

本样章直接来自笔者原稿, 因此在个别文字及排版上会与正式出版物有所差异。

关于本书的更多内容及下载请访问: www.StatStar.com

@文彤老师

第 1 章 方差分析模型

通过对 SPSS 统计分析基础教程的学习, 读者们已经全面掌握了 IBM SPSS Statistics 软件的基本操作、图表绘制、统计描述技术和单因素统计分析方法。但是, 真实的世界复杂多变, 各种变量间的联系错综复杂, 仅仅依靠统计描述或者简单的统计推断方法往往无法满足分析需求, 必须依赖更强大的统计模型来解决问题。本书将进一步向读者们详细介绍各种高级统计分析模型, 而本章将要介绍的方差分析模型就是各种多变量分析模型中最为基础和常用的一种。



一见模型二字, 很多读者就会顿觉血压升高, 心跳加速, 大有战战栗栗, 汗不敢出之势。实际上, 所谓模型, 无非是对复杂现实世界的一种简化描述而已, 一个出色的模型必然比现实世界更为简明易懂, 就像那个著名的公式: 无极=无聊的平方, 大家一看就明白这个模型是什么意思。而模型实际上也不一定是枯燥的公式表达, 比如说当其以一种赏心悦目的姿态出现在 T 台上的时候, 我们就称其为“模特”。怎么样, 这样解释之后, 各位读者对模型的感觉变好一点了吗?

1.1 模型简介

在实际项目中, 研究者在分析数据时往往需要同时考察多个因素对因变量的影响情况, 比如要研究性别对身高的影响, 显然就应当控制年龄、遗传、营养状况等因素的作用。这时单因素分析方法是无能为力的, 而以方差分析模型为代表的多因素分析方法可以在控制其他因素影响的同时研究两者之间的关系, 分析的效率更高, 适用范围更广。

此外, 许多时候各自变量之间还会存在交互作用, 如研究催化剂对化学反应的催化能力, 如果该催化剂只在某个温度范围内效果最佳, 则单独研究该催化剂的催化作用并无实际意义, 此时这种交互作用也成为了研究的重点, 即必须要研究在什么温度条件下该催化剂的催化能力最佳。对交互作用的分析也是方差分析模型的特长。

1.1.1 模型入门

1. 单因素方差分析模型的结构

为了能够让读者能够对方差分析模型有更为清晰的了解, 让我们从一个最简单的统计模型开始, 假设现在希望描述某个人群的月收入状况, 那么根据统计学知识, 均数能够表示集中趋势, 标准差能够表示离散趋势, 则任何一位受访者 i 的月收入 Y_i 可以被表达为如下形式:

$$Y_i = \mu + \varepsilon_i$$

其中 Y_i 代表第 i 位受访者的具体收入。显然, 此时总体均数 μ 是 Y_i 的最佳估计值, 而 ε_i 则表示因各种原因导致的第 i 位受访者实际收入和均数之差, 或者说反映了抽样中的随机误差。为便于推导, 模型一般假定 ε_i 服从均数为 0, 标准差为某个定值的正态分布 $N(0, \sigma^2)$ 。

下面开始扩展模型框架, 假设现在希望比较三种职业: 医生、律师和软件工程师的月收入有无差异, 那么最简单的做法就是在这三类人群中都进行随机抽样, 各自得到一组受访者, 收集它们的月收入数据, 然后进行检验。则在此问题中, 每一位受访者月收入的平均估计值 Y_{ij} 可以被表达为如下形式:

$$Y_{ij} = \mu_i + \varepsilon_{ij}$$

其中 Y_{ij} 代表第 i 个职业组中第 j 位受访者的具体收入。显然, 在此表达式中 μ_i 表示某一个职业组的平均收入, i 的取值范围在 1~3, 分别代表三种职业之一; 而 ε_{ij} 表示第 i 组的第 j 位受访者收入相对于本组均数的随机误差。

下面来看模型中对 ε_{ij} 的设定, 模型中假设各组的 ε_{ij} 服从同一个正态分布, 即无论 i 取值是多少, ε_{ij} 均服从同一个均数为 0, 标准差为某个定值的正态分布 $N(0, \sigma^2)$ 。这样一来, 如果三种职业收入无差异, 则它就应当等于总体均数 (平均水平) 再加上一个随机误差项, 实际上就变成了同一个变量的分布 $N(\mu, \sigma^2)$ 。为了能够对收入水平进行预测, 人们又规定 $E(Y) = \mu_i$, 即第 i 组个体的收入估计值等于该组的平均水平, 结合模型结构, 这应当不难理解。实际上, 如果对应样本数据的话, 该预测值就是各组的样本均数。

为了统计推断的需要, 以上模型往往被改写成如下形式:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

其中 μ 表示不考虑职业时收入的总平均水平; α_i 表示职业为 i 类时的附加效应, 即在 i 职业时平均收入水平的改变情况。例如 $\alpha_i = 1000$, 表明当职业为 i 类时, 平均收入要比总的平均水平高 1000 元。如果职业 1 和职业 3 的平均收入不相等, 则应当有 $\alpha_1 \neq \alpha_3$ 。反之, 如果三种职业的平均收入无差异, 则因为各类均不存在附加效应, 应当有 $\alpha_1 = \alpha_2 = \alpha_3 = 0$ 。因此, 如果要检验职业种类是否对收入有无影响, 就是检验如下假设:

