2}-\overline{f}_2)(e_{O6}-\overline{e}_6)\\
&=(f_{A2}-0)(e_{A6}-0)+\cdots+(f_{O2}-0)(e_{O6}-0)\\
&=f_{A2}e_{A6}+\cdots+f_{O2}e_{O6}\\
&=0
\end{aligned}
$$

请参见步骤 2。

步骤 4 假定 f_1 和 f_2 的单相关系数的值为 0，也就意味着

$$\frac{f_1\text{ 和 }f_2\text{ 的离差积和}}{\sqrt{f_1\text{ 的离差平方和 }\times f_2\text{ 的离差平方和}}}=0$$

还要假定
公共因子之间
不相关。

换句话说，就是

$$\begin{aligned}
&f_1\text{ 和 }f_2\text{ 的离差积和}\\
&=(f_{A1}-\overline{f}_1)(f_{A2}-\overline{f}_2)+\cdots+(f_{O1}-\overline{f}_1)(f_{O2}-\overline{f}_2)\\
&=(f_{A1}-0)(f_{A2}-0)+\cdots+(f_{O1}-0)(f_{O2}-0)\\
&=f_{A1}f_{A2}+\cdots+f_{O1}f_{O2}\\
&=0
\end{aligned}$$

请参见步骤 2。

“任意两公共因子之间的单相关系数的值为零”，这种假定的思考方法被称为**正交因子模型**。不作这种假定的思考方法则被称为**斜交因子模型**。

之前，由于“计算相对简单”这一原因，人们通常先设想成正交因子模型再进行分析。但是现在，随着计算机性能的不断进步，而且“任意两公共因子的单相关系数的值为 0”，这样的假设本来就没有什么现实依据。

所以，设想成斜交因子模型的分析也渐渐开始被人们接受。这个例子中我们以正交因子模型来分析。

步骤 5　例如，可以将 $Q1b$ 的标准值 u_2 和 $Q1f$ 的标准值 u_6 的单相关系数写成以下形式，请您思考这一过程并进行确认。

$$\frac{u_2\text{ 和 }u_6\text{ 的离差积和}}{\sqrt{u_2\text{ 的离差平方和 }\times u_6\text{ 的离差平方和}}}$$

$$=\frac{\dfrac{u_2\text{ 和 }u_6\text{ 的离差积和}}{15-1}}{\dfrac{\sqrt{u_2\text{ 的离差平方和 }\times u_6\text{ 的离差平方和}}}{15-1}}$$

分子和分母同时除以“数据个数 -1”。

$$=\frac{\dfrac{u_2\text{ 和 }u_6\text{ 的离差积和}}{15-1}}{\sqrt{\dfrac{u_2\text{ 的离差平方和}}{15-1}}\times\sqrt{\dfrac{u_6\text{ 的离差平方和}}{15-1}}}$$

将分母分开写成 u_2 的标准差和 u_6 的标准差。

$$=\frac{\dfrac{u_2\text{ 和 }u_6\text{ 的离差积和}}{15-1}}{1\times1}$$

请参见步骤 1。

$$=\frac{u_2\text{ 和 }u_6\text{ 的离差积和}}{15-1}$$

$$=\frac{(a_{21}f_{A1}+a_{22}f_{A2}+e_{A2})(a_{61}f_{A1}+a_{62}f_{A2}+e_{A6})+\cdots+(a_{21}f_{O1}+a_{22}f_{O2}+e_{O2})(a_{61}f_{O1}+a_{62}f_{O2}+e_{O6})}{15-1}$$

整理分子

$$(a_{21}f_{A1}+a_{22}f_{A2}+e_{A2})(a_{61}f_{A1}+a_{62}f_{A2}+e_{A6})+\cdots+(a_{21}f_{O1}+a_{22}f_{O2}+e_{O2})(a_{61}f_{O1}+a_{62}f_{O2}+e_{O6})$$

$$\begin{array}{lllllllll}
=a_{21}f_{A1}a_{61}f_{A1} & +a_{21}f_{A1}a_{62}f_{A2} & +a_{21}f_{A1}e_{A6} & +a_{22}f_{A2}a_{61}f_{A1} & +a_{22}f_{A2}a_{62}f_{A2} & +a_{22}f_{A2}e_{A6} & +e_{A2}a_{61}f_{A1} & +e_{A2}a_{62}f_{A2} & +e_{A2}e_{A6} \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
+a_{21}f_{O1}a_{61}f_{O1} & +a_{21}f_{O1}a_{62}f_{O2} & +a_{21}f_{O1}e_{O6} & +a_{22}f_{O2}a_{61}f_{O1} & +a_{22}f_{O2}a_{62}f_{O2} & +a_{22}f_{O2}e_{O6} & +e_{O2}a_{61}f_{O1} & +e_{O2}a_{62}f_{O2} & +e_{O2}e_{O6}
\end{array}$$

$$\begin{aligned}
&=a_{21}a_{61}(f_{A1}^2+\cdots+f_{O1}^2) &&+a_{21}a_{62}(f_{A1}f_{A2}+\cdots+f_{O1}f_{O2}) &&+a_{21}(f_{A1}e_{A6}+\cdots+f_{O1}e_{O6})\\
&+a_{22}a_{61}(f_{A2}f_{A1}+\cdots+f_{O2}f_{O1}) &&+a_{22}a_{62}(f_{A2}^2+\cdots+f_{O2}^2) &&+a_{22}(f_{A}e_{A6}+\cdots+f_{O2}e_{O6})\\
&+a_{61}(f_{A1}e_{A2}+\cdots+f_{O1}e_{O2}) &&+a_{62}(f_{A2}e_{A2}+\cdots+f_{O2}e_{O2}) &&+(e_{A2}e_{A6}+\cdots+e_{O2}e_{O6})
\end{aligned}$$

$$
\begin{aligned}
= \ & a_{21}a_{61}(f_1\text{ 的离差平方和}) + a_{21}a_{62}(f_1\text{ 和 }f_2\text{ 的离差积和}) + a_{21}(f_1\text{ 和 }e_6\text{ 的离差积和})\\
& + a_{22}a_{61}(f_1\text{ 和 }f_2\text{ 的离差积和}) + a_{22}a_{62}(f_2\text{ 的离差平方和}) + a_{22}(f_2\text{ 和 }e_6\text{ 的离差积和})\\
& + a_{61}(f_1\text{ 和 }e_2\text{ 的离差积和}) + a_{62}(f_2\text{ 和 }e_2\text{ 的离差积和}) + (e_2\text{ 和 }e_6\text{ 的离差积和})
\end{aligned}
$$

$$
\begin{aligned}
= \ & a_{21}a_{61}(f_1\text{ 的离差平方和}) + 0 && && + 0\\
& + 0 && + a_{22}a_{62}(f_2\text{ 的离差平方和}) && + 0\\
& + 0 && + 0 && + 0
\end{aligned}
$$

请参见步骤 3 和步骤 4。

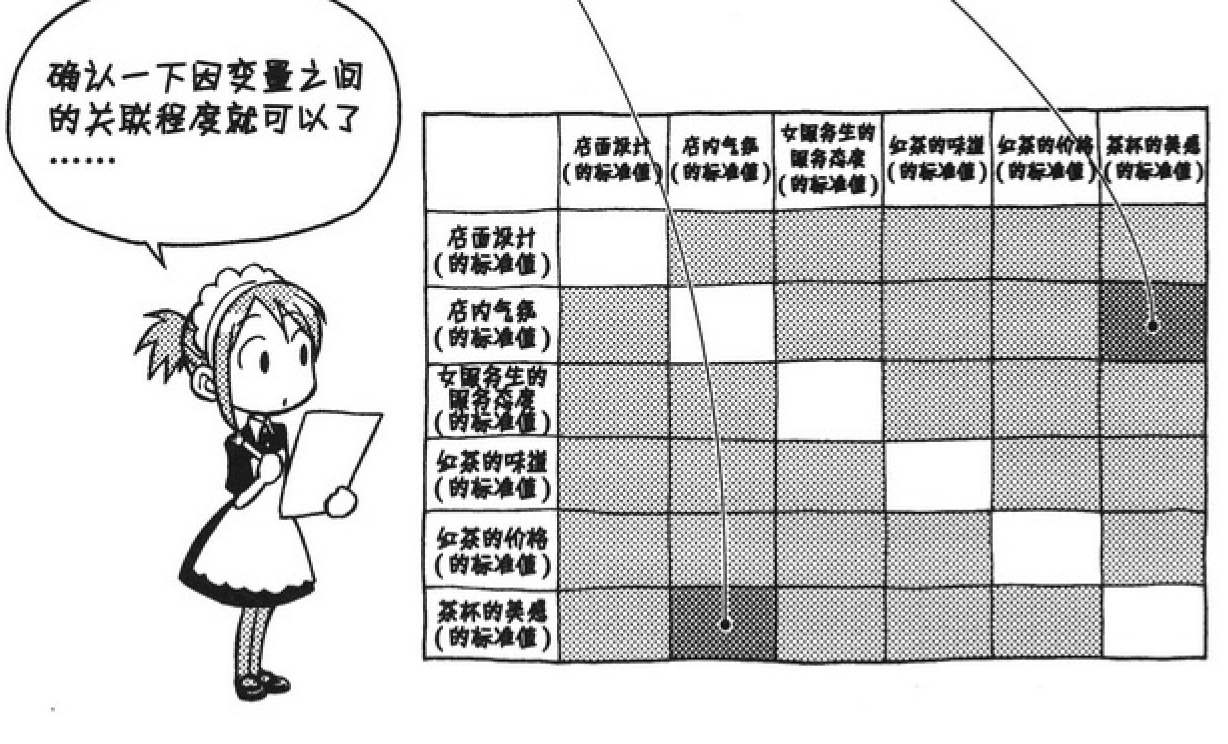
$$= \frac{a_{21}a_{61}(f_1\text{ 的离差平方和}) + a_{22}a_{62}(f_2\text{ 的离差平方和})}{15-1}$$

$$= a_{21}a_{61} \times \frac{f_1\text{ 的离差平方和}}{15-1} + a_{22}a_{62} \times \frac{f_2\text{ 的离差平方和}}{15-1}$$

$$= a_{21}a_{61} \times (f_1\text{ 的方差}) + a_{22}a_{62} \times (f_2\text{ 的方差})$$

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

请参见步骤 2。



步骤 6 例如，可以将 u_1 和 u_2 的单相关系数改写成以下形式，请您思考这一过程并进行确认。

$$\frac{u_2\text{ 和 }u_2\text{ 的离差积和}}{\sqrt{u_2\text{ 的离差平方和 }\times u_2\text{ 的离差平方和}}}$$

$$=\frac{\dfrac{u_2\text{ 和 }u_2\text{ 的离差积和}}{15-1}}{\sqrt{\dfrac{u_2\text{ 的离差平方和 }\times u_2\text{ 的离差平方和}}{15-1}}}$$

分子和分母同时除以“数据个数 -1”。

$$=\frac{\dfrac{u_2\text{ 和 }u_2\text{ 的离差积和}}{15-1}}{u_2\text{ 的方差}}$$

$$=\frac{\dfrac{u_2\text{ 和 }u_2\text{ 的离差积和}}{15-1}}{1}$$

根据方差 =（标准差）^{2}，同时请参见步骤 1。

$$=\frac{u_2\text{ 和 }u_2\text{ 的离差积和}}{15-1}$$

$$= u_2\text{ 的方差}$$

$$=\frac{(a_{21}f_{A1}+a_{22}f_{A2}+e_{A2})^2+\cdots+(a_{21}f_{O1}+a_{22}f_{A2}+e_{O2})^2}{15-1}$$

整理分子

$$(a_{21}f_{A1}+a_{22}f_{A2}+e_{A2})^2+\cdots+(a_{21}f_{O1}+a_{22}f_{O2}+e_{O2})$$

$$= (a_{21}f_{A1})^2 + (a_{22}f_{A2})^2 + (e_{A2})^2 + 2(a_{21}f_{A1})(a_{22}f_{A2}) + 2(a_{22}f_{A2})(e_{A2}) + 2(e_{A2})(a_{21}f_{A1})$$

$$+ (a_{21}f_{O1})^2 + (a_{22}f_{O2})^2 + (e_{O2})^2 + 2(a_{21}f_{O1})(a_{22}f_{O2}) + 2(a_{22}f_{O2})(e_{O2}) + 2(e_{O2})(a_{21}f_{O1})$$

$$= a_{21}^2(f_{A1}^2+\cdots+f_{O1}^2) \quad + a_{22}^2(f_{A2}^2+\cdots+f_{O2}^2) \quad + (e_{A2}^2+\cdots+e_{O2}^2)$$
$$+ 2a_{21}a_{22}(f_{A1}f_{A2}+\cdots+f_{O1}f_{O2}) \quad + 2a_{22}(f_{A2}e_{A2}+\cdots+f_{O1}e_{O2}) \quad + 2a_{21}(f_{A1}e_{A2}+\cdots+f_{O1}e_{O2})$$

$=a_{21}^2$(f_1 的离差平方和) $+a_{22}^2$(f_2 的离差平方和) +(e_2 的离差平方和)
$+2a_{21}a_{22}$(f_1 和 f_2 的离差积和)$+2a_{22}$(f_2 和 e_2 的离差积和)$+2a_{21}$(f_1 和 e_2 的离差积和)

$$= a_{21}^2(f_1\text{ 的离差平方和}) + a_{22}^2(f_2\text{ 的离差平方和}) + (e_2\text{ 的离差平方和})$$

$$+0 \qquad\qquad +0 \qquad\qquad +0$$

请参见步骤 3 和步骤 4。

$$= \frac{a_{21}^2(f_1\text{ 的离差平方和}) + a_{22}^2(f_2\text{ 的离差平方和}) + (e_2\text{ 的离差平方和})}{15-1}$$

$$= a_{21}^2 \times \frac{f_1\text{ 的离差平方和}}{15-1} + a_{22}^2 \times \frac{f_2\text{ 的离差平方和}}{15-1} + \frac{e_2\text{ 的离差平方和}}{15-1}$$

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

$$= a_{21}^2 + a_{22}^2 + d_2^2$$

请参见步骤 2。

	店面设计(的标准值)	店内气氛(的标准值)	女服务生的服务态度(的标准值)	红茶的味道(的标准值)	红茶的价格(的标准值)	茶杯的美感(的标准值)
店面设计(的标准值)						
店内气氛(的标准值)						
女服务生的服务态度(的标准值)						
红茶的味道(的标准值)						
红茶的价格(的标准值)						
茶杯的美感(的标准值)						

步骤 7 由步骤 5 和步骤 6 可知，可以将相关矩阵改写成以下形式，请您思考这一过程并进行确认。

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

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

步骤 8 整理步骤 7 中的式子。在计算的过程中出现的、等式左边主对角线上的 $1-d_i^2$ 被称为**共性方差**，有时记作“h_i^2”。

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

将步骤 7 中的式子进行移项。

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

将 r_{ii} 改写成 1。

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

整理右边

$$\begin{bmatrix} a_{11}^2+a_{12}^2 & a_{11}a_{21}+a_{12}a_{22} & \cdots & a_{11}a_{61}+a_{12}a_{62} \\ a_{21}a_{11}+a_{22}a_{12} & a_{21}^2+a_{22}^2 & \cdots & a_{21}a_{61}+a_{22}a_{62} \\ \vdots & \vdots & \ddots & \vdots \\ a_{61}a_{11}+a_{62}a_{12} & a_{61}a_{21}+a_{62}a_{22} & \cdots & a_{61}^2+a_{62}^2 \end{bmatrix}$$

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

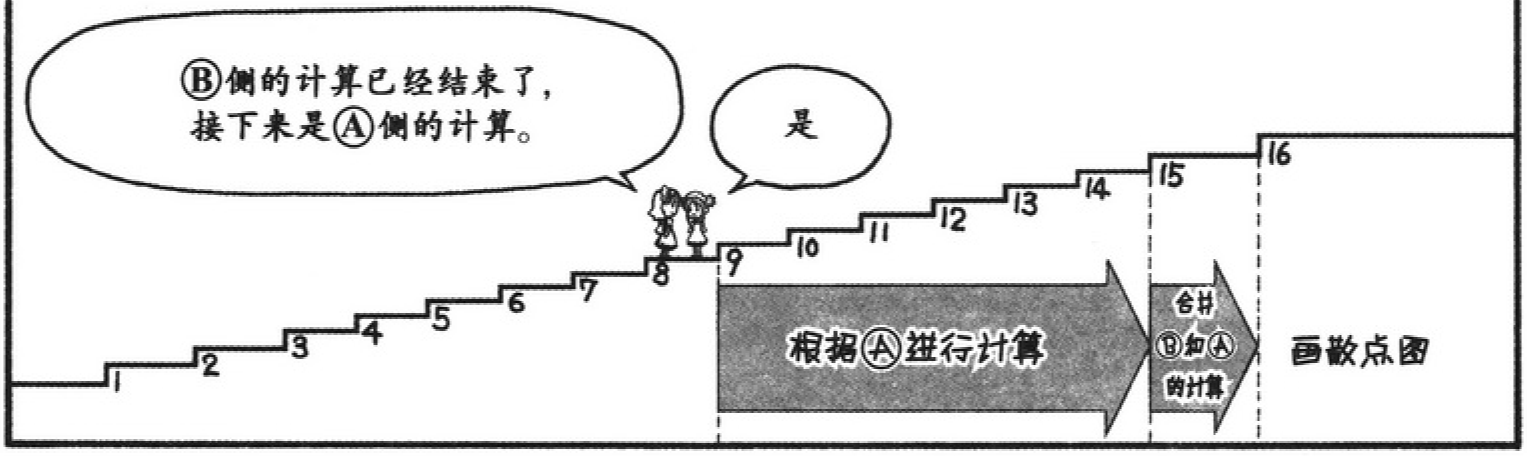
$$+ \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{61} \\ a_{12} & a_{22} & \cdots & a_{62} \end{bmatrix}$$

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

8

	Q1a 的标准值 u_1	…	Q1f 的标准值 u_6
A	1.2	…	−0.9
⋮	⋮		⋮
O	−1.2	…	0.6
平均	0	…	0
标准差	1	…	1

	Q1a 的标准值 u_1	…	Q1f 的标准值 u_6
A	$a_{11}f_{A1}+a_{12}f_{A2}+e_{A1}$	…	$a_{61}f_{A1}+a_{62}f_{A2}+e_{A6}$
⋮	⋮		⋮
O	$a_{11}f_{O1}+a_{12}f_{O2}+e_{O1}$	…	$a_{61}f_{O1}+a_{62}f_{O2}+e_{O6}$
平均	0	…	0
标准差	1	…	1



步骤 9

由分析对象的数据来求解实际的相关矩阵，再减去$\begin{bmatrix} d_1^2 & 0 & \cdots & 0 \\ 0 & d_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_6^2 \end{bmatrix}$。

$$\begin{bmatrix} 1 & 0.65 & 0.80 & 0.11 & 0.01 & 0.14 \\ 0.65 & 1 & 0.59 & 0.02 & 0.19 & 0.01 \\ 0.80 & 0.89 & 1 & 0.02 & 0.04 & 0.10 \\ 0.11 & 0.02 & 0.02 & 1 & 0.82 & 0.77 \\ 0.01 & 0.19 & 0.04 & 0.82 & 1 & 0.64 \\ 0.14 & 0.01 & 0.10 & 0.77 & 0.64 & 1 \end{bmatrix} - \begin{bmatrix} d_1^2 & 0 & 0 & 0 & 0 & 0 \\ 0 & d_2^2 & 0 & 0 & 0 & 0 \\ 0 & 0 & d_3^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & d_4^2 & 0 & 0 \\ 0 & 0 & 0 & 0 & d_5^2 & 0 \\ 0 & 0 & 0 & 0 & 0 & d_6^2 \end{bmatrix}$$

$$= \begin{bmatrix} 1-d_1^2 & 0.65 & 0.80 & 0.11 & 0.01 & 0.14 \\ 0.65 & 1-d_2^2 & 0.89 & 0.02 & 0.19 & 0.01 \\ 0.80 & 0.89 & 1-d_3^2 & 0.02 & 0.04 & 0.10 \\ 0.11 & 0.02 & 0.02 & 1-d_4^2 & 0.82 & 0.77 \\ 0.01 & 0.19 & 0.04 & 0.82 & 1-d_5^2 & 0.64 \\ 0.14 & 0.01 & 0.10 & 0.77 & 0.64 & 1-d_6^2 \end{bmatrix}$$

步骤 10 步骤 9 中主对角线上的值，也就是共性方差 $1-d_i^2$ 的值代入，具体如下：

$$
\begin{cases}
1-d_1^2 = \text{以 } u_1 \text{ 为因变量、以 } u_2\text{、}u_3\text{、}u_4\text{、}u_5\text{、}u_6 \text{ 为自变量的重回归式的贡献率 } R_1^2 \\
1-d_2^2 = \text{以 } u_2 \text{ 为因变量、以 } u_1\text{、}u_3\text{、}u_4\text{、}u_5\text{、}u_6 \text{ 为自变量的重回归式的贡献率 } R_2^2 \\
1-d_3^2 = \text{以 } u_3 \text{ 为因变量、以 } u_1\text{、}u_2\text{、}u_4\text{、}u_5\text{、}u_6 \text{ 为自变量的重回归式的贡献率 } R_3^2 \\
1-d_4^2 = \text{以 } u_4 \text{ 为因变量、以 } u_1\text{、}u_2\text{、}u_3\text{、}u_5\text{、}u_6 \text{ 为自变量的重回归式的贡献率 } R_4^2 \\
1-d_5^2 = \text{以 } u_5 \text{ 为因变量、以 } u_1\text{、}u_2\text{、}u_3\text{、}u_4\text{、}u_6 \text{ 为自变量的重回归式的贡献率 } R_5^2 \\
1-d_6^2 = \text{以 } u_6 \text{ 为因变量、以 } u_1\text{、}u_2\text{、}u_3\text{、}u_4\text{、}u_5 \text{ 为自变量的重回归式的贡献率 } R_6^2
\end{cases}
$$

$$
\begin{bmatrix}
1-d_1^2 & 0.65 & 0.80 & 0.11 & 0.01 & 0.14 \\
0.65 & 1-d_2^2 & 0.59 & 0.02 & 0.19 & 0.01 \\
0.80 & 0.89 & 1-d_3^2 & 0.02 & 0.04 & 0.10 \\
0.11 & 0.02 & 0.02 & 1-d_4^2 & 0.82 & 0.77 \\
0.01 & 0.19 & 0.04 & 0.82 & 1-d_5^2 & 0.64 \\
0.14 & 0.01 & 0.10 & 0.77 & 0.64 & 1-d_6^2
\end{bmatrix}
=
\begin{bmatrix}
0.68 & 0.65 & 0.80 & 0.11 & 0.01 & 0.14 \\
0.65 & 0.88 & 0.89 & 0.02 & 0.19 & 0.01 \\
0.80 & 0.89 & 0.91 & 0.02 & 0.04 & 0.10 \\
0.11 & 0.02 & 0.02 & 0.81 & 0.82 & 0.77 \\
0.01 & 0.19 & 0.04 & 0.82 & 0.81 & 0.64 \\
0.14 & 0.01 & 0.10 & 0.77 & 0.64 & 0.66
\end{bmatrix}
$$