H_0 : 对任意的 i 取值, 都有 $\alpha_i = 0$, H_1 : 至少有一个 $\alpha_i \neq 0$

但是, μ 、 α_i 等显然应当是一个相对的大小, 例如职业 1 比职业 3 的平均收入高 1000 元, 则当 α_3 为 500 时, α_1 就应当是 1500。 α_3 为 100 时, α_1 就应当是 1100, 总之加上 1000 即可。为了能够在实际问题中得到对 μ 、 α_i 具体的估计值, 模型拟合中又会对它们有一些附加的设定, 这被称为模型拟合时的约束条件, 详后。

在基础教程中, 我们已经学习了方差分析的基本思想是变异分解, 例如在单因素方差分析中总变异被分解为如下两部分: 总变异 = 处理因素导致的变异 + 随机变异。现在对照上述模型表达式, 大家就会发现实际上 α_i 就对应了处理因素导致的变异, 而 ε_{ij} 就对应了相应的随机变异。在多因素方差分析模型中, 这一原理没有任何的变化, 只是模型中考虑的因素更多而已。

2. 两因素方差分析模型的基本结构

下面开始对单因素模型进行扩展。同样是上面的问题, 有研究人员提出: 性别应当也对收入水平有影响, 也许正是因为 1 组中男性比例要高于 2 组, 才导致 1 组的收入均数高于后者, 为什么不考虑控制性别的作用? 如果要同时考虑性别和职业对收入的影响, 则模型扩展为:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \varepsilon_{ijk}$$

该模型对应了如下变异分解方式:

总变异=职业导致的变异+性别导致的变异+随机变异

其中 α_i 、 β_j 分别表示职业为 i 类、性别为 j 类时的附加效应, ε_{ijk} 仍为服从某个正态分布的随机误差变量。此时如果要说明职业种类对收入有无影响, 就是检验如下假设:

$H_0: \alpha_i=0$, H_1 : 至少有一个 $\alpha_i \neq 0$

此时性别的影响因被包含在了 β_j 中, 从而不会干扰对职业的检验。

如果要说明性别因素有无影响, 就是检验如下假设:

$H_0: \beta_i=0$, H_1 : 至少有一个 $\beta_i \neq 0$

同理, 此时职业的影响因被包含在 α_i 中, 从而不会影响到对性别差异的检验。在本模型中, 模型无显著性应当是指上面两个 H_0 同时成立 (均不能被拒绝), 而不是只有其中一个成立。

3. 包括交互项的两因素方差分析模型

但是, 以上模型并非两因素模型的完整形式, 继续考虑如下的情形: 也许在某些职业中, 男、女性的收入没有差异, 而在如足球运动员等职业中, 不同性别的收入是有差别的, 即 A 因素是否有作用需要看 B 因素的具体取值而定, 这种情况在统计中被称为两个因素存在交互作用, 此时需要在模型中加入交互项, 具体如下:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$$

其中 α_i 、 β_j 分别表示 A 因素 i 水平和 B 因素 j 水平的附加效应。 γ_{ij} 则代表 A 在 i 水平, B 在 j 水平时两者的交互效应, 在有的模型表达式中也可能被写为 $(\alpha\beta)_{ij}$, 含义相同。

当然, 聪明的读者朋友们还可以想到更多的影响因素, 如学历等级、工作年限, 甚至于血型、民族等等, 那么, 只要依次将相应的效应项 (及其交互作用项) 加入模型表达式即可, 这里不再详述。

4. 模型中效应的检验

在模型表达式列出后, 下面要关心的是: 如何实现对每个因素作用的检验? 简单的说, 根据上面的变异分解式, 可以将总的样本离均差平方和分解成如上这些部分, 随后各个离均差平方和除以自由度可得到均方, 进而将各效应的均方和误差均方相比较, 就得到了 F 统计量。写成公式如下:

$$SS_{\text{总}} = SS_{\text{因素 1}} + SS_{\text{因素 2}} + \dots + SS_{\text{误差}}$$

$$MS_{\text{因素 1}} = SS_{\text{因素 1}} / DF_{\text{因素 1}}, \quad MS_{\text{误差}} = SS_{\text{误差}} / DF_{\text{误差}}$$

$$F_{\text{因素 1}} = \text{MS}_{\text{因素 1}} / \text{MS}_{\text{误差}}$$

借助 F 分布, 计算在 H_0 成立的情况下得到当前这样大 (以及更大) 的 F 值的概率是多少, 从而做出推断, 即可了解该因素对结果变量的影响是否的确存在。可能的朋友对这一部分计算的原理不太理解, 这没有关系, 读者只要记住方差分析的原理是变异分解, 而相应的模型表达式完全能够和变异的具体分解相对应即可, 其余的运算只是为了求得 P 值以做出统计结论。