从数学上讲，如果我们不假定共性方差 $1-d_i^2$ 的值，接下来的计算就无法进行。
关于如何假定共性方差 $1-d_i^2$ 的值，存在着许多方法。其中比较著名的一种，便是上面我们所提到的这种方法。

1. 请参照《漫画统计学之回归分析》第三章。

步骤 11

求满足$\begin{bmatrix} 0.68 & 0.65 & 0.80 & 0.11 & 0.01 & 0.14 \\ 0.65 & 0.88 & 0.89 & 0.02 & 0.19 & 0.01 \\ 0.80 & 0.89 & 0.91 & 0.02 & 0.04 & 0.10 \\ 0.11 & 0.02 & 0.02 & 0.81 & 0.82 & 0.77 \\ 0.01 & 0.19 & 0.04 & 0.82 & 0.81 & 0.64 \\ 0.14 & 0.01 & 0.10 & 0.77 & 0.64 & 0.66 \end{bmatrix}\begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{bmatrix} = \lambda\begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{bmatrix}$的特征值 λ 和特征向量$\begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{bmatrix}$，

再对特征向量进行单位化处理，即令其满足$t_1^2 + t_2^2 + t_3^2 + t_4^2 + t_5^2 + t_6^2 = 1$。

通过数据分析软件，就可得到以下结果。

特征值 λ	特征向量 $\begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{bmatrix}$
2.55	$\begin{bmatrix} 0.43 \\ 0.48 \\ 0.50 \\ 0.34 \\ 0.34 \\ 0.32 \end{bmatrix}$
2.11	$\begin{bmatrix} -0.28 \\ -0.34 \\ -0.38 \\ 0.51 \\ 0.47 \\ 0.43 \end{bmatrix}$

在这个例子中，实际上应该求出 6 组特征值和特征向量。但是，第 3 大及其以后的特征值和特征向量，在之后的内容中不会出现，我们就不做特别介绍了。

步骤 12 将第 3 大及其以后的特征值看作 0，以下的关系依然成立。请您思考这一过程并进行确认。

$$
\begin{bmatrix}
0.68 & 0.65 & 0.80 & 0.11 & 0.01 & 0.14 \\
0.65 & 0.88 & 0.89 & 0.02 & 0.19 & 0.01 \\
0.80 & 0.89 & 0.91 & 0.02 & 0.04 & 0.10 \\
0.11 & 0.02 & 0.02 & 0.81 & 0.82 & 0.77 \\
0.01 & 0.19 & 0.04 & 0.82 & 0.81 & 0.64 \\
0.14 & 0.01 & 0.10 & 0.77 & 0.64 & 0.66
\end{bmatrix}
$$

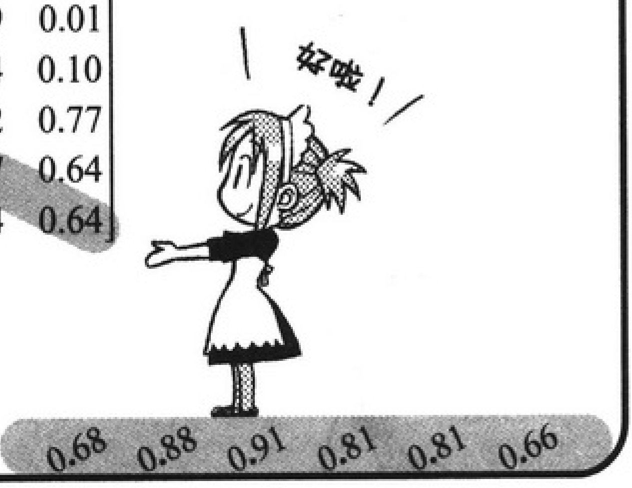
$$
\approx
\begin{bmatrix}
\sqrt{2.55}\times 0.43 & \sqrt{2.11}\times(-0.28) \\
\sqrt{2.55}\times 0.48 & \sqrt{2.11}\times(-0.34) \\
\sqrt{2.55}\times 0.50 & \sqrt{2.11}\times(-0.38) \\
\sqrt{2.55}\times 0.34 & \sqrt{2.11}\times 0.51 \\
\sqrt{2.55}\times 0.34 & \sqrt{2.11}\times 0.47 \\
\sqrt{2.55}\times 0.32 & \sqrt{2.11}\times 0.43
\end{bmatrix}
\begin{bmatrix}
\sqrt{2.55}\times 0.43 & \sqrt{2.55}\times 0.48 & \sqrt{2.55}\times 0.50 & \sqrt{2.55}\times 0.34 & \sqrt{2.55}\times 0.34 & \sqrt{2.55}\times 0.32 \\
\sqrt{2.11}\times(-0.28) & \sqrt{2.11}\times(-0.34) & \sqrt{2.11}\times(-0.38) & \sqrt{2.11}\times 0.51 & \sqrt{2.11}\times 0.47 & \sqrt{2.11}\times 0.43
\end{bmatrix}
$$

请参照第 76~78 页。

$$
\begin{bmatrix}
0.64 & 0.72 & 0.77 & 0.07 & 0.10 & 0.10 \\
0.72 & 0.83 & 0.88 & 0.05 & 0.08 & 0.08 \\
0.77 & 0.88 & 0.94 & 0.03 & 0.06 & 0.06 \\
0.07 & 0.05 & 0.03 & 0.85 & 0.81 & 0.74 \\
0.10 & 0.08 & 0.06 & 0.81 & 0.77 & 0.70 \\
0.10 & 0.08 & 0.06 & 0.74 & 0.70 & 0.64
\end{bmatrix}
$$

步骤 13 将步骤 11 的矩阵中主对角线上的值，替换为步骤 12 的矩阵中主对角线上的值。

$$
\begin{bmatrix}
0.64 & 0.65 & 0.80 & 0.11 & 0.01 & 0.14 \\
0.65 & 0.83 & 0.89 & 0.02 & 0.19 & 0.01 \\
0.80 & 0.89 & 0.94 & 0.02 & 0.04 & 0.10 \\
0.11 & 0.02 & 0.02 & 0.85 & 0.82 & 0.77 \\
0.01 & 0.19 & 0.04 & 0.82 & 0.77 & 0.64 \\
0.14 & 0.01 & 0.10 & 0.77 & 0.64 & 0.64
\end{bmatrix}
$$



步骤 14

求满足 $\begin{bmatrix} 0.64 & 0.65 & 0.80 & 0.11 & 0.01 & 0.14 \\ 0.65 & 0.83 & 0.89 & 0.02 & 0.19 & 0.01 \\ 0.80 & 0.89 & 0.94 & 0.02 & 0.04 & 0.10 \\ 0.11 & 0.02 & 0.02 & 0.85 & 0.82 & 0.77 \\ 0.01 & 0.19 & 0.04 & 0.82 & 0.77 & 0.64 \\ 0.14 & 0.01 & 0.10 & 0.77 & 0.64 & 0.64 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{bmatrix} = \lambda \begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{bmatrix}$ 的特征值 λ 和特征向量 $\begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{bmatrix}$，

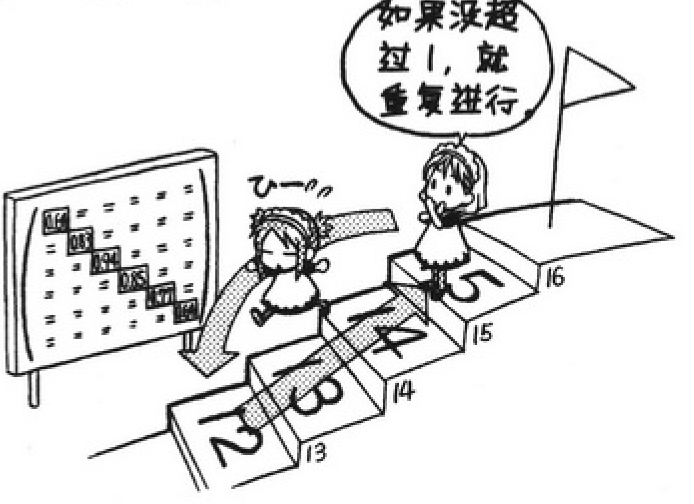
再对特征向量进行单位化处理，即令其满足 $t_1^2 + t_2^2 + t_3^2 + t_4^2 + t_5^2 + t_6^2 = 1$。

通过数据分析软件，就可得到以下结果。

特征值 λ	特征向量 $\begin{bmatrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \end{bmatrix}$
2.54	$\begin{bmatrix} 0.42 \\ 0.47 \\ 0.50 \\ 0.36 \\ 0.35 \\ 0.32 \end{bmatrix}$
2.11	$\begin{bmatrix} -0.28 \\ -0.34 \\ -0.40 \\ 0.52 \\ 0.46 \\ 0.42 \end{bmatrix}$

步骤 15 如果步骤 12 中矩阵的主对角线上没有任何一个值，也就是没有任何一个共性方差 $1-d_i^2$ 的值超过 1，那么循环操作步骤 12 到步骤 14，直到步骤 12 中有任何一个共性方差 $1-d_i^2$ 的值超过 1 为止。

将最后一个循环中的 $\begin{bmatrix} \sqrt{\lambda_1}\times t_{11} & \sqrt{\lambda_2}\times t_{12} \\ \sqrt{\lambda_1}\times t_{21} & \sqrt{\lambda_2}\times t_{22} \\ \sqrt{\lambda_1}\times t_{31} & \sqrt{\lambda_2}\times t_{32} \\ \sqrt{\lambda_1}\times t_{41} & \sqrt{\lambda_2}\times t_{42} \\ \sqrt{\lambda_1}\times t_{51} & \sqrt{\lambda_2}\times t_{52} \\ \sqrt{\lambda_1}\times t_{61} & \sqrt{\lambda_2}\times t_{62} \end{bmatrix}$ 看作 $\begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \\ a_{13} & a_{32} \\ a_{14} & a_{42} \\ a_{15} & a_{52} \\ a_{16} & a_{62} \end{bmatrix}$。



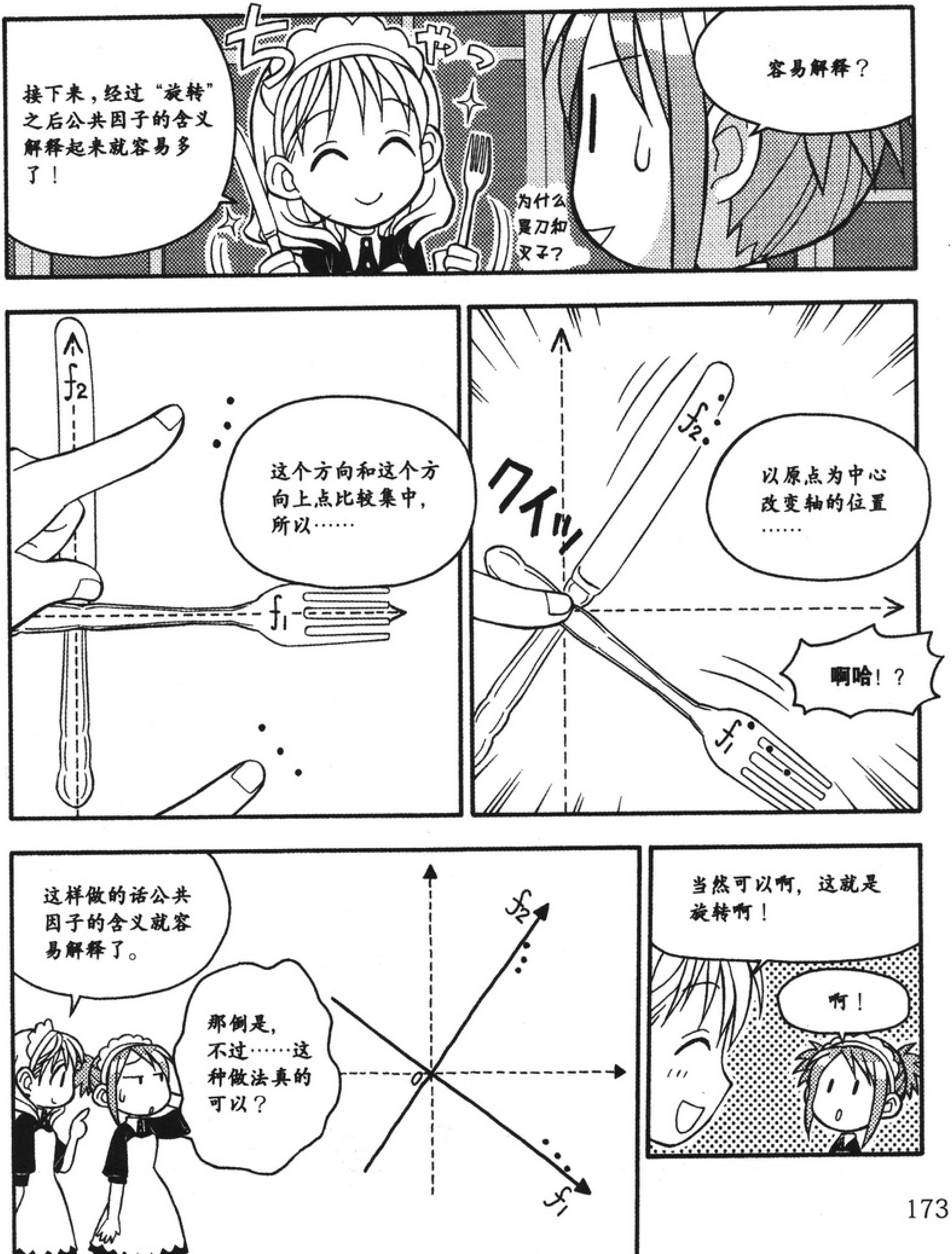
在这个例子中，由 169 页结果可知，还需在进行一轮步骤 12 到步骤 14，在接下来的一轮中的步骤 12 中的共性方差 $1-d_3^2$ 的值超过 1。所以将上一页的

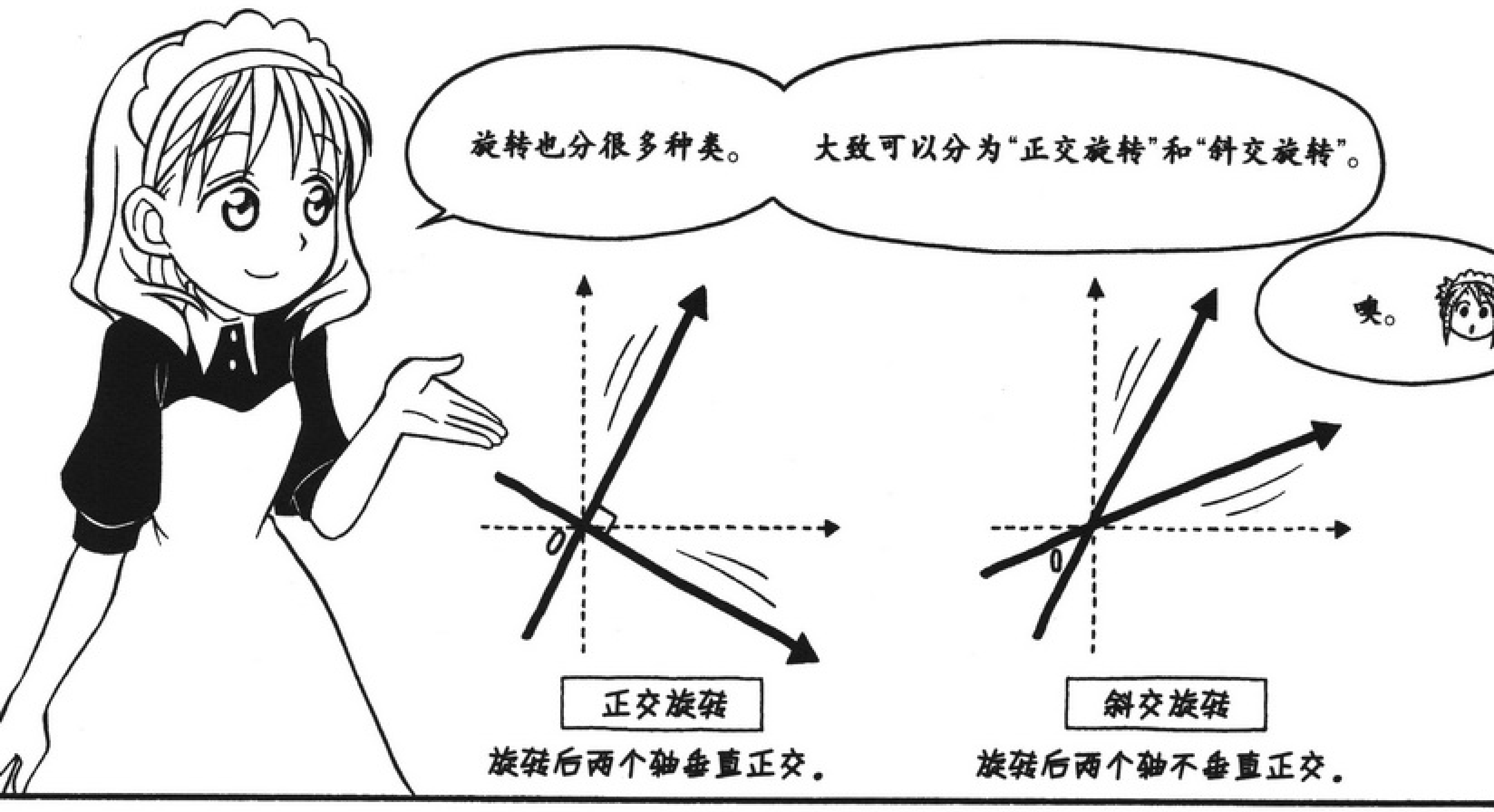
$\begin{bmatrix} \sqrt{2.55}\times 0.42 & \sqrt{2.11}\times(-0.28) \\ \sqrt{2.55}\times 0.47 & \sqrt{2.11}\times(-0.34) \\ \sqrt{2.55}\times 0.50 & \sqrt{2.11}\times(-0.40) \\ \sqrt{2.55}\times 0.36 & \sqrt{2.11}\times\quad 0.52 \\ \sqrt{2.55}\times 0.35 & \sqrt{2.11}\times\quad 0.46 \\ \sqrt{2.55}\times 0.32 & \sqrt{2.11}\times\quad 0.42 \end{bmatrix}$ 看作 $\begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \\ a_{13} & a_{32} \\ a_{14} & a_{42} \\ a_{15} & a_{52} \\ a_{16} & a_{62} \end{bmatrix}$，也就是说可以得出如下结论：

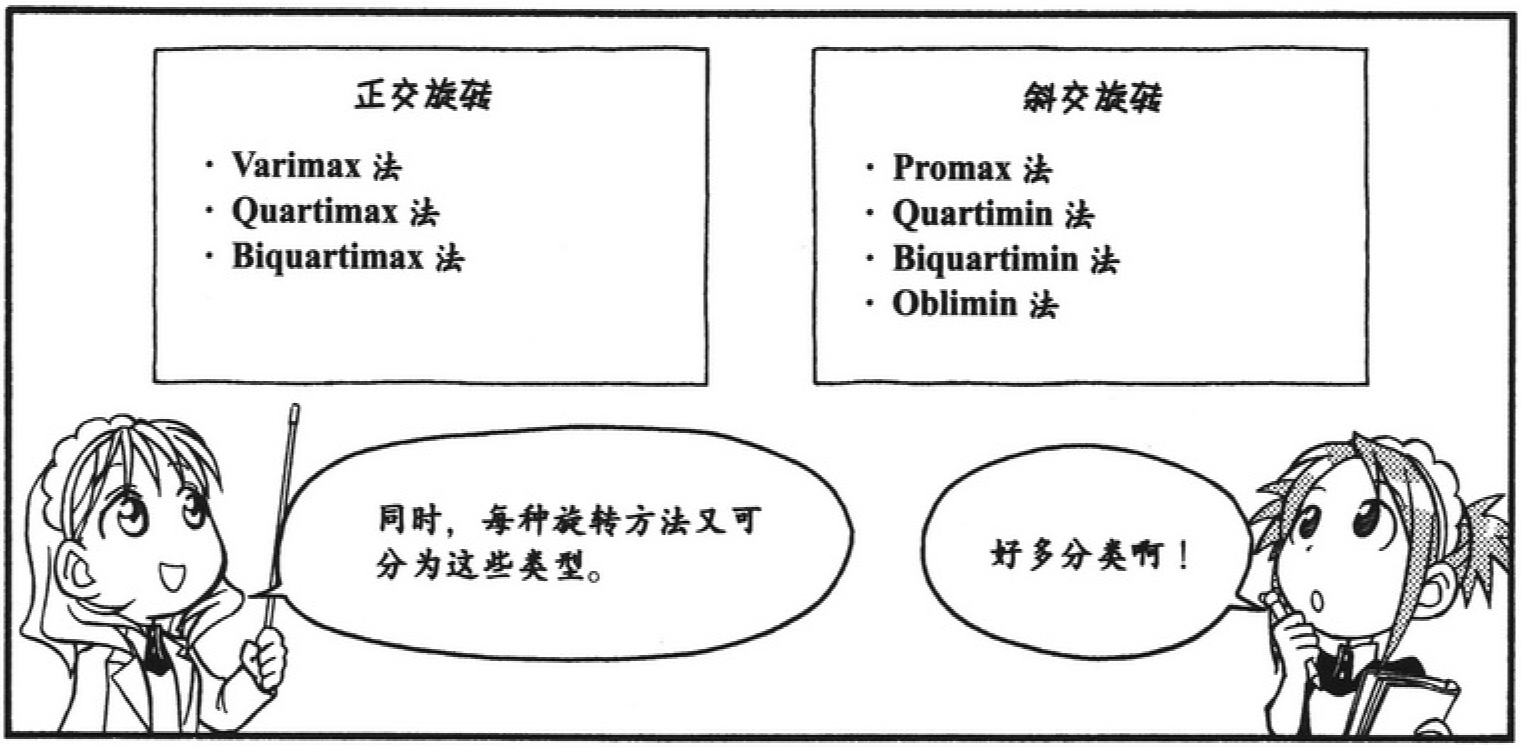
$$\begin{bmatrix} a_{11} & a_{12} \\ a_{12} & a_{22} \\ a_{13} & a_{32} \\ a_{14} & a_{42} \\ a_{15} & a_{52} \\ a_{16} & a_{62} \end{bmatrix} = \begin{bmatrix} \sqrt{\lambda_1}\times t_{11} & \sqrt{\lambda_2}\times t_{12} \\ \sqrt{\lambda_1}\times t_{21} & \sqrt{\lambda_2}\times t_{22} \\ \sqrt{\lambda_1}\times t_{31} & \sqrt{\lambda_2}\times t_{32} \\ \sqrt{\lambda_1}\times t_{41} & \sqrt{\lambda_2}\times t_{42} \\ \sqrt{\lambda_1}\times t_{51} & \sqrt{\lambda_2}\times t_{52} \\ \sqrt{\lambda_1}\times t_{61} & \sqrt{\lambda_2}\times t_{62} \end{bmatrix} = \begin{bmatrix} \sqrt{2.54}\times 0.42 & \sqrt{2.11}\times(-0.28) \\ \sqrt{2.54}\times 0.47 & \sqrt{2.11}\times(-0.34) \\ \sqrt{2.54}\times 0.50 & \sqrt{2.11}\times(-0.40) \\ \sqrt{2.54}\times 0.36 & \sqrt{2.11}\times\quad 0.52 \\ \sqrt{2.54}\times 0.35 & \sqrt{2.11}\times\quad 0.46 \\ \sqrt{2.54}\times 0.32 & \sqrt{2.11}\times\quad 0.42 \end{bmatrix} = \begin{bmatrix} 0.67 & -0.41 \\ 0.74 & -0.49 \\ 0.80 & -0.57 \\ 0.57 & 0.75 \\ 0.55 & 0.66 \\ 0.51 & 0.60 \end{bmatrix}$$