1.1.2 常用术语

在了解了方差分析模型的基本结构后, 现在来学习一下方差分析中的常用术语。

(1) 因素 (Factor) 与水平 (Level): 因素也被称为因子, 指可能对因变量有影响的分类变量, 而分类变量的不同取值等级 (类别) 就被称为水平。显然, 一个进入分析的因素会有不止一个水平, 例如性别有男、女两个水平, 而分析目的就是考察或比较各个水平对因变量的影响是否相同。在方差分析中, 因素的取值范围不能无限, 只能有若干个水平, 但需要注意的是有时候水平是人为划分出来的, 比如身高被分为高、中、低三个水平。

2. 单元 (Cell): 也被称为水平组合, 指各因素各个水平的组合, 例如在研究性别 (二水平)、血型 (四水平) 对成年人身高的影响时, 最多可以有 $2 \times 4 = 8$ 个单元。注意在一些特殊的试验设计中, 可能有的单元在样本中并不会出现, 如拉丁方设计, 详见第 2 章。

3. 元素 (Element): 指用于测量因变量值的最小单位, 比如上文所提到的收入的例子中, 元素就是每一位受访者。而在配伍设计等重复测量问题中, 元素可能是受试者每一次具体的测量。根据具体的实验设计, 一个单元格内可以有一个或多个元素, 甚至也可以没有元素。

4. 均衡 (Balance): 如果在一个实验设计中任一因素各水平在所有单元格中出现的次数相同, 且每个单元格内的元素数均相同, 则该试验是均衡的, 否则, 就被称为不均衡。不均衡的实验设计在分析时较为复杂, 需要对方差分析模型作特别设置才能得到正确的结果, 详后。

5. 协变量 (Covariates): 指对因变量可能有影响, 需要在分析时对其作用加以控制的连续性变量, 实际上, 可以简单的把因素和协变量分别理解为分类自变量和连续性自变量。当模型中存在协变量时, 一般是通过找出它与因变量的回归关系来控制其影响, 详见第 2 章。

6. 交互作用 (Interaction): 如果一个因素的效应大小在另一个因素不同水平下明显不同, 则称为两因素间存在交互作用。当存在交互作用时, 单纯研究某个因素的作用没有意义, 必须分另一个因素的不同水平研究该因素的作用大小。



如果所有单元格内都至多只有一个元素, 则交互作用无法进行分析, 只能不予考虑, 最典型的例子就是配伍设计的方差分析。

7. 固定因素 (Fixed Factor) 与随机因素 (Random Factor): 两者都是因素的不同种类, 固定因素指的是该因素在样本中所有可能的水平都出现了。换言之, 该因素的所有可能水平仅此几种, 针对该因素而言, 从样本的分析结果中就可以得知所有水平的状况, 无需进行外推。比如要研究三种促销手段的效果有无差别, 所有样本只会是三种促销方式之一, 不存在第四种促销手段的问题, 则此时该因素就被认为是固定因素。

和固定因素相对应的是随机因素, 他指的是该因素所有可能的取值在样本中没有都出现, 或不可能都出现。换言之, 目前在样本中的这些水平是从总体中随机抽样而来, 如果重复本研究, 则可能得到的因素水平会和现在完全不同! 这时, 研究者显然希望得到的是一个能够“泛化”, 即对所有可能出现的水平均适用的结果。例如研究广告类型和投放的城市对产品销量是否有影响, 在设计中随机抽取了 20 个城市进行研究, 显然, 研究者希望分析结果能够外推到抽样所对应的全国城市总体, 此时就涉及到将结果外推到抽样未包括的城市中的问题, 在这种情况下, 城市就应当是一个随机因素。



一般来说固定因素和随机因素在分析时应分别指定, 如果将随机因素按固定因素来分析, 则可能得出错误的分析结果。但是, 在许多时候, 判断一个因素究竟是固定因素还是随机因素并不是件容易的事情。在这里需要提醒各位读者的是: 区别这两者的并非是该因素本身的特性, 而是分析目的, 假如将其看成是固定因素, 则结论就不应当外推到未出现的其他水平中去; 否则, 就应当考虑按照随机因素来分析。

1.1.3 适用条件

1. 理论上的适用条件

在上文对模型结构的介绍中大家就可以看到, 作为一种统计模型, 方差分析也有自己的适用条件, 比如各组的随机误差项被设定为服从一个相同的正态分布, 又如各组的效应是可加的。具体而言, 方差分析模型的适用条件有以下几点:

(1) 独立性: 只有样本中的各元素相互独立, 来自真正的随机抽样, 才能保证变异能够按照模型表达式那样具有可加性(可分解性)。

(2) 正态性: 由于各组的随机误差项 ε 被设定为服从正态分布, 因此模型要求各单元格的残差必须服从正态分布。

(3) 方差齐: 同样是 ε 惹的祸, 由于在模型中无论何种组合, ε 都被假定服从相同的正态分布, 因此模型要求各单元格都满足方差齐(变异程度相同)的要求。

2. 实际操作中对适用条件的把握

显然, 以上要求还是比较严格的, 那么在实际操作时如何掌握? 首先在以上条件中, 对独立性的要求是最严的, 但除了重复测量等特殊情况下, 该条件一般都可以满足。下面来看对正态性和方差齐性在不同情况下的考虑:

(1) 单因素方差分析: 因模型中只有一个因素, 设计较为简单, 样本有充足的信息量对正态性和方差齐性进行考察, 这已经成为了标准分析步骤。但是许多人误将正态性理解为因变量应当正态分布, 显然这种想法和实际的要求不是一回事, 真正应当考察的是模型残差是否服从正态分布。不过从经验上讲, 由于模型有一定的稳健性, 只要因变量分布不是明显偏态, 分析结果一般都是较稳定的。



关于方差齐性, 根据 Box 的研究结果, 在单因素方差分析中, 如果各组的例数相同(即均衡), 或总体呈正态分布, 则模型对方差不齐有一定的耐受性, 只要最大/最小

方差之比小于 3, 分析结果都是稳定的。

(2) 单元格内无重复数据的方差分析: 以配伍设计的方差分析最为典型, 此时在建模前不需要考虑正态性和方差齐性问题, 原因在于正态性和方差齐性的考察是以单元格为基本单位, 此时每个格子中只有一个元素, 当然没法分析! 除配伍设计的方差分析外, 交叉设计、正交设计等也可以出现无重复数据的情况。但必须指出, 这里只是因条件不足, 无法考察适用条件, 而不是说可以完全忽视这两个问题。如果根据专业知识认为可能在不同单元格内正态性/方差齐性有问题, 则应当避免使用这种无重复数据的设计方案。



从模型的角度讲, 建模完成后可以通过残差分析来考察适用条件, 例如绘制残差分布图, 如果残差呈随机分布, 且均值不随预测值波动, 则可知模型符合适用条件。

(3) 有重复数据的多因素方差分析: 由于正态性/方差齐性的考察是以单元格为基本单位, 此时单元格数目往往很多, 平均每个单元格内的样本粒数实际上比较少。例如样本量为 500, 供分析四个因素, 每个因素 3 个水平, 则共有 $3^4=81$ 个单元格, 平均一个格子里只有 5 例左右的样本, 此时实际上很难检验出差别; 另一方面, 这时也可能因为极个别单元格方差不齐而导致检验不能通过。根据实际经验, 实际上在多因素方差分析中, 极端值的影响远大于方差齐性等问题的影响, 因此在分析中可以直接考察因变量的分布情况, 如果数据分布不是明显偏态, 不存在极端值, 则一般而言方差齐性和正态齐性不会有太大问题, 而且也可以基本保证单元格内无极端值。因此, 在多因素方差分析中, 方差齐性往往只限于理论探讨。但对于较重要的研究, 则建模后的残差分析是非常重要的, 这方面的知识请读者参考第 6 章的内容。

1.2 案例: 胶合板磨损深度的比较

例 1.1 现希望比较四种胶合板的耐磨性, 分别从这四个品牌的胶合板中抽取了 5 个样品, 在相同的转速下磨损相同时间, 测量其被磨损的深度 (mm), 数据见文件 veneer.sav。

表 1.1 胶合板磨损数据

品牌名称	磨损深度/mm				
A 品牌	2.3	2.32	2.4	2.45	2.58
B 品牌	2.35	2.3	2.42	2.6	2.35
C 品牌	2.2	2.0	1.9	2.1	2.03
D 品牌	2.54	2.61	2.6	2.57	2.54

1.2.1 操作说明

本例中希望比较不同种类胶合板的磨损深度, 换用统计学的思路来考虑, 即希望分析品牌这个分类因素是否对磨损深度这个连续性因变量有影响, 这显然是方差分析模型的适用范围。由于只有一个影响因素, 因此需要建立的是单因素方差分析模型, 根据前面学到的知识, 相应的模型结构如下:

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

其中 μ 表示不考虑具体的品牌时胶合板的平均磨损深度, 而 α_i 代表 i 品牌和总平均水平相比其平均磨损深度的差异, 此处的分析目的就是要检验是否有 α_i 不等于 0, 操作如下:

1. 选择“分析”→“一般线性模型”→“单变量”菜单项。
2. 将 wear 选入“因变量”框。
3. 将 brand 选入“固定因子”框。
4. 在“选项”子对话框中, 选中“描述统计”和“方差齐性检验”复选框。
5. 单击“确定”按钮。