前 15 步都结束啦！
a11 a12 / a21 a22 / a31 a32 / a41 a42 / a51 a52 / a61 a62 = 0.67 −0.41 / 0.74 −0.49 / 0.80 −0.57 / 0.57 0.75 / 0.55 0.66 / 0.51 0.60
辛苦你了！
a11 a12 / a21 a22 / ⋮ ⋮ / a61 a62 被称为
“因子载荷量矩阵”或
“因子模式矩阵”。
噢！
接下来是
步骤 16 ！
将步骤 15 前求出的
因子载荷量矩阵看作
坐标，画出散点图。
坐标
Q1a. 店面设计 (0.67, −0.41)
Q1b. 店内气氛 (0.74, −0.49)
Q1c. 女服务生的服务态度 (0.80, −0.57)
Q1d. 红茶的味道 (0.57, 0.75)
Q1e. 红茶的价格 (0.55, 0.66)
Q1f. 茶杯的美感 (0.51, 0.60)
是！
这个就是旋转前的
因子载荷量了！
成功啦！
16
15
f2
1.0 0.8 0.6 0.4 0.2 −0.2 −0.4 −0.6
−0.6 −0.4 −0.2 0 0.2 0.4 0.6 0.8 1.0
f1
Q2d. 红茶的味道
Q2e. 红茶的价格
Q2f. 茶杯的美感
Q2a. 店面设计
Q2b. 店内气氛
Q2c. 女服务生的服务态度

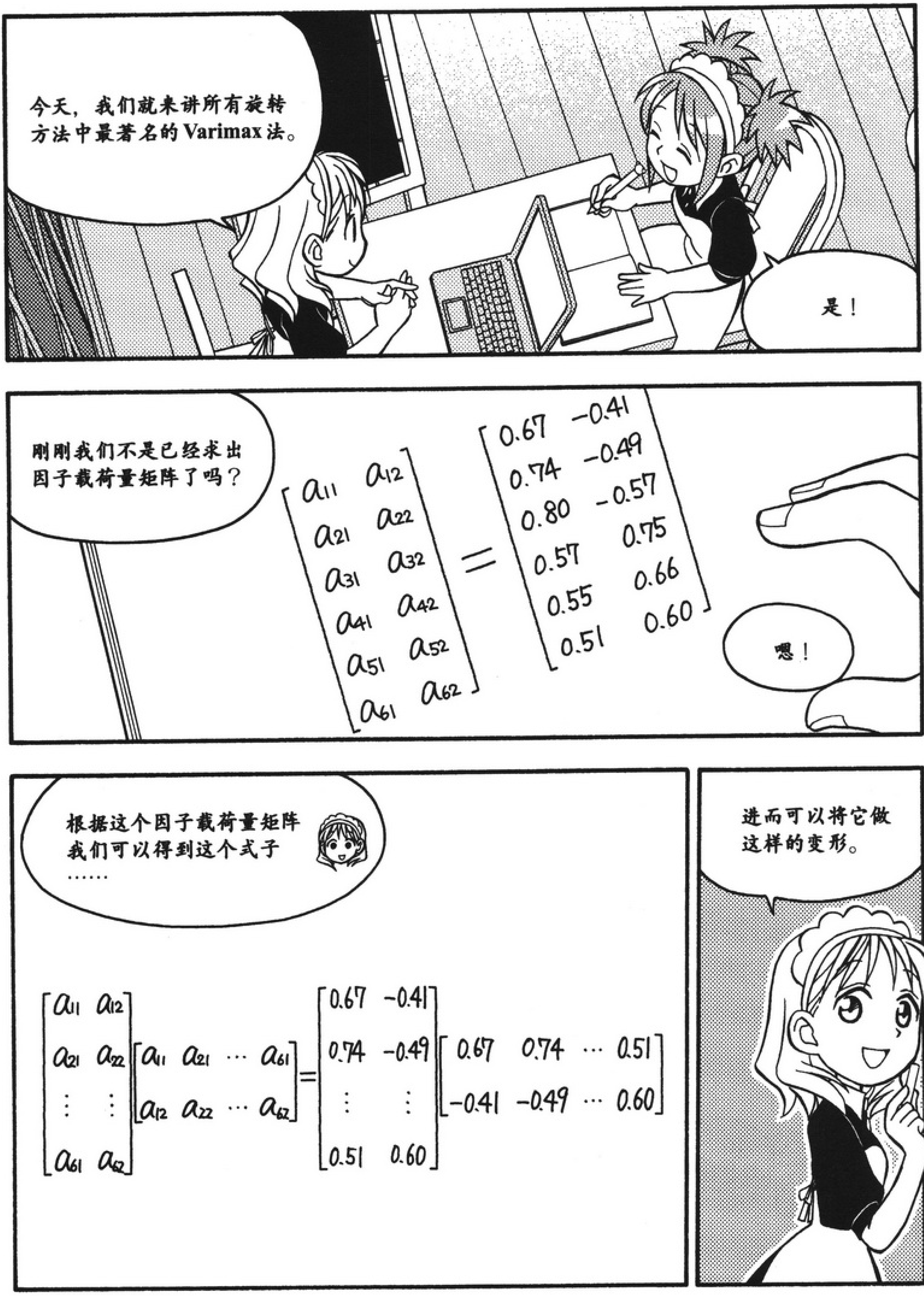
② 求解旋转后的因子载荷量




旋转也分很多种类。
大致可以分为“正交旋转”和“斜交旋转”。
噢。
正交旋转
旋转后两个轴垂直正交。
斜交旋转
旋转后两个轴不垂直正交。


正交旋转
· Varimax 法
· Quartimax 法
· Biquartimax 法
斜交旋转
· Promax 法
· Quartimin 法
· Biquartimin 法
· Oblimin 法
同时，每种旋转方法又可分为这些类型。
好多分类啊！


正交旋转中，最著名的是 Varimax 法，也称最大方差正交旋转法。
斜交旋转中，最著名的则是 Promax 法。
正交旋转
Varimax
プスプス
好酷的名字啊！


今天，我们就来讲所有旋转方法中最著名的Varimax法。
是！
刚刚我们不是已经求出因子载荷量矩阵了吗？
$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \\ a_{41} & a_{42} \\ a_{51} & a_{52} \\ a_{61} & a_{62} \end{bmatrix} = \begin{bmatrix} 0.67 & -0.41 \\ 0.74 & -0.49 \\ 0.80 & -0.57 \\ 0.57 & 0.75 \\ 0.55 & 0.66 \\ 0.51 & 0.60 \end{bmatrix}$
嗯！
根据这个因子载荷量矩阵我们可以得到这个式子……
$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{bmatrix} \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{61} \\ a_{12} & a_{22} & \cdots & a_{62} \end{bmatrix} = \begin{bmatrix} 0.67 & -0.41 \\ 0.74 & -0.49 \\ \vdots & \vdots \\ 0.51 & 0.60 \end{bmatrix} \begin{bmatrix} 0.67 & 0.74 & \cdots & 0.51 \\ -0.41 & -0.49 & \cdots & 0.60 \end{bmatrix}$
进而可以将它做这样的变形。

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

$$= \begin{bmatrix} 0.67 & -0.41 \\ 0.74 & -0.49 \\ \vdots & \vdots \\ 0.51 & 0.60 \end{bmatrix} \begin{bmatrix} 0.67 & 0.74 & \cdots & 0.51 \\ -0.41 & -0.49 & \cdots & 0.60 \end{bmatrix}$$

$$= \begin{bmatrix} 0.67 & -0.41 \\ 0.74 & -0.49 \\ \vdots & \vdots \\ 0.51 & 0.60 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.67 & 0.74 & \cdots & 0.51 \\ -0.41 & -0.49 & \cdots & 0.60 \end{bmatrix}$$

请参见 69 页。

$$= \begin{bmatrix} 0.67 & -0.41 \\ 0.74 & -0.49 \\ \vdots & \vdots \\ 0.51 & 0.60 \end{bmatrix} \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 0.67 & 0.74 & \cdots & 0.51 \\ -0.41 & -0.49 & \cdots & 0.60 \end{bmatrix}$$

请参见 72 页。

$$= \left\{ \begin{bmatrix} 0.67 & -0.41 \\ 0.74 & -0.49 \\ \vdots & \vdots \\ 0.51 & 0.60 \end{bmatrix} \begin{bmatrix} \cos(-\theta) & -\sin(-\theta) \\ \sin(-\theta) & \cos(-\theta) \end{bmatrix} \right\} \left\{ \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 0.67 & 0.74 & \cdots & 0.51 \\ -0.41 & -0.49 & \cdots & 0.60 \end{bmatrix} \right\}$$

用括号括起来。

$$= \left\{ \begin{bmatrix} 0.67 & -0.41 \\ 0.74 & -0.49 \\ \vdots & \vdots \\ 0.51 & 0.60 \end{bmatrix} \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \right\} \left\{ \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 0.67 & 0.74 & \cdots & 0.51 \\ -0.41 & -0.49 & \cdots & 0.60 \end{bmatrix} \right\}$$

$\begin{bmatrix} \cos(-\theta) = \cos\theta \\ \sin(-\theta) = -\sin\theta \end{bmatrix}$，解释从略。

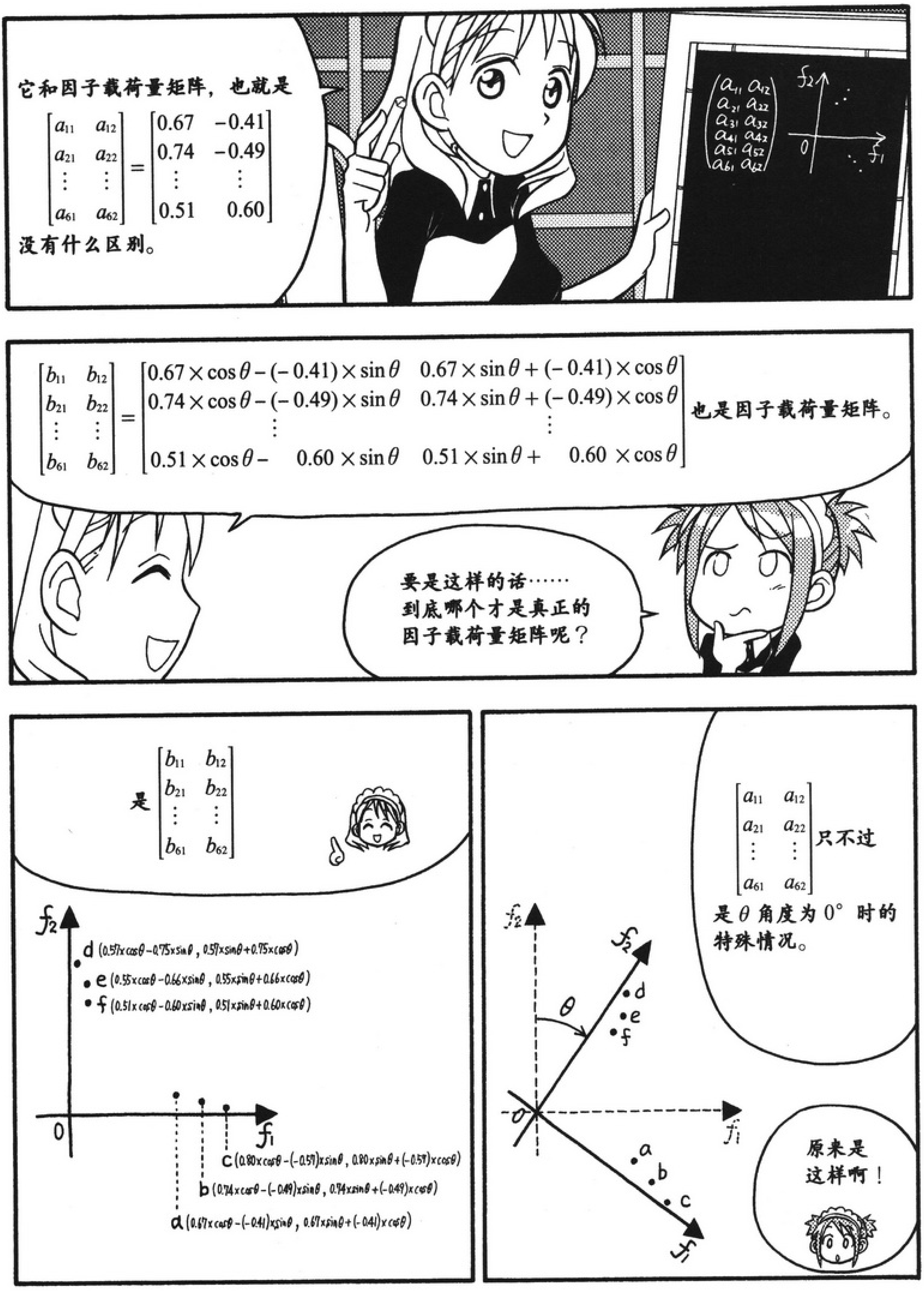
$$= \begin{bmatrix} 0.67\times\cos\theta-(-0.41)\times\sin\theta & 0.67\times\sin\theta+(-0.41)\times\cos\theta \\ 0.74\times\cos\theta-(-0.49)\times\sin\theta & 0.74\times\sin\theta+(-0.49)\times\cos\theta \\ \vdots & \vdots \\ 0.51\times\cos\theta-\ \ 0.60\times\sin\theta & 0.51\times\sin\theta+\ \ 0.60\times\cos\theta \end{bmatrix} \begin{bmatrix} 0.67\times\cos\theta-(-0.41)\times\sin\theta & 0.74\times\cos\theta-(-0.49)\times\sin\theta & \cdots & 0.51\times\cos\theta-0.60\times\sin\theta \\ 0.67\times\sin\theta+(-0.41)\times\cos\theta & 0.74\times\sin\theta+(-0.49)\times\cos\theta & \cdots & 0.51\times\sin\theta+0.60\times\cos\theta \end{bmatrix}$$

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

为了便于观察，我们将

$$\begin{bmatrix} 0.67\times\cos\theta-(-0.41)\times\sin\theta & 0.67\times\sin\theta+(-0.41)\times\cos\theta \\ 0.74\times\cos\theta-(-0.49)\times\sin\theta & 0.74\times\sin\theta+(-0.49)\times\cos\theta \\ \vdots & \vdots \\ 0.51\times\cos\theta-\ \ 0.60\times\sin\theta & 0.51\times\sin\theta+\ \ 0.60\times\cos\theta \end{bmatrix}$$ 记为

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


它和因子载荷量矩阵，也就是
$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{bmatrix} = \begin{bmatrix} 0.67 & -0.41 \\ 0.74 & -0.49 \\ \vdots & \vdots \\ 0.51 & 0.60 \end{bmatrix}$
没有什么区别。
$\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{61} & b_{62} \end{bmatrix} = \begin{bmatrix} 0.67\times\cos\theta-(-0.41)\times\sin\theta & 0.67\times\sin\theta+(-0.41)\times\cos\theta \\ 0.74\times\cos\theta-(-0.49)\times\sin\theta & 0.74\times\sin\theta+(-0.49)\times\cos\theta \\ \vdots & \vdots \\ 0.51\times\cos\theta-\ 0.60\times\sin\theta & 0.51\times\sin\theta+\ 0.60\times\cos\theta \end{bmatrix}$ 也是因子载荷量矩阵。
要是这样的话……
到底哪个才是真正的
因子载荷量矩阵呢？
是 $\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{61} & b_{62} \end{bmatrix}$
f_2
d (0.57×cosθ−0.75×sinθ, 0.57×sinθ+0.75×cosθ)
e (0.55×cosθ−0.66×sinθ, 0.55×sinθ+0.66×cosθ)
f (0.51×cosθ−0.60×sinθ, 0.51×sinθ+0.60×cosθ)
0
f_1
c (0.80×cosθ−(−0.57)×sinθ, 0.80×sinθ+(−0.57)×cosθ)
b (0.74×cosθ−(−0.49)×sinθ, 0.74×sinθ+(−0.49)×cosθ)
a (0.67×cosθ−(−0.41)×sinθ, 0.67×sinθ+(−0.41)×cosθ)
$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{bmatrix}$ 只不过
是 θ 角度为 0° 时的
特殊情况。
原来是
这样啊！

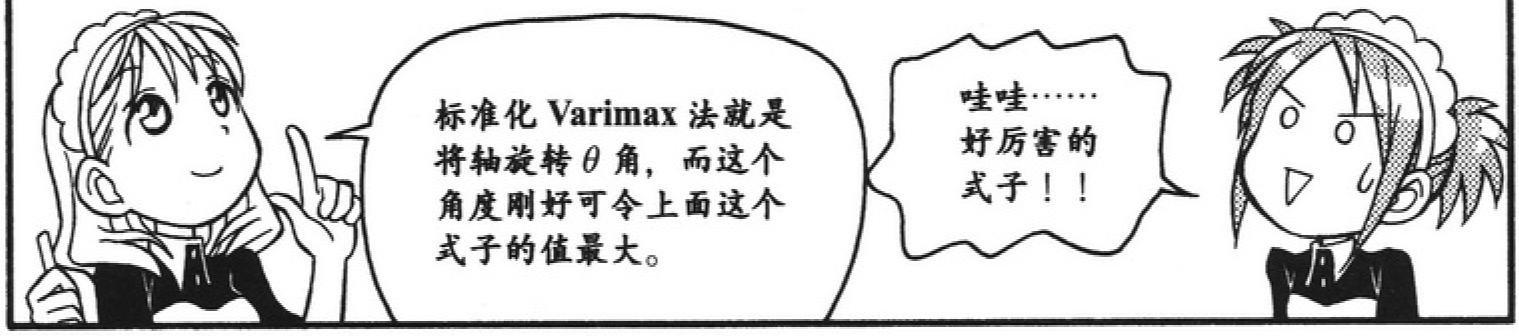

所谓 Varimax 法，拿这个例子来说，就是“第 1 公共因子的因子载荷量平方”的离差平方和同“第 2 公共因子的因子载荷量平方”的离差平方和相加。
嗯……由x的离差平方和$(x_i-\bar{x})^2$之和可知……
具体讲，这个方法就是将轴旋转θ度，而这个角度刚好可令此式子的值最大。
哦……原来是这样！
$$\left(b_{11}^2-\frac{b_{11}^2+b_{21}^2+\cdots+b_{61}^2}{6}\right)^2+\cdots+\left(b_{61}^2-\frac{b_{11}^2+b_{21}^2+\cdots+b_{61}^2}{6}\right)^2$$
$$+\left(b_{12}^2-\frac{b_{12}^2+b_{22}^2+\cdots+b_{62}^2}{6}\right)^2+\cdots+\left(b_{62}^2-\frac{b_{12}^2+b_{22}^2+\cdots+b_{62}^2}{6}\right)^2$$
……话虽如此，实际上 Varimax 法可分为“原始 Varimax 法”和“标准化 Varimax 法”两种……
诶？
raw
标准化
刚才讲的是原始 Varimax 法。
目前，通常所说的 Varimax 法指的是标准化 Varimax 法。
那要怎么做呢？
カッ
カッ

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

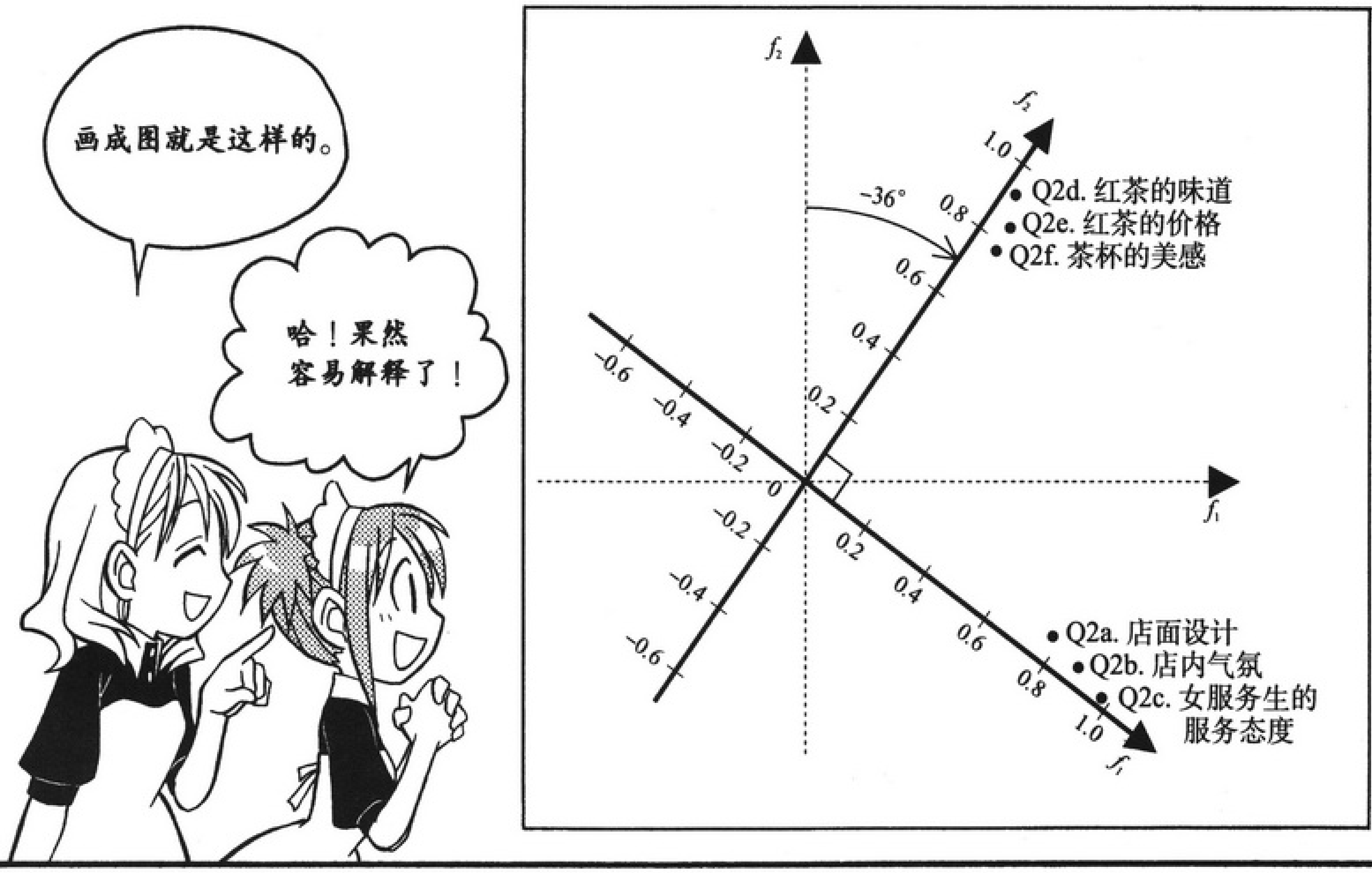
$$+\left(\frac{b_{12}^2}{b_{11}^2+b_{12}^2}-\frac{\frac{b_{12}^2}{b_{11}^2+b_{12}^2}+\cdots+\frac{b_{62}^2}{b_{61}^2+b_{62}^2}}{6}\right)^2+\cdots+\left(\frac{b_{62}^2}{b_{61}^2+b_{62}^2}-\frac{\frac{b_{12}^2}{b_{11}^2+b_{12}^2}+\cdots+\frac{b_{62}^2}{b_{61}^2+b_{62}^2}}{6}\right)^2$$