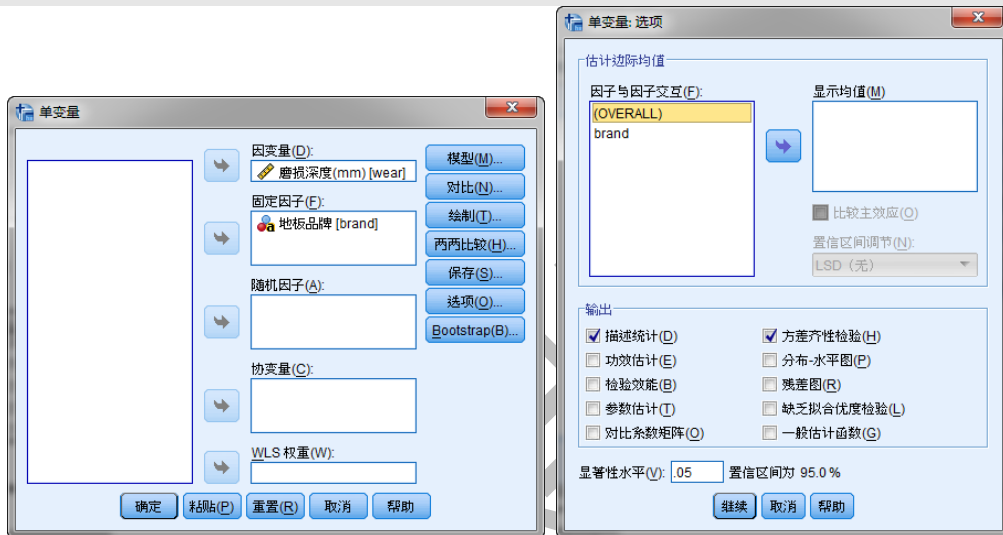


图 1.1 单变量过程的对话框界面



此处过程的名称“单变量”对应的英文为 Univariate, 其实际含义是“只有一个因变量的方差分析模型”。

1.2.2 结果解释

首先输出的是模型中所有因素各水平的取值情况列表, 此处略。表格显示模型中共有一个因素: 地板品牌。它有四个水平, 构成了四个单元格, 每个单元格内共有 5 个元素。

因变量: 磨损深度(mm)

地板品牌	均值	标准 偏差	N
A	2.4100	.11269	5
B	2.4040	.11760	5
C	2.0460	.11216	5
D	2.5720	.03271	5
总计	2.3580	.21771	20

图 1.2 描述性统计量

图 1.2 为对四组样本进行的统计描述, 分别给出了样本均数和标准差, 从标准差可见除 D 组较小外, 其余三组标准差非常接近, 究竟有无方差不齐的问题需要看随后的齐性检验结果。另外请大家注意一下各组的均数大小, 在后面阅读模型参数的估计结果中将会用到。

因变量: 磨损深度(mm)				
F	df1	df2	Sig.	
1.292	3	16	.311	

检验零假设, 即在所有组中因变量的误差方差均相等。
a. 设计 : 截距 + brand

图 1.3 误差方差等同性的 Levene 检验

图 1.3 为方差齐性检验结果, 用来检验所有单元格内的方差是否齐同, 本例中共有 4 个单元格, 因此第一自由度为 4-1=3。可见使用的是 Levene 方差齐性检验, F 统计量为 1.292, 在当前自由度下对应的 P 值为 0.311, 可以认为各单元格所代表总体的方差齐。

图 1.4 给出的是对模型进行方差分析的结果表格, 可见共进行了三个检验:

(1) 第 1 行校正模型 (Corrected Model) 进行的是整个方差分析模型的检验, 其原假设为: 模型中所有的因素均对因变量无影响, 所有的系数 (α 、 β 、 γ ...等) 均等于 0。可见 F 值为 24.550, $P < 0.001$, 因此所用的模型有统计学意义, 其中有系数不等于 0。由于当前模型中只有品牌这一个因素, 因此该结论等价于说品牌间有差异, 有 α_i 不等于 0;

(2) 第 2 行是截距, 其原假设为 $\mu=0$, 即当不考虑品牌时, 胶合板的平均磨损深度均数为 0, 显然检验结果拒绝了该假设, 但由于它在本分析中没有实际意义, 忽略即可;

(3) 从第 3 行开始就应当是对模型中各因素的检验, 由于本模型中只有一个因素, 因此只能见到对因素 brand 的检验, 其原假设为: brand 这一因素对因变量无影响, 其所有的系数 α_i 均等于 0, 可见该检验的 F 值, P 值均和前面的一行对模型的检验完全相同, 结论也完全等价。对此大家应当不难理解。

因变量: 磨损深度(mm)					
源	III 型平方和	df	均方	F	Sig.
校正模型	.740 ^a	3	.247	24.550	.000
截距	111.203	1	111.203	11070.511	.000
brand	.740	3	.247	24.550	.000
误差	.161	16	.010		
总计	112.104	20			
校正的总计	.901	19			

a. R 方 = .822 (调整 R 方 = .788)

图 1.4 主体间效应的检验



如果 SPSS 结果表格中的 P 值显示为 “.000”, 这只是表明 $P < 0.001$, 但由于表格数值格式设定的原因无法显示精确数值, 将其理解为 $P=0.000$ 是严重的误解。要知道假设检验的 P 值是不可能等于 0 的, 至多只会无限接近于 0。