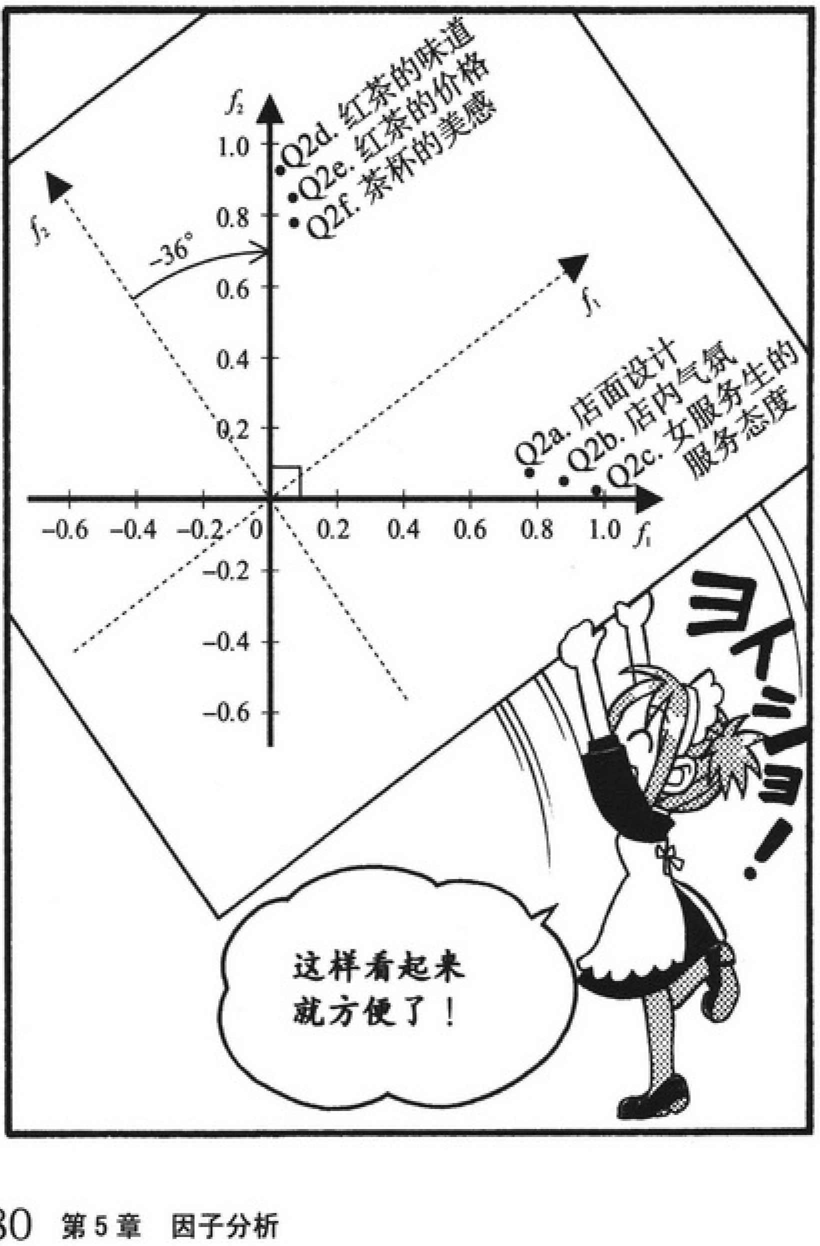
共性方差 $1-d_1^2$ $(=h_1^2=a_{11}^2+a_{12}^2)$

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

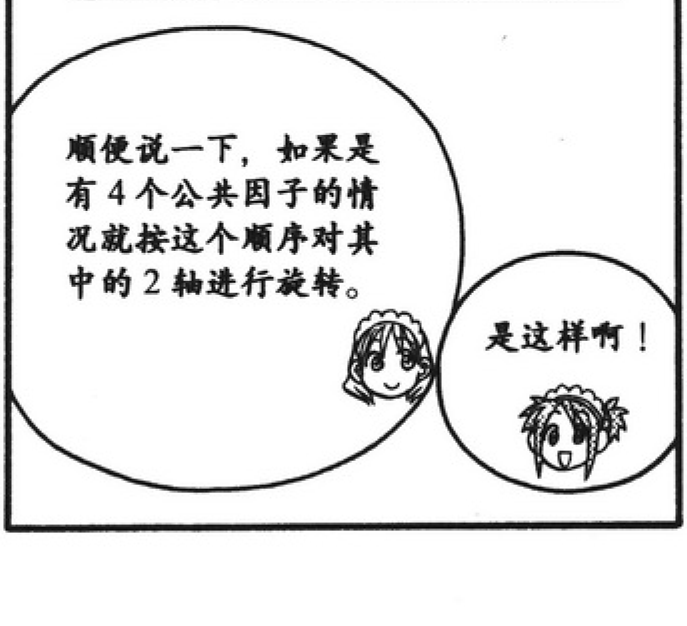


$$\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{61} & b_{62} \end{bmatrix}=\begin{bmatrix} 0.67\times\cos(-36^\circ)-(-0.41)\times\sin(-36^\circ) & 0.67\times\sin(-36^\circ)+(-0.41)\times\cos(-36^\circ) \\ 0.74\times\cos(-36^\circ)-(-0.49)\times\sin(-36^\circ) & 0.74\times\sin(-36^\circ)+(-0.49)\times\cos(-36^\circ) \\ \vdots & \vdots \\ 0.51\times\cos(-36^\circ)-\quad 0.60\times\sin(-36^\circ) & 0.51\times\sin(-36^\circ)+\quad 0.60\times\cos(-36^\circ) \end{bmatrix}=\begin{bmatrix} 0.78 & 0.07 \\ 0.89 & 0.04 \\ 0.99 & 0.01 \\ 0.01 & 0.94 \\ 0.05 & 0.86 \\ 0.06 & 0.79 \end{bmatrix}$$

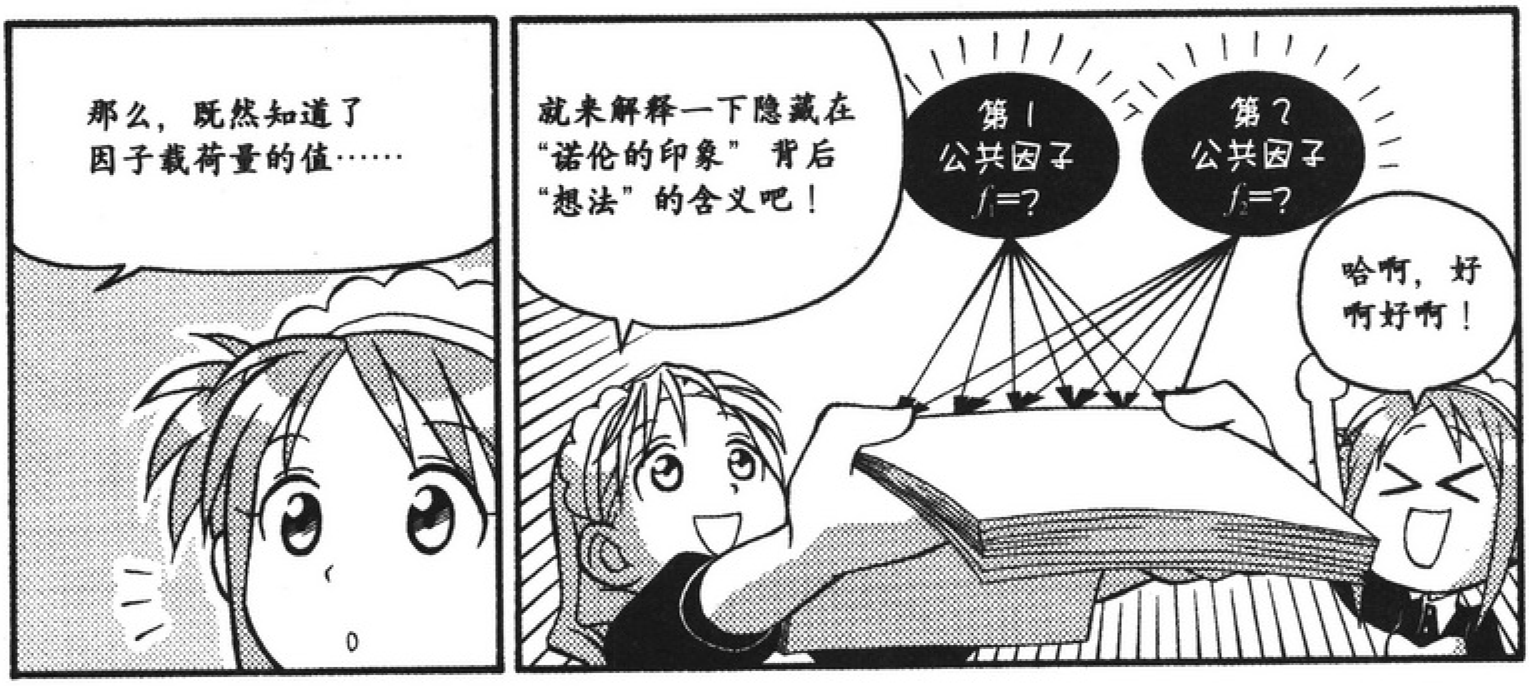




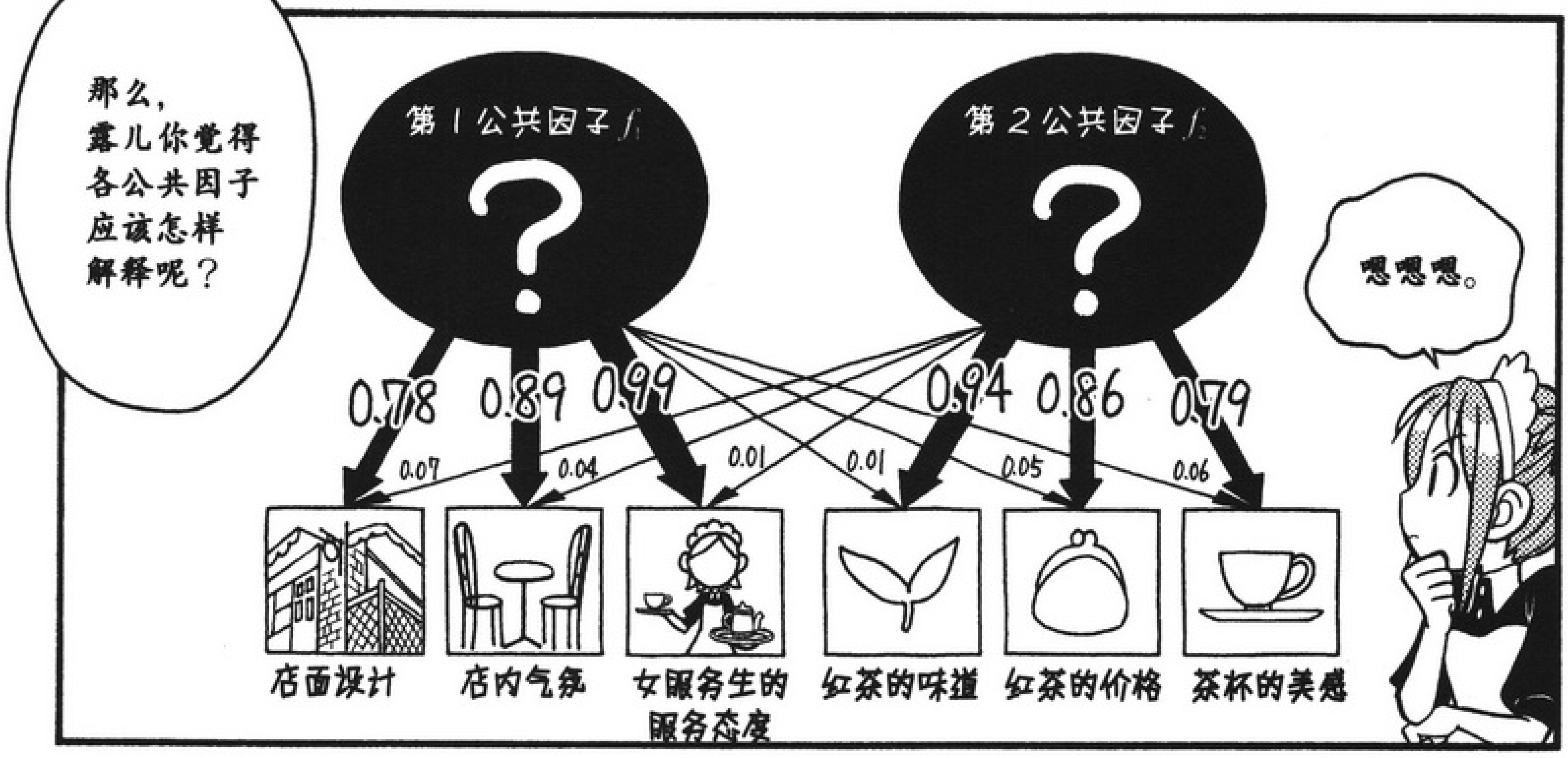
步骤 1	仅旋转	第 1 公共因子 第 2 公共因子
步骤 2	仅旋转	第 1 公共因子 第 3 公共因子
步骤 3	仅旋转	第 1 公共因子 第 4 公共因子
步骤 4	仅旋转	第 2 公共因子 第 3 公共因子
步骤 5	仅旋转	第 2 公共因子 第 4 公共因子
步骤 6	仅旋转	第 3 公共因子 第 4 公共因子



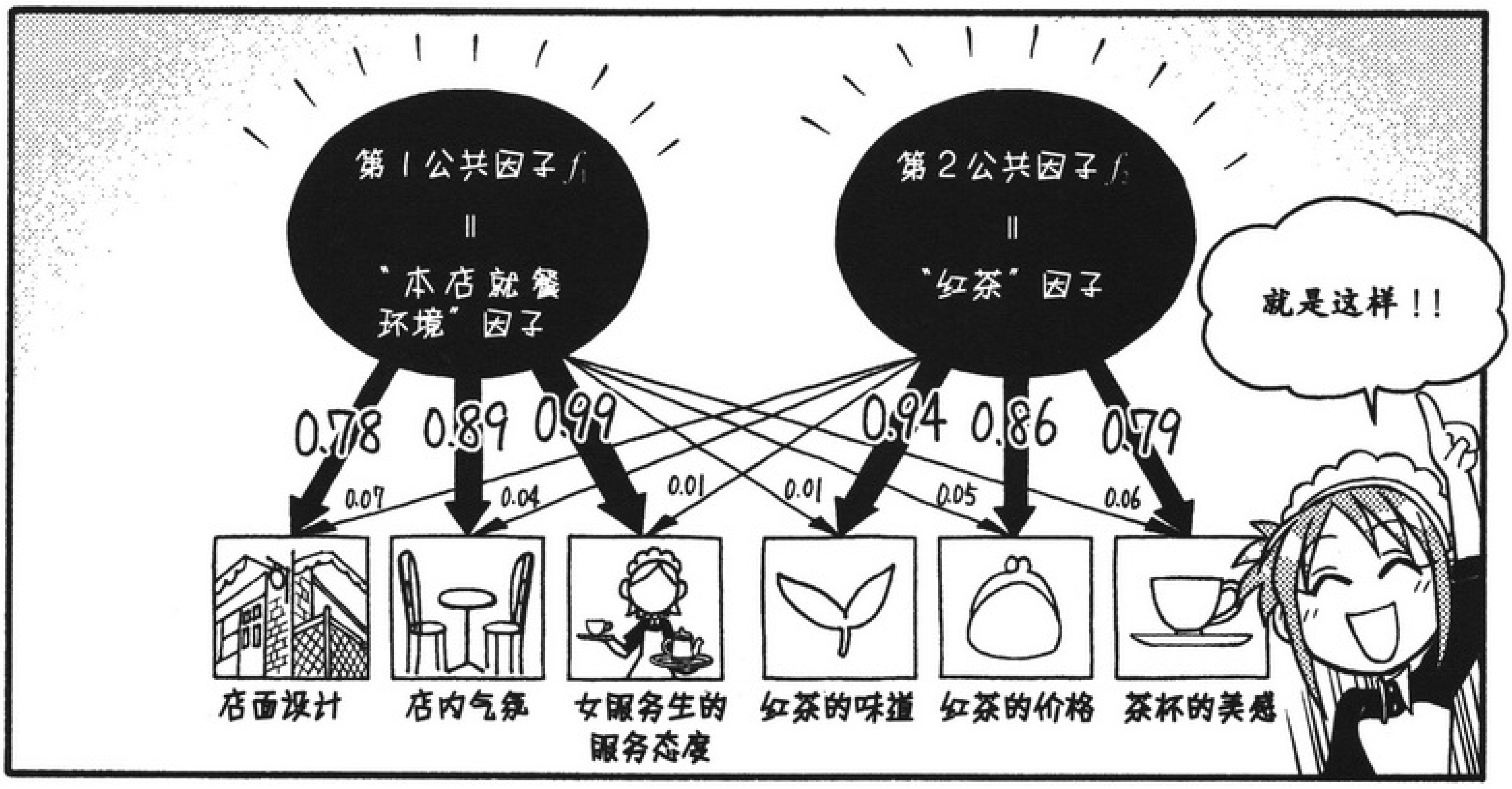
③ 解释各公共因子的含义



	第 1 公共因子	第 2 公共因子
Q1a. 店面设计	0.78	0.07
Q1b. 店内气氛	0.89	0.04
Q1c. 女服务生的服务态度	0.99	0.01
Q1d. 红茶的味道	0.01	0.94
Q1e. 红茶的价格	0.05	0.86
Q1f. 茶杯的美感	0.06	0.79

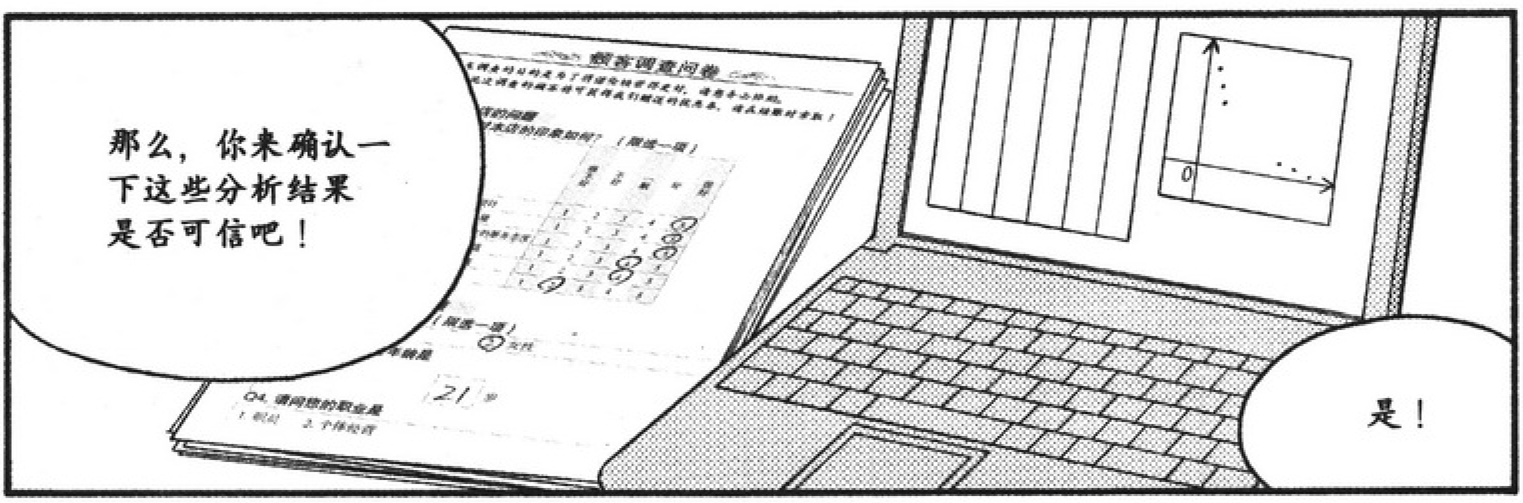

那么，露儿你觉得各公共因子应该怎样解释呢？
第1公共因子f_1
第2公共因子f_2
0.78
0.89
0.99
0.94
0.86
0.79
0.07
0.04
0.01
0.01
0.05
0.06
店面设计
店内气氛
女服务生的服务态度
红茶的味道
红茶的价格
茶杯的美感
嗯嗯嗯。


一个对“店面设计”、“店内气氛”、“女服务生的服务态度”影响较大。
另一个对“红茶的味道”、“红茶的价格”、“茶杯的美感”影响较大，所以……


第1公共因子f_1
=
“本店就餐环境”因子
第2公共因子f_2
=
“红茶”因子
0.78
0.89
0.99
0.94
0.86
0.79
0.07
0.04
0.01
0.01
0.05
0.06
店面设计
店内气氛
女服务生的服务态度
红茶的味道
红茶的价格
茶杯的美感
就是这样！！


“诺伦的印象”背后所隐藏的是对“店内就餐环境”和“红茶”的看法，对不对？
Tea Room NORNS
是这样的！
搞定了！我会了，我会了！
啊！露儿，还没完呢。
还没确认分析结果的精度呢，之后还要求出因子得分呢！
是这样啊！