注意图 1.4 的最下方还会给出决定系数与校正的决定系数, 这些指标的含义请参见回归分析一章, 这里不作展开讨论。

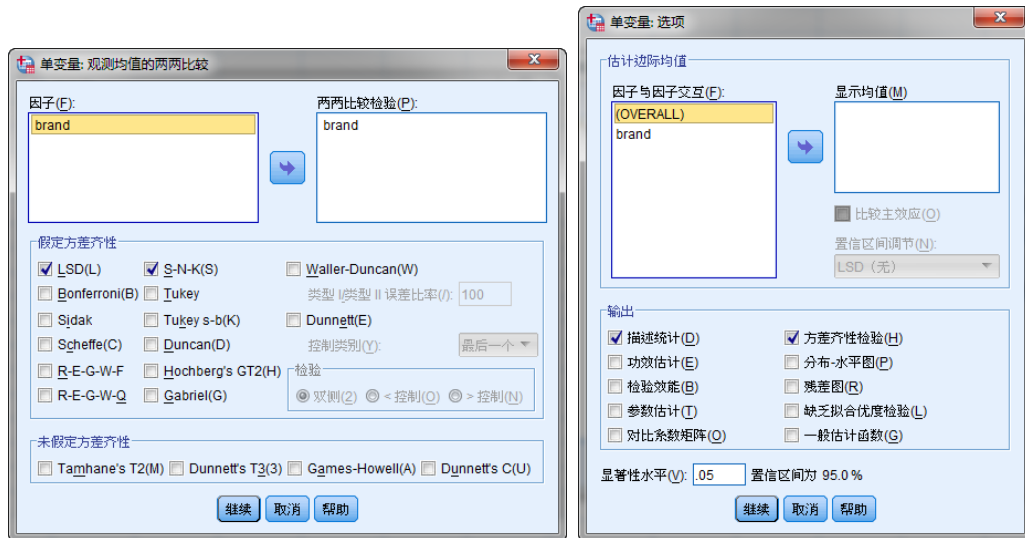


图 1.5 “两两比较”子对话框和“选项”子对话框

1.2.3 模型参数的估计值

经过对上面分析结果的解释, 大家会发现分析结果就是和模型表达式完全对应的, 那么, 能否直接对模型中的参数给出估计值? 结论显然是肯定的, 对话框中相应的操作如下:

1. 单击“选项”按钮。
2. 选中“参数估计”复选框和“对比系数矩阵”复选框。
3. 单击“继续”按钮。

因变量: 磨损深度 (mm)

参数	B	标准 误差	t	Sig.	95% 置信区间	
					下限	上限
截距	2.572	.045	57.383	.000	2.477	2.667
[brand=A]	-.162	.063	-2.556	.021	-.296	-.028
[brand=B]	-.168	.063	-2.650	.017	-.302	-.034
[brand=C]	-.526	.063	-8.298	.000	-.660	-.392
[brand=D]	0 ^a

a. 此参数为冗余参数, 将被设为零。

图 1.6 参数估计

分析结果中会新增 3 个表格, 首先给出的是模型中各参数的估计值, 如图 1.6 所示。所谓截距 (Intercept) 即为总均数 μ , 可见其估计值为 2.572, 即不考虑品牌时平均的磨损深度为 2.572mm。对照前面的样本均数, 我们会发现这实际上就是 D 组的均数! 为什么会这样? 这里需要结合对 α_i 的估计来一起理解。从第二行开始就是对各水平参数的估计, 可见针对四个品牌模型中一共有四个参数, 但是如前所述, 由于这些参数之间存在着数量上的关联, 必须要加上一定的限制条件才能进行估计, 在本例中, 模型默认将编号取值最高的一类, 即 D 类作为参照水平, 这相当于强迫 $\alpha_4=0$, 从而另外三个参数的估计值和检验结果实际上就相当于该水平和品牌 D 相比的结果, 例如 A 类的参数实际上就是两组的均数之差 $2.41-2.572=-0.162$ 。可见 A、B、C 的参数均小于 0 且有统计学意义, 即它们的磨损深度均小于品牌 D。

参数	对比
	L1
截距	1
[brand=A]	.250
[brand=B]	.250
[brand=C]	.250
[brand=D]	.250

此矩阵的缺省显示是相应的 L 矩阵的转置。
基于 III 型平方和。

(a)

参数	对比		
	L2	L3	L4
截距	0	0	0
[brand=A]	1	0	0
[brand=B]	0	1	0
[brand=C]	0	0	1
[brand=D]	-1	-1	-1

此矩阵的缺省显示是相应的 L 矩阵的转置。
基于 III 型平方和。

(b)

图 1.7 截距的设计矩阵表和参数的设计矩阵表

参数估计值表格之后给出的是进行参数估计时使用的设计矩阵, 被称为 L 矩阵, 首先是估计常数项 (截距) 时使用的 L 阵, 如图 1.7(a)所示。可见矩阵中关于 brand 四个水平的参数均为 0.25, 即总样本的平均水平是按照四种品牌等量混合的情况计算的, 但是由于后面又限制了 $\alpha_4=0$, 所以最终会使得常数项的估计值等于 D 组的均数。