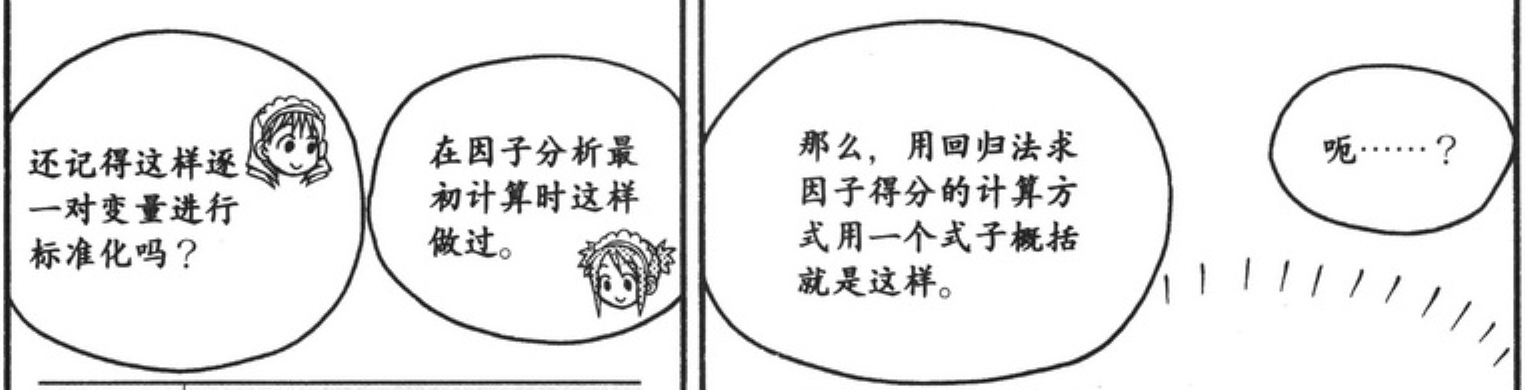
④ 确认分析结果的精度


那么，你来确认一下这些分析结果是否可信吧！
是！


由p个变量组成的相关矩阵的特征值之和就是p，你还记得吗？
记得！
每个变量应该分得的特征值是1吧？
特征值
$$\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix}$$
第1个
第2个
第p个
共p个
完全正确！
所以，同主成分分析一样，因子分析的成功与否也是通过累积贡献度的大小进行判断的。
哦！
但是，同主成分分析中贡献度的定义有所区别，
カッ カッ
拿眼前的这个例子来说，“第i公共因子的贡献度”就是这样。
哦……
用因子载荷量来定义啊！
$$第i公共因子的贡献度=\frac{b_{1i}^2+b_{2i}^2+b_{3i}^2+b_{4i}^2+b_{5i}^2+b_{6i}^2}{6}\times100$$

	$b_{1i}^2+b_{2i}^2+\cdots+b_{6i}^2$	贡献度	累积贡献度
第1公共因子	2.39	$\frac{2.39}{6}\times 100=39.8(\%)$	$\frac{2.39}{6}\times 100=39.8(\%)$
第2公共因子	2.26	$\frac{2.26}{6}\times 100=37.7(\%)$	$\frac{2.39}{6}\times 100+\frac{2.26}{6}\times 100=77.4(\%)$

⑤ 求出因子得分，充分理解每个个体的特征



	Q1a 的标准值 u_1		Q1f 的标准值 u_6
A	1.2	⋯	−0.9
⋮	⋮	⋮	⋮
O	−1.2	⋯	0.6

A 的第 1 公共因子得分	$f_{A1} = w_{11} \times 1.2 + \cdots + w_{61} \times (-0.9)$
O 的第 2 公共因子得分	$f_{O2} = w_{12} \times (-1.2) + \cdots + w_{62} \times 0.6$

变量逐一标准化之后的数据（参见 155 页）

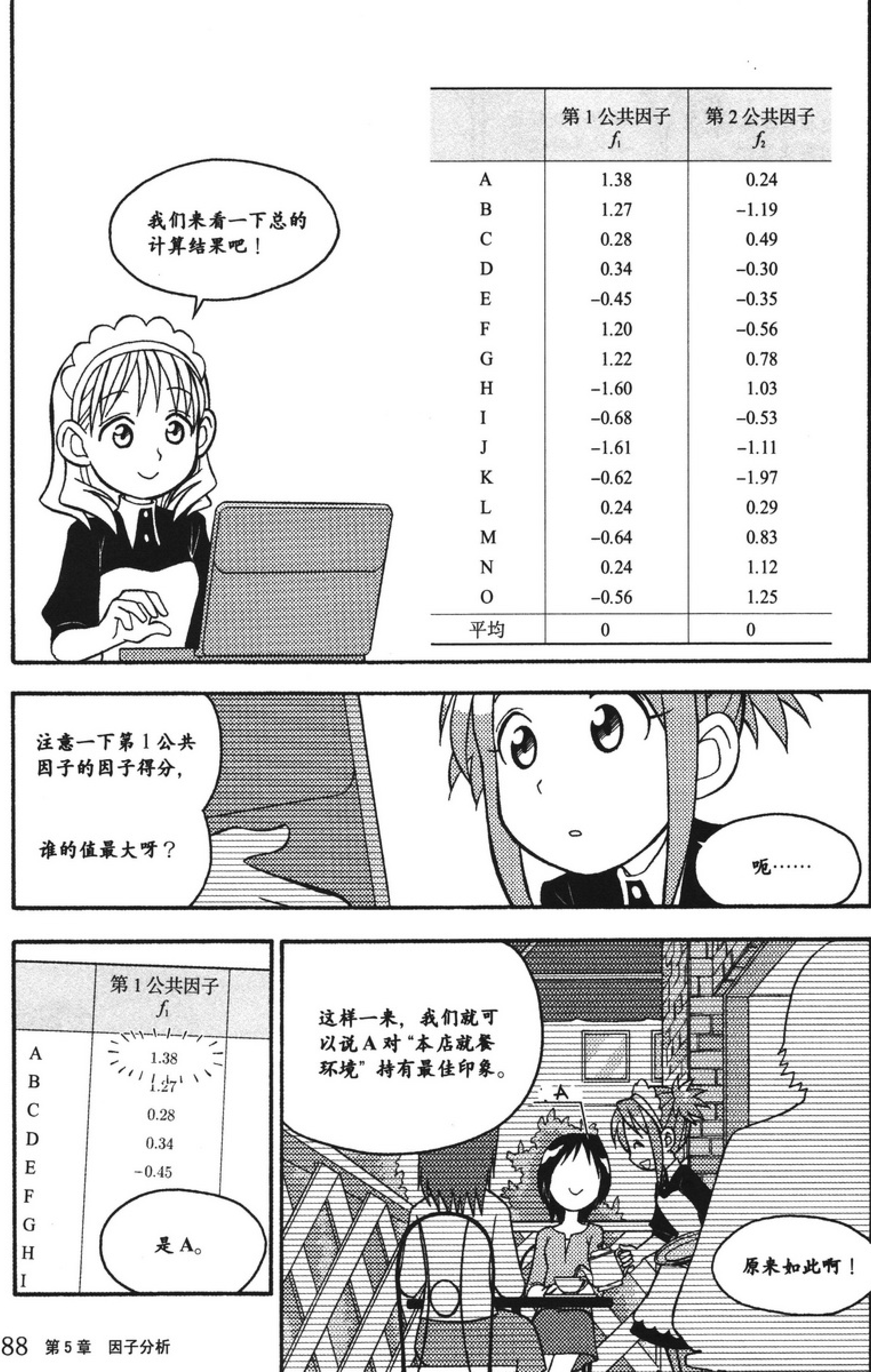
相关矩阵的逆矩阵（参见 166 页）

旋转后因子载荷量矩阵（参见 179 页）

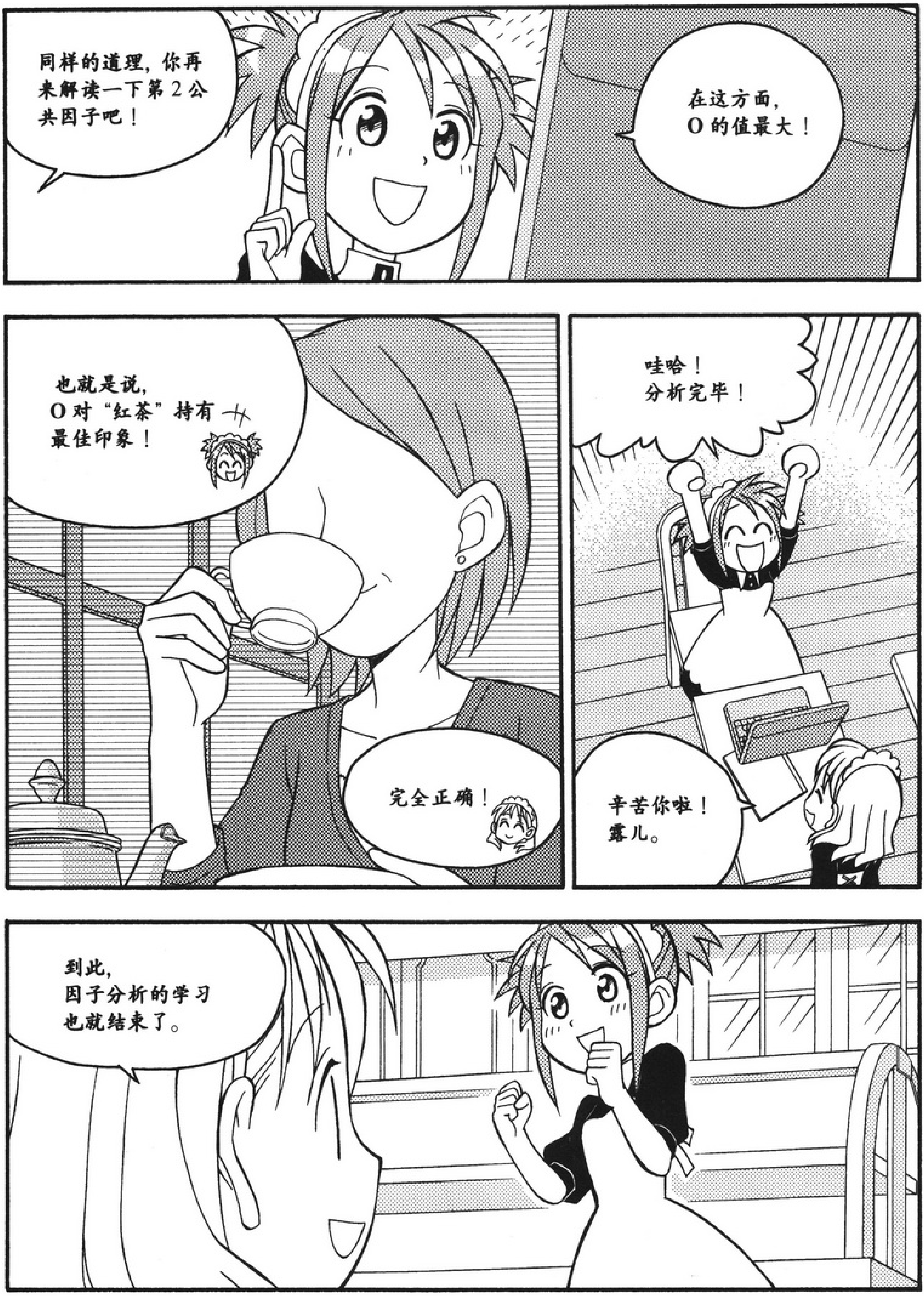
$$\begin{bmatrix} f_{A1} & f_{A2} \\ f_{B1} & f_{B2} \\ f_{C1} & f_{C2} \\ f_{D1} & f_{D2} \\ f_{E1} & f_{E2} \\ f_{F1} & f_{F2} \\ f_{G1} & f_{G2} \\ f_{H1} & f_{H2} \\ f_{I1} & f_{I2} \\ f_{J1} & f_{J2} \\ f_{K1} & f_{K2} \\ f_{L1} & f_{L2} \\ f_{M1} & f_{M2} \\ f_{N1} & f_{N2} \\ f_{O1} & f_{O2} \end{bmatrix} = \begin{bmatrix} 1.2 & 1.6 & 1.3 & 0.4 & 0.5 & -0.9 \\ 1.2 & 0.8 & 1.3 & -1.2 & -1.0 & -0.9 \\ 0.4 & 0.8 & 0.3 & 0.4 & 0.5 & 0.6 \\ -1.2 & 0.1 & 0.3 & -0.4 & -0.3 & -0.1 \\ -0.4 & 0.1 & -0.6 & -0.4 & 0.5 & -1.6 \\ 1.2 & 0.8 & 1.3 & -0.4 & -1.0 & -0.1 \\ 1.2 & 1.6 & 1.3 & 0.4 & 1.3 & 1.3 \\ -0.4 & -1.5 & -1.6 & 1.3 & 0.5 & 0.6 \\ 0.4 & -1.5 & -0.6 & -0.4 & -1.0 & -0.1 \\ -2.0 & -0.7 & -1.6 & -1.2 & -1.0 & -0.9 \\ -0.4 & -0.7 & -0.6 & -2.1 & -1.8 & -1.6 \\ 0.4 & 0.1 & 0.3 & 0.4 & -0.3 & 0.6 \\ -0.4 & -0.7 & -0.6 & 0.4 & 1.3 & 1.3 \\ 0.4 & 0.1 & 0.3 & 1.3 & 0.5 & 1.3 \\ -1.2 & -0.7 & -0.6 & 1.3 & 1.3 & 0.6 \end{bmatrix} \boxed{\begin{bmatrix} 1 & 0.65 & 0.80 & 0.11 & 0.01 & 0.14 \\ 0.65 & 1 & 0.89 & 0.02 & 0.19 & 0.01 \\ 0.80 & 0.89 & 1 & 0.02 & 0.04 & 0.10 \\ 0.11 & 0.02 & 0.02 & 1 & 0.82 & 0.77 \\ 0.01 & 0.19 & 0.04 & 0.82 & 1 & 0.64 \\ 0.14 & 0.01 & 0.10 & 0.77 & 0.64 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.78 & 0.07 \\ 0.89 & 0.04 \\ 0.99 & 0.01 \\ 0.01 & 0.94 \\ 0.05 & 0.86 \\ 0.06 & 0.79 \end{bmatrix}}$$

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

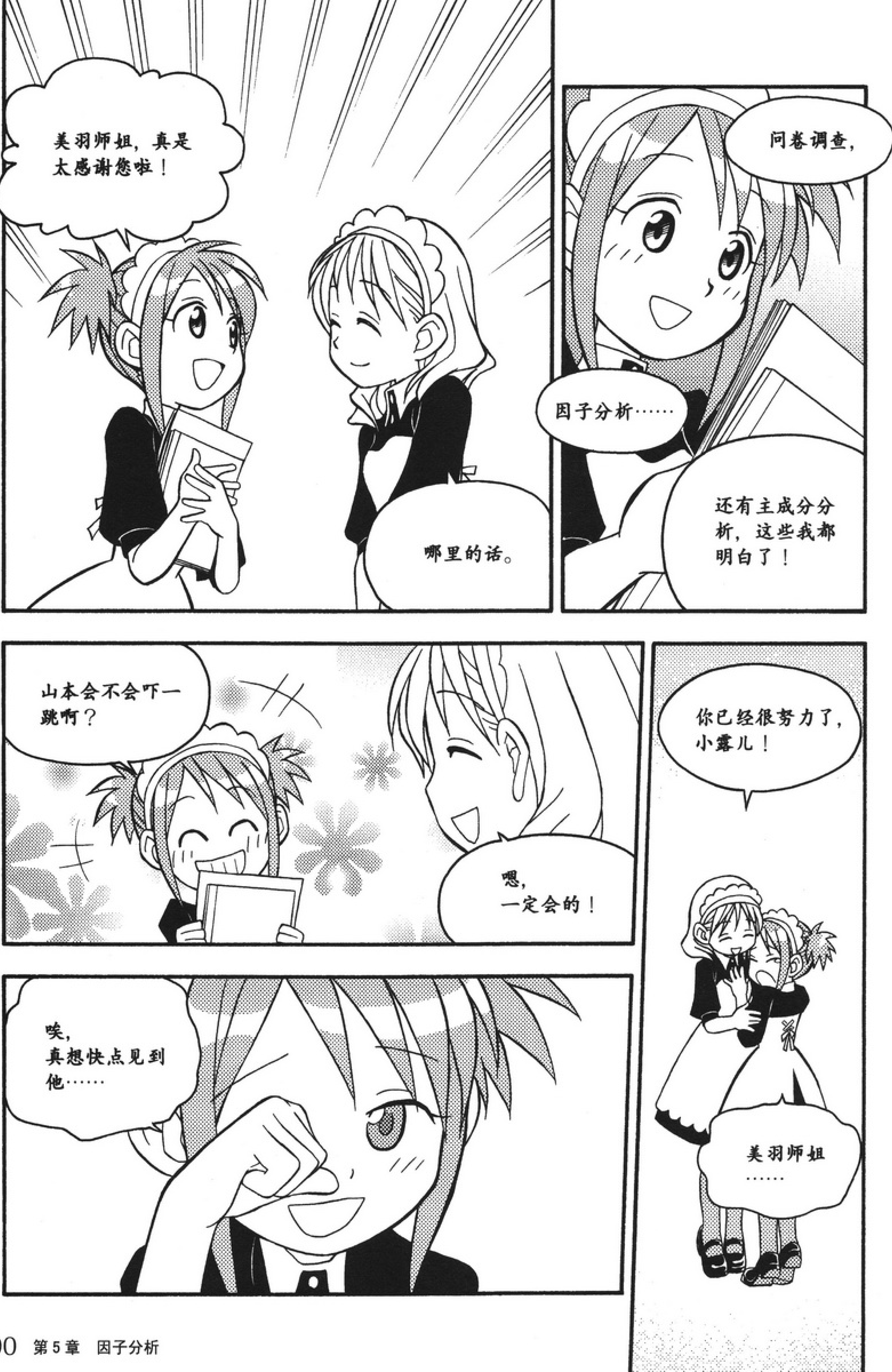
$$= \begin{bmatrix} 1.2 & 1.6 & 1.3 & 0.4 & 0.5 & -0.9 \\ 1.2 & 0.8 & 1.3 & -1.2 & -1.0 & -0.9 \\ 0.4 & 0.8 & 0.3 & 0.4 & 0.5 & 0.6 \\ -1.2 & 0.1 & 0.3 & -0.4 & -0.3 & -0.1 \\ -0.4 & 0.1 & -0.6 & -0.4 & 0.5 & -1.6 \\ 1.2 & 0.8 & 1.3 & -0.4 & -1.0 & -0.1 \\ 1.2 & 1.6 & 1.3 & 0.4 & 1.3 & 1.3 \\ -0.4 & -1.5 & -1.6 & 1.3 & 0.5 & 0.6 \\ 0.4 & -1.5 & -0.6 & -0.4 & -1.0 & -0.1 \\ -2.0 & -0.7 & -1.6 & -1.2 & -1.0 & -0.9 \\ -0.4 & -0.7 & -0.6 & -2.1 & -1.8 & -1.6 \\ 0.4 & 0.1 & 0.3 & 0.4 & -0.3 & 0.6 \\ -0.4 & -0.7 & -0.6 & 0.4 & 1.3 & 1.3 \\ 0.4 & 0.1 & 0.3 & 1.3 & 0.5 & 1.3 \\ -1.2 & -0.7 & -0.6 & 1.3 & 1.3 & 0.6 \end{bmatrix} \begin{bmatrix} -0.014 & -0.024 \\ -0.047 & -0.005 \\ 1.048 & 0.001 \\ -0.001 & 0.611 \\ 0.092 & 0.259 \\ -0.104 & 0.155 \end{bmatrix}$$



	第 1 公共因子 f_1	第 2 公共因子 f_2
A	1.38	0.24
B	1.27	−1.19
C	0.28	0.49
D	0.34	−0.30
E	−0.45	−0.35
F	1.20	−0.56
G	1.22	0.78
H	−1.60	1.03
I	−0.68	−0.53
J	−1.61	−1.11
K	−0.62	−1.97
L	0.24	0.29
M	−0.64	0.83
N	0.24	1.12
O	−0.56	1.25
平均	0	0



同样的道理，你再来解读一下第2公共因子吧！
在这方面，O的值最大！
也就是说，O对“红茶”持有最佳印象！
完全正确！
哇哈！分析完毕！
辛苦你啦！露儿。
到此，因子分析的学习也就结束了。


美羽师姐，真是太感谢您啦！
哪里的话。
问卷调查，
因子分析……
还有主成分分析，这些我都明白了！
山本会不会吓一跳啊？
嗯，一定会的！
你已经很努力了，小露儿！
唉，真想快点见到他……
美羽师姐……

某营销公司
高津部长，
您的电话。
好，
我知道了……
什么！？
已经到机场了！？
怎么会？
工作应该不会
这么快就结束啊
……
……
呀，
快把眼泪擦了。
啊哈哈，
真对不起。
啊，
不好意思，
是手机……
喂，您好……
啊，
是哥哥啊！
啊……
真的吗！？
露儿！
山本到机场了！！
啊……

✿ 4. 本章例子中的样本 ✿

在本章的例子中，总体和样本是如下定义的：

总体	所有光临诺伦的顾客
样本	× 月 × 日 15 时至 16 时之间，参加美羽和露儿所进行的问卷调查的全体顾客

无论怎么说，这些样本都不能算是通过无作为抽样法抽出的样本，而是由美羽和露儿根据自己的随意判断所进行的有意抽样法抽出的样本。

第一章曾讲过“如果样本不是‘总体的精确缩影’就没有意义了”，可是在介绍本书的主要内容——因子分析时，怎么会举出了一个与其相矛盾的例子呢？这肯定会给一些读者带来疑惑。也许您会认为这是强词夺理，不过在市场营销等领域进行数据分析时，上述情况确实是不可避免的。也就是说，“本来是通过有意抽样法得到的样本，却将它假设成是通过无作为抽样法得到的样本进行分析”，这种情况是不可避免的。如果不这样做的话，不用说因子分析，几乎什么分析都没有办法进行。

“本来是通过有意抽样法得到的样本，却将它假设成是通过无作为抽样法得到的样本进行分析”，如此做法，笔者认为，只要不是太混乱不堪的样本，并且事先对周围的人讲明具体情况，在实际操作中也是允许的。但是在学术研究中，是不是也允许这样做呢？允许与否，还要请您仔细留意样本的形成过程，再做决定。

✿ 5. 补充注意事项 ✿

以下总结了从 139 页到 149 页之间所讲过的因子分析的注意事项。

注意事项 1	各公共因子的含义，只能是在做完分析之后，由分析者主观地进行推断。
注意事项 2	同主成分分析中的各主成分有所不同，各公共因子是平等的，排序不分先后。
注意事项 3	对于因子分析的计算，分析者必须在分析前先假定出公共因子的个数。
注意事项 4	对同一组数据进行分析时，对于公共因子的个数有几种假定就会得到几种分析结果。这时使用哪种结果作为“最终的分析结果”，就要根据分析者自己的喜好而定了。
注意事项 5	虽然说在因子分析中，可以隐藏着多个公共因子，但是受计算方法所限，实际操作中公共因子的个数最多也不会超过因变量的个数。
注意事项 6	（图形所占空间过大，因此省略）请参见 143 页。
注意事项 7	在因子分析的计算中，首先要对分析对象的数据逐一进行变量标准化。
注意事项 8	（图形所占空间过大，因此省略）请参见 145 页。
注意事项 9	实际上，因子分析并不是一种不需要分析者思考，公共因子就会自动浮现的魔法般的分析方法。
注意事项 10	因子分析是为了确定因子载荷量的值而存在的分析方法。

对于这些注意事项，下面做几点补充。

注意事项 1 的补充

没有什么需要特别补充的。

注意事项 2 的补充

没有什么需要特别补充的。

注意事项 3 的补充

分析者一定要在分析之前假定出公共因子的个数。虽然听起来有些荒谬，但因子分析确实是这样进行的。

事实上，所假定的公共因子的个数是有其数学上的标准的。就是 141 页所讲的“相关矩阵中大于 1 的特征值的个数”。此外，还相当于“在 Scree Plot（碎石图）中，曲线变缓前的特征值的个数”。所谓 Scree Plot 就是将特征值由大到小依次列出后画出的折线图。

在一般的介绍因子分析的书中，大体上都会给出上面所讲的这段文字。不过，笔者还是犹豫，只是为了所谓的“标准”，便把“特征值的个数”、“数学上”这样的话拿出来讲，这样做是否合适。但是，如果不想得复杂些，只是像第 141 页所讲的那样，先假定各式各样的公共因子的个数，再随意地做各种形式的分析，这样做又是不切实际的。

注意事项 4 的补充

同之前所讲的注意事项 3 相比，这个注意事项 4 更为重要。

分析者根据自己的喜好来决定“最终的分析结果”，对此可能有的读者会想“这么做可以吗？”

实际上，还是有一些正规的方法可供我们使用的，比如用以下这些方法来确定“最终的分析结果”会比较合适。

· 以“相关矩阵中大于 1 的特征值的个数”为公共因子个数时的分析结果。

· 以“在 Scree Plot（碎石图）中，曲线变缓前的特征值的个数”为公共因子个数时的分析结果。

· 以“累积贡献度的值达到一定程度[1]时特征值的个数”为公共因子个数时的分析结果。

· 以“拟合优度检验[2]中失去意义的个数”作为公共因子个数时的分析结果。

· 以“拟合优度指标[3]的值达到最佳的个数”作为公共因子个数时的分析结果。

不过，至少就笔者的经验看来，我们不能过于依赖这些方法。也就是说，我们不能过于期待以下这种情况会发生。

> 先假定公共因子的个数为 2、3、4，或者其他，而后随意进行分析。总觉得 3 个公共因子的情况最适合作为分析结果。再查看一下此时的相关矩阵中大于 1 的特征值的个数，正好是 3。到底是正规方法啊，果然信得过。

1. 笔者所谓的“一定程度”指的是“50%”。
2. 稍后再对“拟合优度检验”进行介绍。
3. “拟合优度指标”本书不作介绍。

分析者根据自己的喜好来决定“最终的分析结果”，笔者这么说毕竟还是有些言过其实。但是尽管如此，我们也不要过分依赖现在所讲的这些正规的方法。结论还是要视分析者自己的判断而定。

注意事项 5 的补充

至少在我们所见到的，使用 Varimax 法或 Promax 法旋转来进行因子分析的过程中，所能假定的公共因子个数是有上限的。详细内容以后再讲。

注意事项 6 的补充

没有什么需要特别补充的。

注意事项 7 的补充

没有什么需要特别补充的。

注意事项 8 的补充

在第 4 章的开篇时我们曾讲过，主成分分析和因子分析是两种不同的分析方法。尽管如此，恐怕还是会有一些软件将主成分分析和因子分析混为一谈，存在同样误解的人也不在少数。请您再看一下 138 页的图。无论怎样，都不会有“主成分分析 = 因子分析”的情况。

注意事项 9 的补充

在第 148 页中我们曾讲过，对于因子分析来说，只有当“这些因变量背后应该隐藏着这样的公共因子”这一假设在某种程度上成立时，因子分析才能顺利进行。换句话说，只要这个假设越接近实际情况，那么，分析顺利进行的可能性，即推导出分析者想要的结果的可能性也就越高。这样看来，因子分析这种分析方法比较像大家所说的“没有悬念的比赛”[1]。

1. 硬要说成“放水的比赛”又有些不妥，但不管怎么说，因子分析者往往通过因子分析产生出对自己有利的结果，这是目前因子分析的实际情况。

在做因子分析时，一定要对调查问卷中的问题仔细推敲。不过，即便是仔细推敲过，所得到的结果还是会令您惊讶“这是什么啊”。所以说这是一种万不可麻痹大意的分析方法。“手边刚好有最近做过的问卷调查的数据，就凑合着用这个吧”，存在类似这种想法的话是绝对做不好因子分析的！

注意事项 10 的补充

一般情况下，我们大概会像如下所述的那样对因子分析进行定义。

· 通过少数几个公共因子来说明因变量之间的相关关系的分析方法。
· 找出背后所隐藏的公共因子的分析方法。

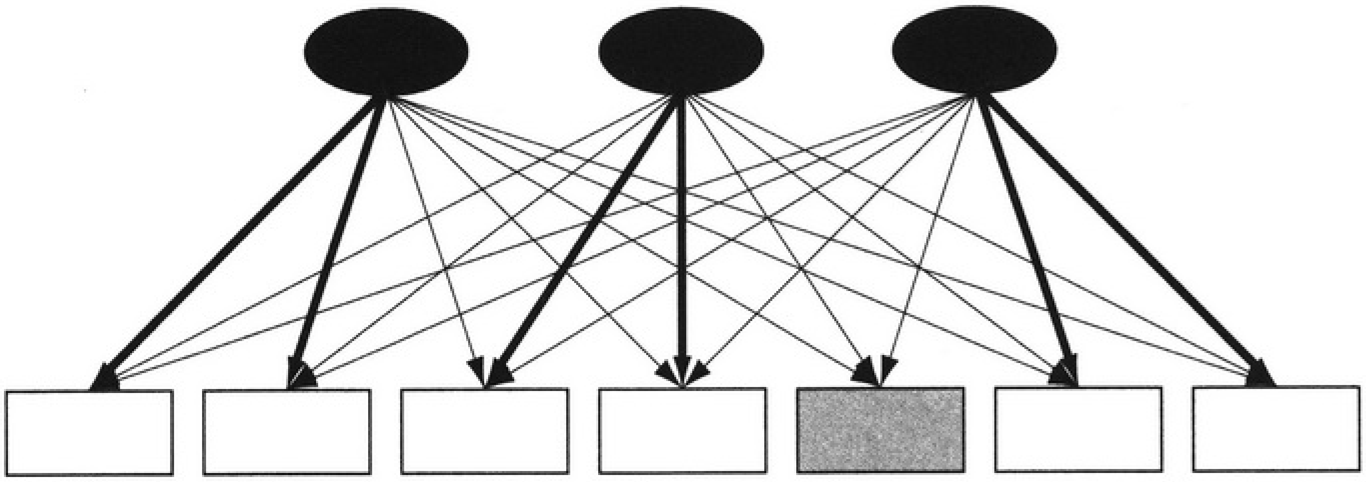
笔者对于哪一种定义都不是十分认同。为什么呢？首先，拿前者来说，确实让人觉得这是一个在数学上很妥当的定义，但除非是“独具慧眼之人”，否则那些有生以来第一次见到这个定义的人，还真难免不会颔首认同“确实如此”。再拿后者来说，正如之前注意事项 9 中所讲的那样，因子分析可谓是一场“没有悬念的比赛”，所以很明显这个定义是不正确的。但是，我认为将直观上非常容易理解的定义作为学习因子分析的“第一步”也并非是一件坏事。

笔者将“为了确定因子载荷量的值而存在的分析方法”作为因子分析的定义。也许会有读者感到疑惑“这难道不是**验证型因子分析**[1]的定义吗？”请不要误会，笔者并没有将因子分析定义为“为了‘精确地’确定因子载荷量的值而存在的分析方法”。

1. 实际上，因子分析可以大致分为**探索型因子分析**和**验证型因子分析**（**确认型因子分析**）。通常我们将前者称为“因子分析”，本书所讲的也正是前者。

✿ 6. 因子载荷量小的变量的处理方法 ✿

在做因子分析时，会有像下图这样意外的情况发生，就是任何一个公共因子都不会对某个因变量产生太大影响。



◆图5.1　任何一个公共因子都不会对某个因变量产生太大影响的情况
（因子载荷量的绝对值在 0.5 以上的，用粗箭头标记）

这种情况下，建议您采用以下任意一种方法进行处理：

· 将这个因变量剔除后再进行因子分析。

· 不剔除这个因变量，而是将“因子载荷量的绝对值为 0.5 以上”这个划分标准，按照“0.5 以上→ 0.45 以上→ 0.4 以上→……”这样的趋势逐步下降，迫使“任何一个公共因子都不会对某个因变量产生太大影响”的情况不再发生。

后者中“因子载荷量的绝对值为 XX 以上”中的 XX，并不存在统计学上的依据，不过通常是 0.3 到 0.5 之间某一个值。

✿ 7. 极大似然法 ✿

7.1 极大似然法概要

在因子载荷量的计算方法中，比较著名的除了我们讲过的主因子法，还有极大似然法[1]。下面通过本章的例子来讲解一下这种计算方法。

L= 因变量的个数 + log（X 的行列式）— X 的主对角线上的值的和，其中

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

$$=\begin{bmatrix} 1 & 0.65 & \cdots & 0.14 \\ 0.65 & 1 & \cdots & 0.01 \\ \vdots & \vdots & \ddots & \vdots \\ 0.14 & 0.01 & \cdots & 1 \end{bmatrix}\left[\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{bmatrix}\begin{bmatrix} a_{11} & a_{21} & \cdots & a_{61} \\ a_{12} & a_{22} & \cdots & a_{62} \end{bmatrix}+\begin{bmatrix} d_1^2 & 0 & \cdots & 0 \\ 0 & d_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_6^2 \end{bmatrix}\right]^{-1}$$

极大似然法就是求解令 L 值最大的$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ \vdots & \vdots \\ a_{61} & a_{62} \end{bmatrix}$的计算方法[2]。

最近好像有消息称“极大似然法似乎比主因子法好”。如此一来，笔者担心会让人们走进一个误区，就是无论什么情况都认为使用极大似然法为佳。请您注意，使用极大似然法的前提是总体中的数据要服从多变量正态分布[3]。

7.2 拟合优度检验

为什么会说极大似然法比主因子法更好一些呢？原因之一就是它能够进行“拟合优度检验”[4]。所谓“拟合优度检验”，大体上讲，就是进行如下讨论的检验。

1. 顺便提一下，还有**最小二乘法**、**正则化最小二乘法**这些计算方法。
2. 本书对**行列式**不做介绍。
3. 多变量正态分布的概率密度函数写起来是一个非常繁杂的式子，同时式中还需要对很多符号做出解释，为了避免这种情况，这部分内容从略。直观上理解，顾名思义，可以认为是正态分布的多变量版本。
4. 认为极大似然法好的其他理由，在本书中不做介绍，但是可以求出“拟合优度指标”的值也是其中一个原因。

原假设	公共因子的个数是 m
备择假设	公共因子的个数不是 m

在“拟合优度检验”中，如果 p 值比有意义的标准小，则说明备择假设成立，也就是说可以得到“公共因子的个数不是 m”这一结论。但如果 p 值大，则说明原假设并没有错，也就是说可以得到这样的结论："公共因子的个数可能是 m 也可能不是 m，再进一步扩大解释的话，就是可以当作 m 左右"。

“拟合优度检验”的优势在于，可以大概地掌握公共因子的个数。但是尽管如此，并不表示最恰当的个数会自动地显现出来，具有如此强大功能的检验是不存在的。

✿ 8. 旋转与 Varimax 法 ✿

正如第 174 页所讲，旋转存在着很多种方法。但是不知为什么，不管是论文也好，报告也好，实际情况却是——只有 Varimax 法被广泛地使用着。

为何只是 Varimax 法？笔者没有从事过“因子分析史”的研究，所以其中原委便不得而知了。不过，数年前曾听某位统计学者这样说过："做旋转的时候，如果用的不是 Varimax 法，周围的人会责问，但如果是 Varimax 法便没有人会说什么了，仅此而已。"因此，我们可以假设发生了如下的演变：

① 很久以前，研究人员（非统计学者）发表了关于 Varimax 法的研究结果。

② 获悉这些研究结果的其它研究人员觉得“因子分析似乎也挺有意思的”，他们自己也发表关于 Varimax 法的研究结果。

③ 随着时间的推移，关注这一领域的研究结果、以及追随①和②的研究结果变得庞大起来。这样一来就给人留下了一个并没有什么根据的印象，认为“因子分析的旋转指的就是 Varimax 法”。

④ “仔细想想，无论做什么研究都要用 Varimax 法”，存在这种想法的研究人员过去并不是没有，只是因为不想引来麻烦，所以即便不是很明白其中的原因，也会顺应潮流地称“因子分析的旋转指的就是 Varimax 法”。

⑤ 随着①和②的发展，研究结果变得日益丰富，逐渐形成了现在的局面。

请您注意，笔者并不是认为 Varimax 法“不妥”“不能用”，只是将“为何一提起旋转就会想到 Varimax 法”作为一个话题进行讨论。

笔者估计如果我们对之后要讲的 Promax 法也稍做研究的话，可能就会同 Varimax 法一样，以相同的模式发展吧。

✿ 9. 因子载荷量矩阵和因子结构矩阵 ✿

本节的内容有些抽象，因此本想说“不擅长数学的读者可以跳过本节不做阅读”，但是如果不理解本节内容的话便无法理解下一节的内容，所以请您尽量克服困难。

正如第 172 页所讲，我们将$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pm} \end{bmatrix}$称为**因子载荷量矩阵**或**因子模式矩阵**。

从第 2 公共因子 f_2 指向因变量 p 的因子载荷量。

将$\begin{bmatrix} r_{1f_1} & r_{1f_2} & \cdots & r_{1f_m} \\ r_{2f_1} & r_{1f_2} & \cdots & r_{2f_m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{pf_1} & r_{pf_2} & \cdots & r_{pf_m} \end{bmatrix}$称为**因子结构矩阵**。

第 2 公共因子 f_2 同因变量 p 的单相关系数。

在正交因子模型的情况下，因子载荷量矩阵和因子结构矩阵相一致。也就是说有这样的关系成立：

$$a_{61} = r_{6f_1}$$

第 1 公共因子 f_1 指向因变量 6 的因子载荷量　第 1 公共因子 f_1 同因变量 6 的单相关系数

如果是斜交因子模型，则这个关系不成立。

值得注意的是，在下表中记录的是第 152 页的数据和第 188 页的因子得分的一部分。

◆表5.1　第152页的数据和第188页的因子得分的一部分

	Q1a 店面的设计	第 1 公共因子 f_1
A	5	1.38
B	5	1.27
C	4	0.28
D	2	0.34
E	3	−0.45
F	5	1.20
G	5	1.22
H	3	−1.60
I	4	−0.68
J	1	−1.61
K	3	−0.62
L	4	0.24
M	3	−0.64
N	4	0.24
O	2	−0.56

求得它们的单相关系数的值为 0.79，这与第 179 页因子载荷量矩阵中的 b_{11}=0.78 并不一致。这是因为上表中的因子得分不是“准确值”，而是“估计值”[1]。

1. 如第 186 页所讲，因子得分的计算方法有“回归法”、“Bartlett 法”和“Anderson-Rubin 法”，等等。计算方法如此多种多样也就意味着，并不存在一种独一无二的方法可以求出因子得分的“准确值”。

✿ 10. Promax 法 ✿

10.1 Promax法的概要

正如之前所讲的那样,斜交旋转中最著名的方法当属 Promax 法。所谓 Promax 法，笼统地讲，就是遵循以下步骤进行的方法。

① 按照 Varimax 法进行旋转。

② 假定一个“真的因子载荷量矩阵”，也就是说“就现有的经验来讲，总体的形式应该是这样”。我们将这个矩阵称为“目标矩阵”。

③ 旋转①中的轴，使其尽可能地靠近②中所假定的目标矩阵。

但是，让我们感到困惑的是我们并不知道总体的情况，这样也就无法找出之前第②步中所假定的目标矩阵。接下来，让我们就本章的例子，来讲一讲如何使用 Promax 法进行旋转。

$$C = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \\ c_{31} & c_{32} \\ c_{41} & c_{42} \\ c_{51} & c_{52} \\ c_{61} & c_{62} \end{bmatrix} = \begin{bmatrix} 0.98599 & 0.00005 \\ 0.99538 & \boxed{0.00001} \\ 0.99966 & 0.00000 \\ 0.00000 & 0.99966 \\ 0.00001 & 0.99270 \\ 0.00003 & 0.98987 \end{bmatrix}$$

$$\frac{\sqrt{b_{21}^2 + b_{22}^2}}{b_{22}} \times \left| \frac{b_{22}}{\sqrt{b_{21}^2 + b_{22}^2}} \right|^{\kappa+1} = \frac{\sqrt{0.89^2 + 0.04^2}}{0.04} \times \left| \frac{0.04}{\sqrt{0.89^2 + 0.04^2}} \right|^{4+1} = 0.00001$$

按照上述计算所求得的矩阵 C，我们就将其看成是目标矩阵。并且运算过程中所出现的 b_{21} 和 b_{22} 等参数，就是通过 Varimax 法旋转后得到的因子载荷量矩阵 B 的值。

$$B=\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \\ b_{41} & b_{42} \\ b_{51} & b_{52} \\ b_{61} & b_{62} \end{bmatrix}=\begin{bmatrix} 0.78 & 0.07 \\ 0.89 & 0.04 \\ 0.99 & 0.01 \\ 0.01 & 0.94 \\ 0.05 & 0.86 \\ 0.06 & 0.79 \end{bmatrix}$$

κ 是要由分析者自己规定的值，这里我们虽然给出的是 4，但通常情况下使用 2、3 或 4 都是可以的。

可能有些读者还是感觉有些模糊，不是非常理解。也就是说，可能存在着如下疑惑：

· 凭什么就能断言“目标矩阵 = 矩阵 C”？

· κ 的值由分析者自己主观地进行规定，这样听起来不是很荒谬吗?

· Promax 法原本应该是斜交旋转，但是为何最初要用 Varimax 法这种正交旋转法进行旋转。

即便您可以理解这种方法，也要注意。请您转换思路想一想，令“目标矩阵 = 矩阵 C”就是 Promax 法的一部分；对 κ 值的主观规定也是 Promax 法的一部分；最初要用 Varimax 法进行旋转还是 Promax 法的一部分。

10.2 因子载荷量矩阵、因子相关矩阵以及因子结构矩阵

就本章的例子来讲，通过 Promax 法进行旋转之后的因子载荷量矩阵、**因子相关矩阵**以及因子结构矩阵，可以通过以下计算求得。所谓因子相关矩阵，就是记录着公共因子相互间的单相关系数的矩阵。

在以下讲解中出现了过多的矩阵，所以，为了便于讲解，我们将矩阵命名为“P”和“Q”。

■ 因子载荷量矩阵 P

$$P = \underbrace{\begin{bmatrix} 0.78 & 0.07 \\ \vdots & \vdots \\ 0.06 & 0.79 \end{bmatrix}}_{\text{前一页中的}B} \underbrace{\begin{bmatrix} 1.11 & -0.05 \\ -0.05 & 1.15 \end{bmatrix}}_{Q} \underbrace{\begin{bmatrix} 0.90 & 0 \\ 0 & 0.88 \end{bmatrix}}_{D} = \begin{bmatrix} 0.78 & 0.03 \\ 0.89 & 0.01 \\ 0.99 & -0.03 \\ -0.03 & 0.94 \\ 0.02 & 0.86 \\ 0.02 & 0.79 \end{bmatrix}$$

$$Q = \left[\underbrace{\begin{bmatrix} 0.78 & \cdots & 0.06 \\ 0.07 & \cdots & 0.79 \end{bmatrix}}_{B\text{ 的转置矩阵}} \underbrace{\begin{bmatrix} 0.78 & 0.07 \\ \vdots & \vdots \\ 0.06 & 0.79 \end{bmatrix}}_{B} \right]^{-1} \underbrace{\begin{bmatrix} 0.78 & \cdots & 0.06 \\ 0.07 & \cdots & 0.79 \end{bmatrix}}_{B\text{ 的转置矩阵}} \underbrace{\begin{bmatrix} 0.98599 & 0.00005 \\ \vdots & \vdots \\ 0.0003 & 0.98987 \end{bmatrix}}_{202\text{ 页中的 }C} = \begin{bmatrix} 1.11 & -0.05 \\ -0.05 & 1.15 \end{bmatrix}$$

$$d = \left[\underbrace{\begin{bmatrix} 1.11 & -0.05 \\ -0.05 & 1.15 \end{bmatrix}}_{Q\text{ 的转置矩阵}} \underbrace{\begin{bmatrix} 1.11 & -0.05 \\ -0.05 & 1.15 \end{bmatrix}}_{Q} \right]^{-1} \begin{bmatrix} 0.81 & 0.07 \\ 0.07 & 0.77 \end{bmatrix}$$

$$D = \underbrace{\begin{bmatrix} \sqrt{0.81} & 0 \\ 0 & \sqrt{0.77} \end{bmatrix}}_{\text{将 }d\text{ 中主对角线上的值取平方根，除此之外的值全部取 0}} = \begin{bmatrix} 0.90 & 0 \\ 0 & 0.88 \end{bmatrix}$$

■ 因子相关矩阵 L

$$L=\underbrace{\begin{bmatrix}0.9991 & 0.0414\\0.0421 & 0.9991\end{bmatrix}}_{T}\underbrace{\begin{bmatrix}0.9991 & 0.0421\\0.0414 & 0.9991\end{bmatrix}}_{T\text{的转置矩阵}}=\begin{bmatrix}1 & 0.08\\0.08 & 1\end{bmatrix}$$

$$T=\left[\underbrace{\begin{bmatrix}1.11 & -0.05\\-0.05 & 1.15\end{bmatrix}}_{Q}\underbrace{\begin{bmatrix}0.90 & 0\\0 & 0.08\end{bmatrix}}_{D}\right]^{-1}=\begin{bmatrix}0.9991 & 0.0414\\0.0421 & 0.9991\end{bmatrix}$$

■ 因子结构矩阵 S

$$S=\underbrace{\begin{bmatrix}0.78 & 0.03\\0.89 & 0.01\\0.99 & -0.03\\-0.03 & 0.94\\0.02 & 0.86\\0.02 & 0.79\end{bmatrix}}_{P}\underbrace{\begin{bmatrix}1 & 0.08\\0.08 & 1\end{bmatrix}}_{L}=\begin{bmatrix}0.78 & 0.10\\0.89 & 0.08\\0.99 & 0.05\\0.05 & 0.94\\0.09 & 0.86\\0.09 & 0.79\end{bmatrix}$$

10.3 分析结果的精度

经过 Promax 法旋转后所得到的分析结果，其精度并不是通过贡献率或累积贡献率来评价的，而是用被称作“**排除其他公共因子影响后的贡献**”或“**忽略其他公共因子后的贡献**”的指标进行评价的，此外便没有什么评价指标了。

“排除其他公共因子影响后的贡献”的概念理解起来比较困难，所以这里我们只介绍“忽略其他公共因子后的贡献”[1]。

“忽略其他公共因子后第 i 公共因子的贡献”，就是将因子结构矩阵中第 i 公共因子所对应的各单相关系数的值取平方再按列相加后便可得到。就本章的例子来讲，如下所示。