图 1.7(b)为计算 $\alpha_1 \sim \alpha_3$ 时的 L 阵, 以 α_1 为例, 它对应的是表中 L2 这一列, 实际上这里的表达式为: $0 * \mu + 1 * \alpha_1 + 0 * \alpha_2 + 0 * \alpha_3 - 1 * \alpha_4 =$, 化简后即等价于 $\alpha_1 = \alpha_4$, 此即前面对 α_1 所作的假设检验的原假设。

除了对比系数矩阵 (Contrast coefficient matrix) 外, 选项子对话框中还有一个一般估计函数 (General estimable function) 复选框, 它输出的内容实质上也是上述模型参数估计时所用的结构, 如图 1.8 所示。对照上面的内容, 其含义不难理解。

参数	对比			
	L1	L2	L3	L4
截距	1	0	0	0
[brand=A]	0	1	0	0
[brand=B]	0	0	1	0
[brand=C]	0	0	0	1
[brand=D]	1	-1	-1	-1

a. 设计 : 截距 + brand

图 1.8 一般估计函数



可能有的朋友会想: 能否对随机误差项 ϵ 也进行估计和检验? 答案是肯定的, 只是由于方差分析模型的重点不在分析 ϵ 上, 因此使用现在的这个分析过程无法自动得出估计值, 除手工计算外, 分析者还可以使用后面将会介绍的混合效应模型直接得到该误差项的估计值及其标准误, 有兴趣的读者可参见第四章的相关内容。

1.2.4 两两比较

由上面的结果可知: 品牌间的磨损深度有差异。但是, 上述检验并未告诉我们究竟是哪些品牌更耐磨, 哪些易磨损, 为此需要进一步做各水平间的两两比较, 操作如下:

1. 单击“两两比较”按钮。
2. 将 brand 选入“两两比较检验”框。
3. 选中“假定方差齐性”框组中的“LSD”复选框和“S-N-K”复选框。

4. 单击“继续”按钮。

因变量: 磨损深度(mm)

(I) 地板品牌		(J) 地板品牌	均值差值 (I-J)	标准 误差	Sig.	95% 置信区间	
						下限	上限
		B	.0060	.06339	.926	-.1284	.1404
A		C	.3640*	.06339	.000	.2296	.4984
		D	-.1620*	.06339	.021	-.2964	-.0276
LSD
		A	.1620*	.06339	.021	.0276	.2964
D		B	.1680*	.06339	.017	.0336	.3024
		C	.5260*	.06339	.000	.3916	.6604

基于观测到的均值。

误差项为均值方 (错误) = .010。

*. 均值差值在 .05 级别上较显著。

图 1.9 多个比较



请大家注意最后一部分将 D 作为参照水平时的分析结果, 大家会发现检验结果和前面参数估计时的检验完全相同, 也就是说, 这里的分析实际上是和前面完全等价的。

分析结果中标题为“在此之后检验”的部分就是两两比较结果, 首先给出的是 LSD 法的输出结果, 如图 1.9 所示。由于 LSD 法实际上是要求将各组均和一个参照水平加以比较, 而操作中并未指定参照水平, 为此 SPSS 假设每一个水平都有可能成为参照, 让其他水平都和该参照进行比较, 因此最终形成了如上所示的四个品牌均作一次参照组的分析结果。但为了便于排版, 我们截去了表格中的部分输出 (表格编辑操作请参见基础教程一书), 只保留了以 A 和 D 作为参照的检验结果。表中会给出两水平间的均数差值、差值的标准误, 95% 可信区间以及检验的 P 值。可见检验结果为 C、D 均与另三个水平有差异, 而 A、B 间无差异。

	地板品牌	N	子集		
			1	2	3
	C	5	2.0460		
	B	5		2.4040	
Student-Newman-Keuls ^{a, b}	A	5		2.4100	
	D	5			2.5720
	Sig.		1.000	.926	1.000

已显示同类子集中的组均值。

基于观测到的均值。

误差项为均值方 (错误) = .010。

a. 使用调和均值样本大小 = 5.000。

b. Alpha = .05。

图 1.10 磨损深度(mm)的两两比较

显然, LSD 法的分析结果并不太容易阅读, 相对而言, 图 1.10 给出的 S-N-K 法两两比较结果则要清楚得多。该方法的输出比较特别, 简单的说, 首先他会将各组在表格的纵向上按照均数大小排序, 如上表中就是按照 CBAD 的顺序进行了排序; 随后, 在表格的横向上各水平被分成了若干个亚组 (Subset), 不同亚组间的 P 值小于 0.05, 而同一亚组内的各组均数则两两

无差别, 比较的 P 值均大于 0.05。从表中可见, 4 种品牌被分在了三个不同的亚组中, 第一亚组仅由 C 组成, 磨损深度最低; 第二亚组由 B、A 组成, 深度居中; 第三亚组由 D 组成, 磨损情况最为严重。最后一行给出的是亚组内部各水平的比较结果, 因第一、四亚组都仅有一个水平, 因此其组内比较的 P 值均等于 1, 第二亚组中 B、A 两水平比较的 P 值则等于 0.926, 显然无统计学差异。