	忽略其他因子后的贡献
第 1 公共因子	$r_{1f_1}^2+r_{2f_1}^2+r_{3f_1}^2+r_{4f_1}^2+r_{5f_1}^2+r_{6f_1}^2$ $=0.78^2+0.89^2+0.99^2+0.05^2+0.09^2+0.09^2$ $=2.40$
第 2 公共因子	$r_{1f_2}^2+r_{2f_2}^2+r_{3f_2}^2+r_{4f_2}^2+r_{5f_2}^2+r_{6f_2}^2$ $=0.10^2+0.08^2+0.05^2+0.94^2+0.86^2+0.79^2$ $=2.28$

这个值越大就意味着其所对应的公共因子与多个因变量的关联性越强。但是它并不能像 Varimax 法那样，求出“忽略其他公共因子后的贡献‘率’”。

“忽略其他公共因子后的贡献”不是绝对的，而是相对的。也就是说，它仅仅是一个只能说出“这个公共因子比那个公共因子大，还是那个公共因子比这个公共因子大”的、十分宽泛的评价指标。因此，不可否认的是，尽管您花费了一番功夫将它求出来，可最终还是不能完全理解它的意义。

1. 说到底，是因为概念不易理解才不做介绍的，但这并不意味着“排除其他公共因子影响后的贡献”不可用。

10.4 因子得分

就本章的例子来讲，经过 Promax 法旋转后，基于回归法的因子得分可以按照以下计算过程求得。

逐一进行变量标准化后的数据（请参见第 155 页）　相关矩阵的逆矩阵（请参见第 166 页）　因子结构矩阵（请参见第 205 页）

$$
\begin{bmatrix}
f_{A1} & f_{A2}\\
f_{B1} & f_{B2}\\
f_{C1} & f_{C2}\\
f_{D1} & f_{D2}\\
f_{E1} & f_{E2}\\
f_{F1} & f_{F2}\\
f_{G1} & f_{G2}\\
f_{H1} & f_{H2}\\
f_{I1} & f_{I2}\\
f_{J1} & f_{J2}\\
f_{K1} & f_{K2}\\
f_{L1} & f_{L2}\\
f_{M1} & f_{M2}\\
f_{N1} & f_{N2}\\
f_{O1} & f_{O2}
\end{bmatrix}
=
\begin{bmatrix}
1.2 & 1.6 & 1.3 & 0.4 & 0.5 & -0.9\\
1.2 & 0.8 & 1.3 & -1.2 & -1.0 & -0.9\\
0.4 & 0.8 & 0.3 & 0.4 & 0.5 & 0.6\\
-1.2 & 0.1 & 0.3 & -0.4 & -0.3 & -0.1\\
-0.4 & 0.1 & -0.6 & -0.4 & 0.5 & -1.6\\
1.2 & 0.8 & 1.3 & -0.4 & -1.0 & -0.1\\
1.2 & 1.6 & 1.3 & 0.4 & 1.3 & 1.3\\
-0.4 & -1.5 & -1.6 & 1.3 & 0.5 & 0.6\\
0.4 & -1.5 & -0.6 & -0.4 & -1.0 & -0.1\\
-2.0 & -0.7 & -1.6 & -1.2 & -1.0 & -0.9\\
-0.4 & -0.7 & -0.6 & -2.1 & -1.8 & -1.6\\
0.4 & 0.1 & 0.3 & 0.4 & -0.3 & 0.6\\
-0.4 & -0.7 & -0.6 & 0.4 & 1.3 & 1.3\\
0.4 & 0.1 & 0.3 & 1.3 & 0.5 & 1.3\\
-1.2 & -0.7 & -0.6 & 1.3 & 1.3 & 0.6
\end{bmatrix}
\begin{bmatrix}
1 & 0.65 & 0.80 & 0.11 & 0.01 & 0.14\\
0.65 & 1 & 0.89 & 0.02 & 0.19 & 0.01\\
0.80 & 0.89 & 1 & 0.02 & 0.04 & 0.10\\
0.11 & 0.02 & 0.02 & 1 & 0.82 & 0.77\\
0.01 & 0.19 & 0.04 & 0.82 & 1 & 0.64\\
0.14 & 0.01 & 0.10 & 0.77 & 0.64 & 1
\end{bmatrix}^{-1}
\begin{bmatrix}
0.78 & 0.10\\
0.89 & 0.08\\
0.99 & 0.05\\
0.05 & 0.94\\
0.09 & 0.86\\
0.09 & 0.79
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
1.39 & 0.29\\
1.22 & -1.14\\
0.30 & 0.50\\
0.33 & -0.29\\
-0.46 & -0.36\\
1.17 & -0.51\\
1.25 & 0.83\\
-1.56 & 0.96\\
-0.70 & -0.23\\
-1.65 & -1.17\\
-0.71 & -2.00\\
0.25 & 0.30\\
-0.61 & 0.80\\
-0.28 & 1.13\\
-0.51 & 1.22
\end{bmatrix}
$$

✿ 11. 能够假定的公共因子个数的上限 ✿

至少我们讲过的使用 Varimax 法或 Promax 法旋转进而完成的因子分析中，所能够假定的公共因子的个数是存在上线的。具体地讲就是，公共因子的个数一定要满足以下关系，请您记住这一点。

$$\text{公共因子的个数} \leqslant \frac{2\times\text{因变量的个数}+1-\sqrt{8\times\text{因变量的个数}+1}}{2}$$

根据以上不等式计算能够假定的公共因子个数的上限，计算结果见下表，请您参考。

◆表5.2　能够假定的公共因子个数的上限

因变量的个数		能够假定的公共因子个数的上限	因变量的个数		能够假定的公共因子个数的上限
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. 主因子法和 Varimax 法真的过时了吗 ✿

可能本节的内容不太容易理解，但是如果是初次接触因子分析的读者，还请您一定要阅读本节内容。

在本章的例子中，因子载荷量的计算方法采用的是主因子法，旋转方法采用的是 Varimax 法。实际上，主因子法也好，Varimax 法也好，都可以说是“过时的方法”了，都已经渐渐地被人们当作“不适合的方法”了。具体地讲就是，如今“主因子法+Varimax 法”已经渐渐地被“极大似然法+Promax 法”所取代。

将主因子法和 Varimax 法看成是“过时的方法”，笔者对这种作法持怀疑态度。理由如下所述。

· 主因子法从整体上来讲，就是所谓的**谱分解**[1]，这种计算方法并无“新旧”、“好坏”之论。

· 同极大似然法相比，主因子法没有像“总体必须要服从多变量正态分布”这样较强的约束条件。

· 主因子法看上去比较复杂，但实际上计算起来比极大似然法容易[2]。

· 再看一看用来代替 Varimax 法的 Promax 法，其目标矩阵、κ 值的取值方法都存在着莫名其妙之处，所以也不能无条件地称颂这种旋转方法。

· 虽然在 Varimax 法中，我们假定“任意两公共因子间的单相关系数的值为 0”的作法也不是很合理，但是，若只是因为“计算比较容易”、“电脑的性能达不到要求”等这些并不影响我们探求真理的理由，就突然将长久以来所默认的 Varimax 法看作是“过时的方法”、“不适合的方法”，这样做是不是太不合理了。

· 如果将主因子法和 Varimax 法当作“过时的方法”、“不适当的方法”，就等于是在说“根据主因子法和 Varimax 法得来的研究结果不值得一看，也没有参考价值，可以忽略不计”。然而现有[3]的研究结果几乎都是由“主因子法+Varimax 法”得来的，要是按照刚才的意思，岂不是在说我们找不到能够依据的资料了[4]。

---

1. 粗略地讲，第 76~78 页所讲的就是谱分解。
2. 笔者主观认为。
3. 这一部分写于 2006 年秋。
4. 如果对周围的人不假思索地说“你怎么还用主因子法、Varimax 法来做因子分子啊？真拿你没办法啊”之类的话，笔者认为，结果反而会令说这些话的人感到难堪。

笔者认为，对于因子载荷量的计算方法来说，主因子法和极大似然法的区别并不在于“新与旧”、“好与坏”，而应当是“流派”的不同。就旋转方法来说，虽然Varimax法中“任意两公共因子间的单相关系数的值为0”的假设确实有些不合理，但是尽管如此，Promax法中也有难以消除的、莫名其妙之处。因此，笔者总觉得同Promax法相比，Varimax法还是有可取之处的。

✿ 13. 因子分析中的术语 ✿

本书始终将因子分析中的因变量称为“因变量”，但是还请读者注意，通常情况下我们也将其称为**观测变量**。

还有，有时也将公共因子称为**潜变量**，将因子载荷量称为**路径系数**。


谢谢您，
司机师傅！
ハア
山本，你
在哪……
山本先生……
ハア
ハア
ずる
ぺた
ハア
ずる
ぺた


露，
露儿……
啊，
呵呵……
还是那身打扮，
一点没变……
呀，怎么
会穿着拖
鞋？
啊！这个……
都是因为回来
得太匆忙了
……
那家伙就是山本
笨蛋！


我害怕再也见不到你了！
怎么都不和我联系啊！？
啊……？
呃……
对了！露儿现在上大学了，有很多事情要做，没什么事就别联系她了！
那，那个……
真对不起……
山本，啊啊啊啊啊！！
啊！
啊！部长！
爸爸？
山本！你的工作怎么样了？
啊，都做完了啊……
做了那么多事就是不想让你再打搅露儿！


您说什么!?
爸爸!!
我不必跟你解释!!
是爸爸您在我上高中时，为了让我学习统计学所以才请来了山本老师!
那也不行!
爸爸您不讲道理!!
虽然山本……穿得很土气，也不太会说话，但是……
他很诚实，也很善良……
实际上，山本的这些优点爸爸您也是知道的!?
呃……
ザグッ

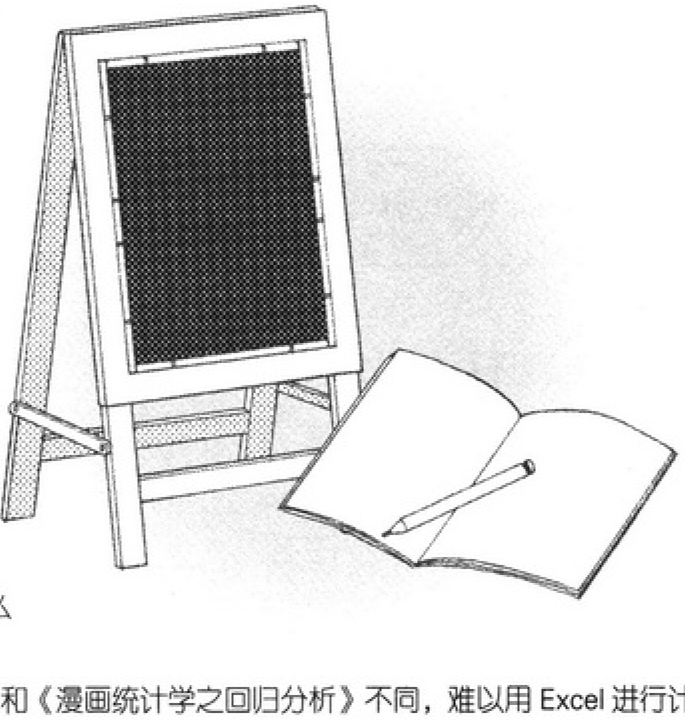
叔叔！
高津部长！
我真……失败啊……
露儿为了要成为山本的助手……
每天都在学因子分析。
啊？是因子分析吗？
厉害！
山本！
啊，什么？
希望可以赶快见到露儿，所以才……
拼命地工作……
没想到山本提前回来了！
这个……我也……嗯……

欢迎你回来，山本！
ばっ
太好了，小露儿！
那家伙，的表情！
山本！要是让我女儿受委屈的话，我不会放过你的！！
啊！不会的……
爸爸……
好吧，各位！那我们就在诺伦餐厅为山本举行一个回国party，怎么样？
哈，好主意！
一起去吧！叔叔！
噢，我就不去了吧！
不管怎么说您都得去啊！
知道啦！别推我啦……
啊，哈哈……


走吧！
山本。
好！
从现在起……
让我们一起学习吧！

附录
各种各样的分析方法

1. 多变量分析
   1.1 多变量分析的概要
   1.2 重回归分析
   1.3 Logistic 回归分析
   1.4 聚类分析
   1.5 对应分析以及数量化Ⅲ类
   1.6 结构方程模型
2. 其　他
   2.1 统计的假设检验
   2.2 Kaplan-Meier 法



这里为大家介绍一些曾在《漫画统计学》和《漫画统计学之回归分析》出现过的、比较著名的分析方法。

- 存在着什么样的分析方法
- 各种分析方法的具体内容
- 各种分析方法可以用来做什么

请您按以上思路进行阅读。

此外，本书与《漫画统计学》和《漫画统计学之回归分析》不同，难以用 Excel 进行计算，所以本书中对使用 Excel 计算的步骤不做介绍。

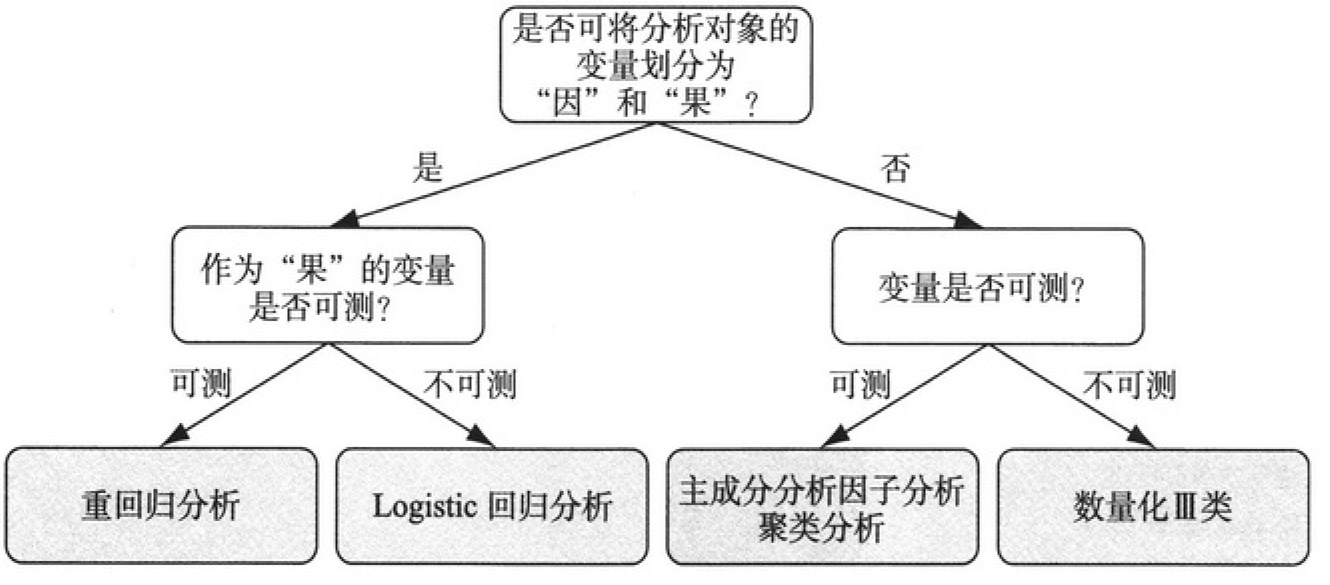
✿ 1. 多变量分析 ✿

1.1　多变量分析的概要

如第 11 页所述，多变量分析是对由多个变量组成的数据进行分析的分析方法的统称，如下表所示。

	变量 1	变量 2	…	变量 p
受访者 1	34	1	…	171.7
受访者 2	27	0	…	156.8
⋮	⋮	⋮	⋮	⋮
受访者 n	19	1	…	178.3

属于“多变量分析”这一范畴内的分析方法，除了本书所讲的主成分分析和因子分析以外，还有很多种类。下图中列出几种代表性的方法。



本节将针对上图中的内容进行如下介绍。

- 重回归分析
- Logistic 回归分析
- 聚类分析
- 数量化Ⅲ类

同时也会介绍**对应分析**以及**结构方程模型**。

1.2 重回归分析

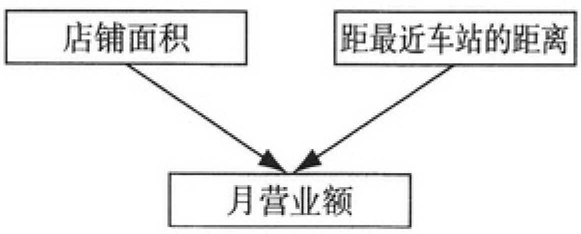
重回归分析是基于多个自变量来预测数值的一种分析方法。

■ 具体实例

下表中记录的是来自“风见面包房”（一家面包连锁店）的数据。

	店铺面积（m^2）	距最近车站的距离（m）	月营业额（万日元）
梦之丘总店	10	80	469
寺井站大楼店	8	0	366
曾根店	8	200	371
桥本大街店	5	200	208
桔梗町店	7	300	246
邮电局前店	8	230	297
水道町站前店	7	40	363
六条站大楼店	9	0	436
若叶川沿线店	6	330	198
美里店	9	180	364

假定各变量间存在如下关系：



然后进行重回归分析，可推导出下式：

$$y = 41.5x_1 - 0.3x_2 + 65.3$$

月营业额　店铺面积　距最近车站的距离

Point!

对重回归分析感兴趣的读者，可以参见本系列图书中的《漫画统计学之回归分析》。该书中对此处所举的具体实例的重回归分析作出了详细介绍。

1.3 Logistic回归分析

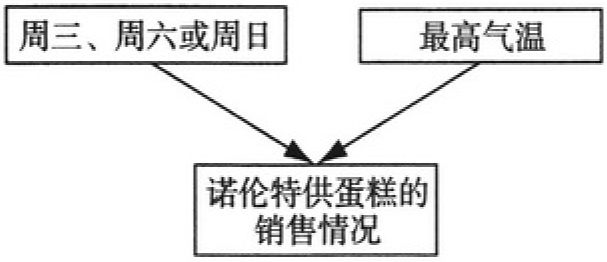
Logistic 回归分析是基于多个自变量来预测概率的一种分析方法。

■ 具体实例

下表记录的是一种称为“诺伦特供”蛋糕的销售情况，这种蛋糕在露儿打工的茶餐厅——诺伦茶餐厅，每日限售 1 个。

	周三、周六或周日	最高气温（℃）	诺伦特供蛋糕的销售情况
5 日（一）	0	28	1
6 日（二）	0	24	0
7 日（三）	1	26	0
8 日（四）	0	24	0
9 日（五）	0	23	0
10 日（六）	1	28	1
11 日（日）	1	24	0
12 日（一）	0	26	1
13 日（二）	0	25	0
14 日（三）	1	28	1
15 日（四）	0	21	0
16 日（五）	0	22	0
17 日（六）	1	27	1
18 日（日）	1	26	1
19 日（一）	0	26	0
20 日（二）	0	21	0
21 日（三）	1	21	1
22 日（四）	0	27	0
23 日（五）	0	23	0
24 日（六）	1	22	0
25 日（日）	1	24	1

假定各变量间存在如下关系：



然后进行 Logistic 回归分析，可以推导出下列：

$$\underset{\substack{\uparrow\\ \text{诺伦特色的销售情况}}}{y}=\frac{1}{1+e^{-(2.44\underset{\substack{\uparrow\\ \text{周三、周六或周日}}}{x_1}+0.54\underset{\substack{\uparrow\\ \text{最高气温}}}{x_2}-15.20)}}$$

代入不同的 x_1 与 x_2 的值，就能够估算出 y 值。

Point!

对 Logistic 回归分析感兴趣的读者，可以参见本系列图书中的《漫画统计学之回归分析》。该书中对此处所举的具体实例的 Logistic 回归分析作出了详细介绍。

1.4 聚类分析

聚类分析（Cluster Analysis），就是基于分析对象间的距离将分析对象分为若干类的一种分析方法。所谓“分析对象”指的就是个体或者变量。

也许有读者会这样认为“不用刻意去做所谓的聚类分析，只做主成分分析或者因子分析不是也能够将个体和变量进行分类吗”。这种想法确实没错，但是主成分分析或因子分析只能进行粗略的分类，也就是说仅仅只是“看着散点图，指出某个受访者和某个受访者好像比较类似”这种程度上的分类。若想进行更为严格的分类，就要使用聚类分析进行分类了。

■ 具体实例

下表记录的是某个补习班上中学三年级学生的测验结果。

	语文	社会	理科	英语	数学
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

假定将其分为 2 类，进行聚类分析，可以得到以下结果。

第 1 类	第 2 类
A	D
B	E
C	F
G	H
J	I

Point!

在阅读具体实例的过程中，也许会有不少读者感到疑惑。实际上，在聚类分析中分类的个数是分析者在分析之前就已经“判断”出来的。这同因子分析中公共因子的个数是分析者在分析之前就已经“假定”出来的情况是一样的。

聚类分析，是基于分析对象间的距离来对分析对象进行分类的，说到底是一种数学分析方法。“第 1 类是由具有怎样特征的人所构成的集合”，这是在进行聚类分析之前所不由知道的。各类的特征只有在进行聚类分析之后，才能由分析者“事后诸葛”般地、“主观地”做出推断，比如“第 1 类是学习好的人，那第 2 类就是学习不好的人吧”。

聚类分析存在着很多“流派”，也就是很多种计算方法。

请您再仔细阅读一下本栏所写的内容。您就会发现，聚类分析——这种分析方法中的许多环节，不免要被指责为“主观臆断”。这些内容在普通的书籍中不会被提到，所以先在这里提示一下。

1.5 对应分析以及数量化Ⅲ类

对应分析（Correspondence Analysis）是一种同**数量化Ⅲ类**非常类似的分析方法。由于对应分析相对容易理解，所以这里先来介绍这种分析方法，之后再来介绍本节的正题——数量化Ⅲ类。此外，对应分析并不属于一般的多变量分析的范畴之内。

对应分析是一种将列联表（又称交叉分类表，Cross-tabulation Table）画成散点图的分析方法。更确切地说，就是将“列联表的行和列”中“能够充分表达列联表信息的值”呈现出来的一种分析方法。也可以理解为像是在为列联表拍摄航空摄影一样的分析方法。

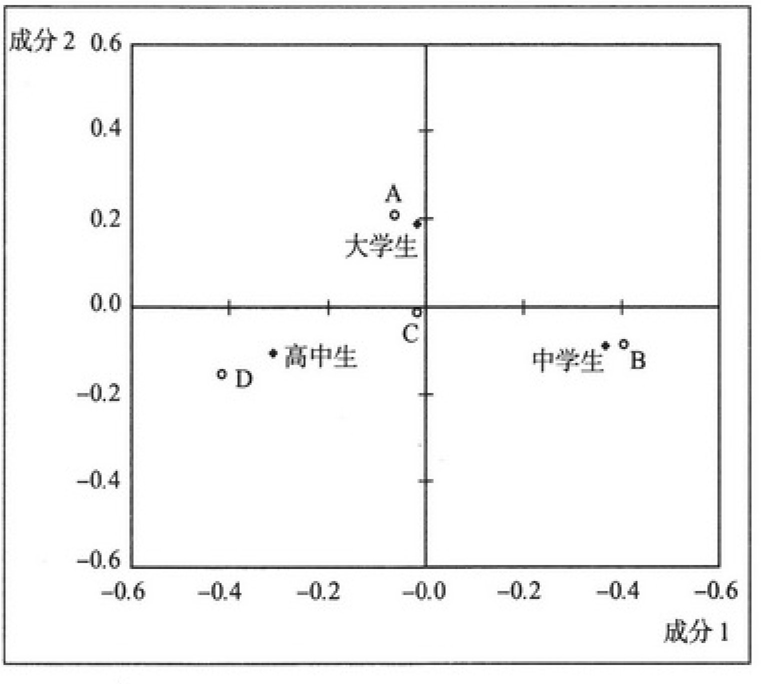
■具体实例

下表是以初中生、高中生和大学生为对象进行的某个问卷调查的结果，并整理成列联表的形式。

（单位：人）

		最喜爱的艺术家				合　计
		A	B	C	D	
学生类型	初中生	10	19	12	5	47
	高中生	13	8	15	16	52
	大学生	18	11	14	8	51
合　计		41	38	42	29	150

对上表进行对应分析，可以得出如下图所示的结果。



一眼就能看出这样的结论：“初中生喜欢 B”，“高中生喜欢 D”，“大学生喜欢 A”。

接下来进入本节的正题。数量化Ⅲ类，可以说是“一种以原始数据为对象的对应分析”。

· 将原始数据画成散点图的分析方法

· 将“作为原始数据的受访者和变量”中“能够充分表达原始数据的信息的值”呈现出来的分析方法

· 像是在为原始数据拍摄航空摄影一样的分析方法

对应分析的对象……

列联表

	甲	乙	丙	丁
A				
B				
C				

数量化Ⅲ类的对象……

原始数据

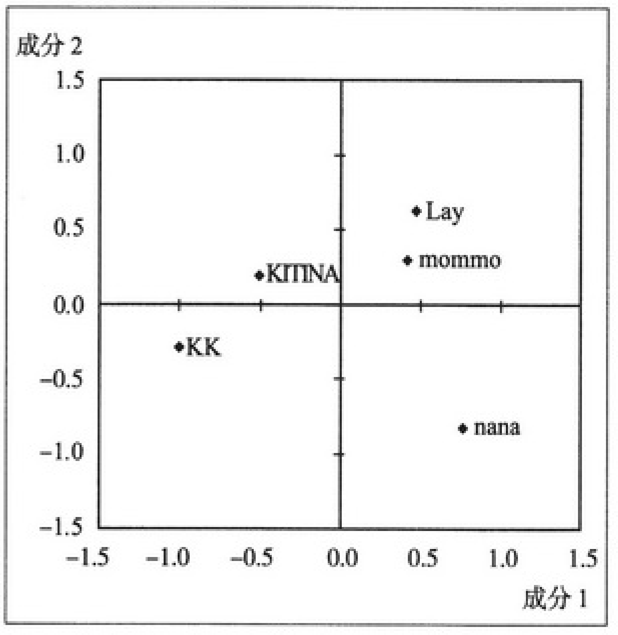
	变量 1	变量 2	变量 3	变量 4
A				
B				
C				
D				
E				
F				
G				
H				
I				
J				

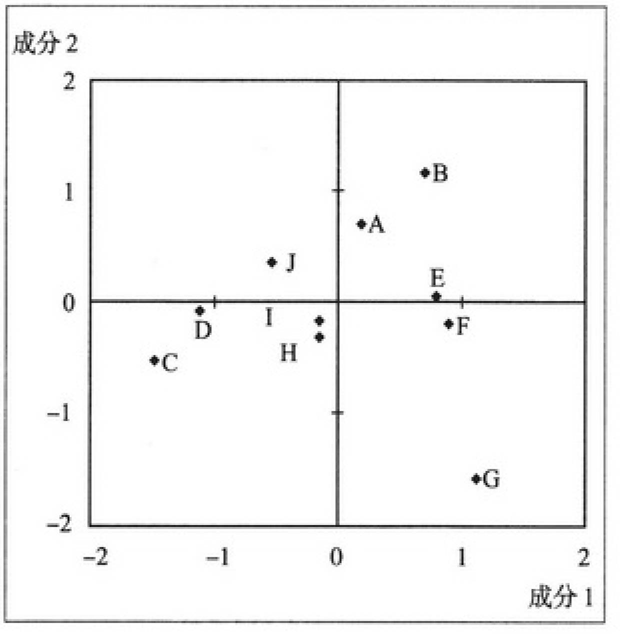
■ 具体实例

下表记录的是对 20 多岁女性所喜爱的杂志的调查结果。

	KK	nana	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

对上表进行数量化Ⅲ类分析，可以得出如下图所示的结果。





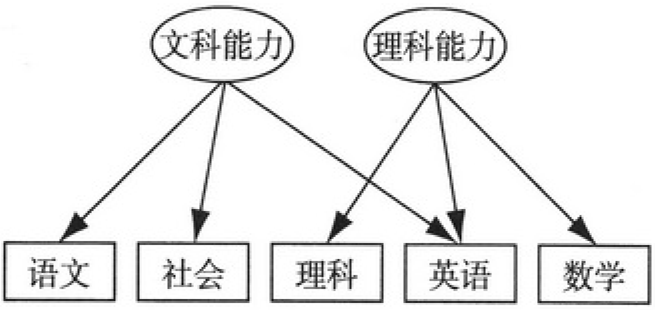
一眼就能看出这样的结论：“喜欢 nana 的是 G”，“喜欢 KK 的是 C”。

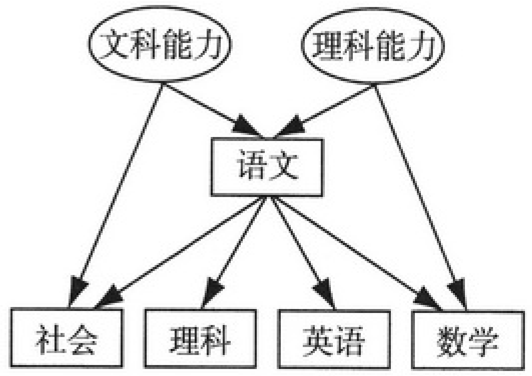
Point!

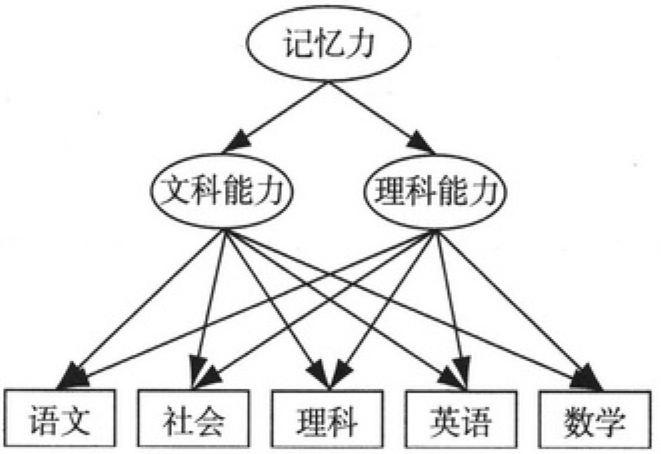
对数量化Ⅲ类和对应分析感兴趣的读者，可以参见对应分析的相关书籍。

1.6 结构方程模型

我们将以下这种图称为路径图。







路径图，是一种用来表示分析者的“主观”假设的图，分析者认为“实际操作中一定存在着图中所示的结构”。矩形表示的变量是可观测变量，椭圆形表示的变量是潜变量。

结构方程模型，这种分析方法是用来验证分析者所画的路径图是否和真实情况相一致，也就是用来验证分析者所做的主观假设——“实际操作中一定存在着图中所示的结构”是否和真实情况相一致。也是为了求出**路径系数**（相当于因子载荷量、即各箭头的具体值）而存在的分析方法。

“结构方程模型”，这个名称是由 Structural Equation Modeling 翻译得来。由于这个名字过长，所以通常将其简称为“SEM”。

通常，结构方程模型更多地被称为**协方差结构**分析。

■ 具体实例

从略。

Point!

结构方程模型，相当于第 40 页所讲过的“验证型”分析方法。也就是说，是一种应当按照如下顺序进行的分析方法。

① 建立假设

② 收集数据

③ 进行分析

但是，我们还是不能充分理解这些步骤，因为在具体操作时，我们可能会陷入以下这些自相矛盾的处境。

· 不知道存在着几个潜变量

· 不知道各潜变量和各观测变量之间存在着怎样的关系

· 不知道箭头应当指向哪里

“在第①步中，即便不知道，也要通过自己的思考将它们确定，这已经成为使用结构方程模型的分析者的工作”，只有充分地认识到这一点之后才能进行分析。

实际操作中，有一些很好的结构方程模型软件。这固然是值得高兴的。但是软件太好用了也会带来弊端，比如说会导致“结构方程模型很简单”——这样的误解蔓延开来。实际上，结构方程模型并不简单。并且对于结构方程模型来说，如果不能求解的话，基本上就意味着这个分析是失败的。因此，特别是对于那些从业人员，还是不要轻易地在周围的人（尤其是客人）面前提出“我想在这次的分析中挑战一下 SEM”为好。

✿ 2. 其　他 ✿

2.1　统计的假设检验

统计的假设检验，是分析者通过样本数据，来推测其建立的关于总体的假设是否正确的一种分析方法，通常简称为“检验”。

正如刚才所讲，统计的假设检验，是分析者通过样本数据，来推测其建立的关于总体的假设是否成立的一种分析方法。但是绝对不能认为“只要‘p 值’小，就意味着‘有意义’，数学上就能讲得通”、这样做“容易得到数学权威人士的认可”。由于存在这种误解的人有很多，所以从现在起，不单是正在学习的人，哪怕是学过的人也要请您予以充分的注意。

“统计的假设检验”并不是某一种分析方法的名称，而是一个统称。统计的假设检验包括

- 总体均值差的检验（即所谓的 t 检验）
- 独立性检验（即所谓的 χ^2 检验）
- 总体比例差的检验
- 总体方差比的检验
- Wilcoxon 检验

■ 具体实例

● 总体均值差的检验

推测“东京地区全体工薪阶层平均每月的零用钱额度”和“大阪地区全体工薪阶层平均每月的零用钱额度”之间存在着怎样的差异。

	地区	零用钱额度（日元）	
A	东京	42500	平均 41060 日元
B	东京	40800	
C	东京	39400	
D	东京	42800	
E	东京	39800	
F	大阪	38700	平均 39260 日元
G	大阪	40000	
H	大阪	38500	
I	大阪	42100	
J	大阪	37000	

● **独立性检验**

推测总体中“学生类型”和“最喜爱的艺术家”之间的 Cramer 关联系数（本书中不做讲解）的值是否为 0，换句话说就是推测“学生类型”和“最喜爱的艺术家”之间是否相关联。

（单位：人）

		最喜爱的艺术家				合　计
		A	B	C	D	
学生类型	初中生	10	19	12	5	47
	高中生	13	8	15	16	52
	大学生	18	11	14	8	51
合　计		41	38	42	29	150

Point!

请您不要认为“从数据中不是可以清楚地知道，东京的工薪阶层收入的多一些嘛！”必须注意的是，表中所记录的是样本信息而并非总体的信息。这里再次重申，统计的假设检验，是分析者通过样本数据，来推测其建立的关于总体的假设是否成立的一种分析方法。

统计学的假设检验虽然很有名，但却不像人们所想的那样简单，所以对于以上问题的分析结果就不做介绍了。

对统计学的假设检验感兴趣的读者，可以参见本系列图书中的《漫画统计学》。

2.2 Kaplan-Meier法

Kaplan-Meier 法是一种估计生存率的方法。Kapla 和 Meier 都是人的名字，前者指的是 Edward Kaplan，后者指的是 Paul Meier。

Kaplan-Meier 法的特点就在于它能够估计生存率。例如，分析者将肺癌患者作为关心对象，在同时参考以下患者数据之后对生存率进行估计。

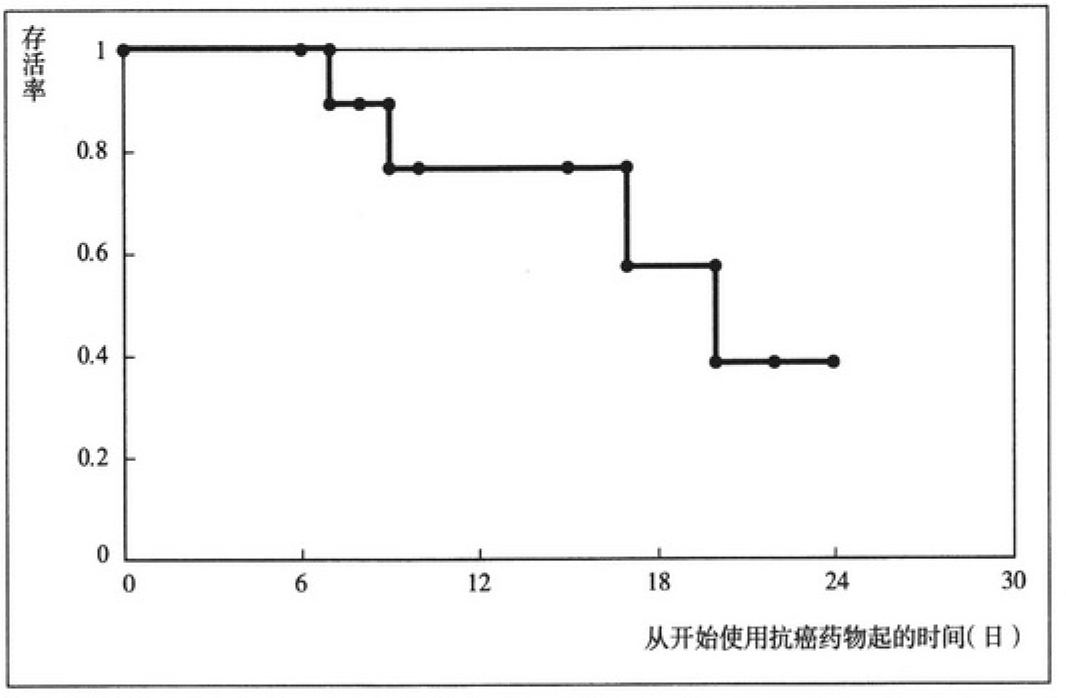
- 在观察期间，死于交通事故等与肺癌无关的原因的患者
- 在观察期间转院的患者
- 在分析者所规定的观察时间内没有死亡的患者

■ 具体实例

下表是对肺癌晚期患者从开始使用抗癌药物起的观察记录。

	从开始使用抗癌药物起的时间（日）	观察结果	
A	17	1	← 死于肺癌
B	10	0	← 至观察结束依然存活
C	15	0	← 死于与肺癌无关的原因
D	20	1	← 死于肺癌
E	7	1	← 死于肺癌
F	6	0	← 转院
G	9	1	← 死于肺癌
H	22	0	← 至观察结束依然存活
I	8	0	← 死于与肺癌无关的原因
J	24	0	← 至观察结束依然存活

使用 Kaplan-Meier 法，可以得到下图所示结果。



Point!

正如此处所举的具体事例那样，利用 Kaplan-Meier 法可以估计某一群体的存活率，虽然说不是一件坏事，但多少有些遗憾。我们可以使用 Kaplan-Meier 法对“服用药剂 X 的患者群”、“服用药剂 Y 的患者群”以及“服用任何药剂的患者群”的存活率进行估计，但是如果要判断它们之间是否存在差异的话，就要使用 Logrank 检验（※ 本书中不做讲解）进行讨论了，这一点需要读者了解。

◆ 参考文献 ◆

■ 第 1 章 / 第 2 章

・内田治『すぐわかる EXCEL によるアンケートの調査・集計・解析』(東京図書) 1997

・大谷信介 / 木下栄二 / 後藤範章 / 小松洋 / 永野武『社会調査へのアプローチ(第 2 版)－論理と方法－』(ミネルヴァ書房) 2005

・大谷信介編『これでいいのか市民意識調査　－大阪府 44 都道府県の実態が語る課題と展望－』(ミネルヴァ書房) 2002

・鎌原雅彦 / 宮下一博 / 大野木裕明 / 中澤潤編『心理学マニュアル　質問紙法』(北大路書房) 1998

・鈴木武 / 山田作太郎『数理統計学　－基礎から学ぶデータ解析－』(内田老鶴圃) 1996

・竹内光悦 / 元治恵子 / 山口和範『図解入門ビジネス アンケート調査とデータ解析の仕組みがよ～くわかる本』(秀和システム) 2005

・谷岡一郎『「社会調査」のウソ　－リサーチ・リテラシーのすすめ』(文藝春秋) 2000

・土屋隆裕『社会教育調査ハンドブック』(文憲堂) 2005

・豊田秀樹『調査法講義』(朝倉書店) 1998

・好井裕明『「あたりまえ」を疑う社会学　質的調査のセンス』(光文社) 2006

■ 第 3 章 / 第 4 章 / 第 5 章

・足立浩平『多変量データ解析法　－心理・教育・社会系のための入門－』(ナカニシヤ出版) 2006

・内田治 / 菅民郎 / 高橋信『文系にもよくわかる多変量解析』(東京図書) 2005

・小塩真司『研究事例で学ぶ SPSS と Amos による心理・調査データ解析』(東京図書) 2005

・小野寺孝義 / 山本嘉一郎編『SPSS 事典　－ BASE 編－』(ナカニシヤ出版) 2004

・菅民郎『多変量解析の実践(上)』(現代数学社) 1993

・芝祐順『因子分析法(第 2 版)』(東京大学出版会) 1979

・高橋信『マンガでわかる統計学【回帰分析編】』(オーム社) 2005

・高橋信『マンガでわかる統計学』(オーム社) 2004

・永田靖 / 棟近雅彦『多変量解析法入門』(サイエンス社) 2001

・南風原朝和『心理統計学の基礎　－統合的理解のために』(有斐閣) 2002

・松尾太加志 / 中村知靖『誰も教えてくれなかった因子分析』(北大路書房) 2002

・柳井晴夫 / 繁桝算男 / 前川眞一 / 市川雅教『因子分析　－その理論と方法－』(朝倉書店) 1990

・柳井晴夫 / 高木廣文 / 市川雅教 / 服部芳明 / 佐藤俊哉 / 丸井英二『多変量解析ハンドブック』(現代数学社) 1986

・山口和範 / 高橋淳一 / 竹内光悦『図解入門　よくわかる　多変量解析の基本と仕組み』(秀和システム) 2004


下次要去哪儿呢？
送给山本吧！

（N-0354.0103）
责任编辑：唐　璐　赵丽艳
责任制作：董立颖　魏　谨
封面制作：【铭轩堂设计：13671110894】

用漫画这种形式讲数学、物理和统计学，十分有利于在广大青少年中普及科学知识。

周恩来、邓颖超秘书，周恩来邓颖超纪念馆顾问
中日友好协会理事，《数理天地》顾问，全国政协原副秘书长

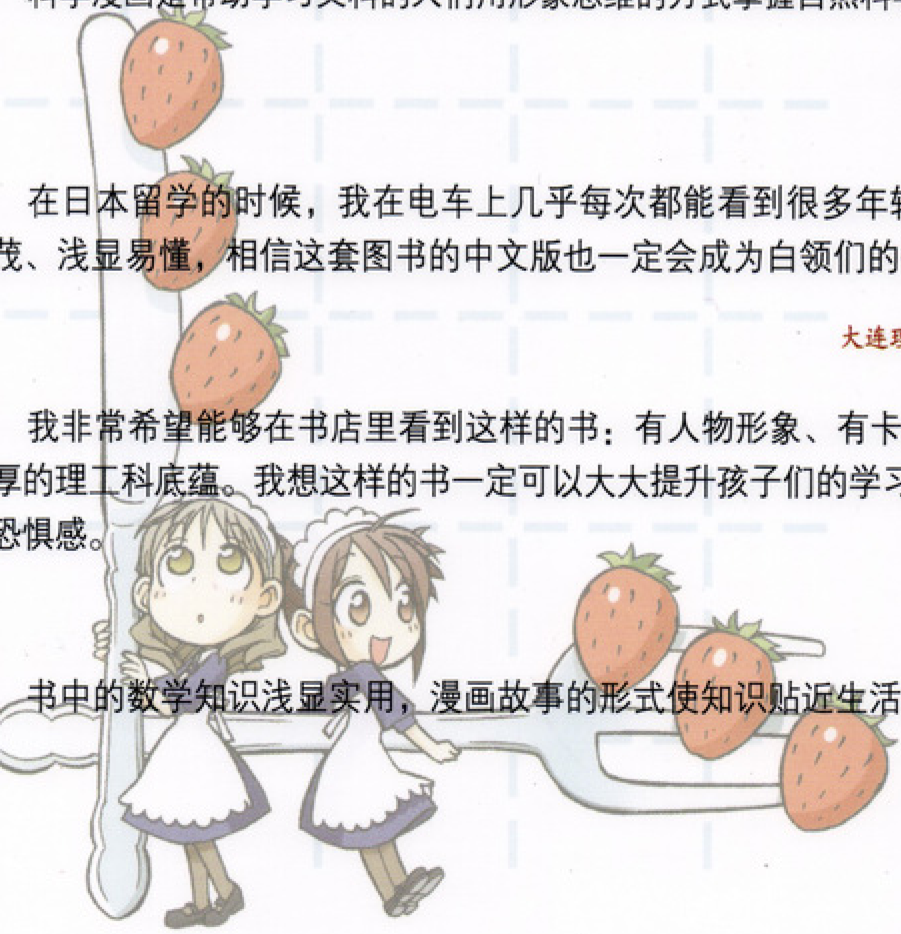
用漫画和说故事的形式讲数学，使面貌冷峻的数学变得亲切、生动、有趣，使学习数学变得容易，这对于提高全民的数学水平无疑是功德无量的事。

《数理天地》杂志社　社长 总编
"希望杯"全国数学邀请赛组委会 命题委员会主任
周国镇

用漫画的形式，讲解日常生活中的数学、物理知识，更能让大家感受到数学殿堂的奥妙与乐趣。

《光明日报》 原副总编辑
中华炎黄文化研究会 常务副会长
鲁谆

科学漫画是帮助学习文科的人们用形象思维的方式掌握自然科学的金钥匙。



中国人民大学外语学院日语专业 主任
大学日语教学研究会 会长
成同社

在日本留学的时候，我在电车上几乎每次都能看到很多年轻的白领看这套图书，经济实惠、图文并茂、浅显易懂，相信这套图书的中文版也一定会成为白领们的手中爱物。

大连理工大学 能源与动力学院 博士 副教授

我非常希望能够在书店里看到这样的书：有人物形象、有卡通图、有故事情节，当然最重要的还有深厚的理工科底蕴。我想这样的书一定可以大大提升孩子们的学习兴趣，降低他们对于高深的理工科知识的恐惧感。

北京启明星培训学校 校长

书中的数学知识浅显实用，漫画故事的形式使知识贴近生活，概念更容易理解。

北京大学 数学科学学院 博士
张磊

上架建议：科普/漫画

ISBN 978-7-03-024963-0

科学出版社 东方科龙
http://www.okbook.com.cn
zhaoliyan@mail.sciencep.com

定价：29.80元