1.2.5 其他常用选项

作为高级教程的第一章, 介绍一些对话框元素的含义将有助于大家今后的学习, 因此这里对本例中未曾用到的一些对话框元素进行一下解释。

1. 主对话框

主对话框中的绝大部分框组名称都在前面的常用术语中出现过, 不需过多说明。比较特殊的是最下方的“WLS 权重”框, 他用于选入加权最小二乘法的权重系数, 以进行加权最小二乘分析, 它实际上和线性回归过程主对话框中 WLS 框的功能相同, 只不过在方差分析中该方法用得比较少, 因此在这里我们不做过多讲解, 请大家参见本书的回归分析相关章节。

2. “保存”子对话框

用于将模型拟合时产生的中间结果或参数保存为新变量供继续分析时用。预测值 (Predicted Values) 复选框组用于存储因变量预测值以及预测值的标准误。除原始的未标准化预测值外, 还有标准化预测值、权重预测值可选, 其中权重的预测值只有在主对话框中选择了加权最小二乘变量 (WLS) 时才可用; 残差 (Residuals) 复选框组则用于存储预测值和实测值之差, 即残差, 同样有多种选择; 诊断 (Diagnostics) 复选框组用于存储一些模型诊断用指标; 系数统计 (Save to New File) 框组则用于选择一个数据文件以保存参数拟合的协方差矩阵。对残差和诊断指标的详细解释请参见线性回归章节的相应内容, 这里不再详述。

3. “选项”子对话框

用于定义模型中的一些附加功能, 上半部的估计边际均值 (Estimated Marginal Means) 框组用于定义输出哪些指标的估计均数、并选择所用的两两比较方法, 详后。下半部的输出 (Display) 复选框组用于设定一些常用的输出指标。除本章将会用到的外, 其余的含义如下:

(1) 功效估计 (Estimates of effects size): 对模型和各因素计算偏 η^2 平方, 它用于表示在控制了其他因素的影响之后, 由该因素所导致的变异占因变量总变异的比率, 总模型的偏 η^2 平方等价于回归分析中的决定系数, 而各因素的偏 η^2 平方等价于相应偏相关系数的平方。偏相关分析的介绍请读者参见本丛书的基础教程, 这里不再重复。

(2) 检验效能 (Observed power): 为模型和所有因素/交互项的检验计算检验效能, 通过该数值我们可以得知试验设计的样本量是否充足, 以及接近检验水准的因素有无必要继续研究, 这是一个常被人忽视的选项。

(3) 分布-水平图 (Spread vs. level plot): 绘出所有单元格的分布——水平图 (该图的详情请参见基础教程中探索过程的相应内容)。

4. “Bootstrap”子对话框

要求在相应的参数估计和假设检验中使用 Bootstrap 方法进行估计, 该方法由 Efron 于 1979 年提出, 是基于大量计算的一种模拟抽样统计推断方法, 可以解决经典统计学较难完成的一些统计推断任务。目前对该方法的使用主要出于两种目的: 1. 判断原参数估计值是否准确; 2. 计算出更准确的可信区间, 判断得出的统计学结论是否正确。对 Bootstrap 方法及相应子对话框的详细介绍请读者参见基础教程第 7 章, 这里不再重复。

1.3 两因素方差分析模型

上一节中大家通过一个单因素的分析实例, 对方差分析模型的基本操作和结果解释进行了了解。但在实际问题中, 更多的是多个影响因素的情形。下面就来对多因素方差分析模型作进一步的学习。

1.3.1 案例: 超市规模、货架位置与销量的关系

例 1.2 相信大家都有过在超级市场购物的经历, 这里就来看一个超市研究的例子: 对超市中销售的某种商品而言, 是否其销售额会受到货架上摆放位置的影响, 除此以外, 超市的规模是否也会有作用? 甚或两者间还会存在交互作用? Berenson 和 Levine (1992) 着手研究了此问题, 他们按照超市的大小 (三水平)、摆放位置 (四水平) 各随机选取了两个点, 记录其同一周内该货物的销量, 数据文件见 `twoway.sav`。

显然, 在本例中所关心的变量是销售量, 可能对其有影响的有两个分类变量: 超市规模和货物摆放位置, 因此本例仍然可以使用方差分析模型来分析, 只是影响因素为两个而已, 相应的模型表达式为: $Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}$ 。其中 α_i 、 β_j 分别表示超市规模 i 水平和货物摆放位置 j 水平的附加效应。而 γ_{ij} 则为两者的交互作用。

表 1.2 超市购物数据

超市规模	货物摆放位置			
	A	B	C	D
小型	45、50	56、63	65、71	48、53
中型	57、65	69、78	73、80	60、57
大型	70、78	75、82	82、89	71、75

1. 初步拟合模型

1. 选择“分析”→“一般线性模型”→“单变量”菜单项。
2. 将 `sale` 选入“因变量”框。
3. 将 `size`、`position` 选入“固定因子”框。
4. 在“选项”子对话框中选中“方差齐性检验”复选框。
5. 单击“确定”按钮。

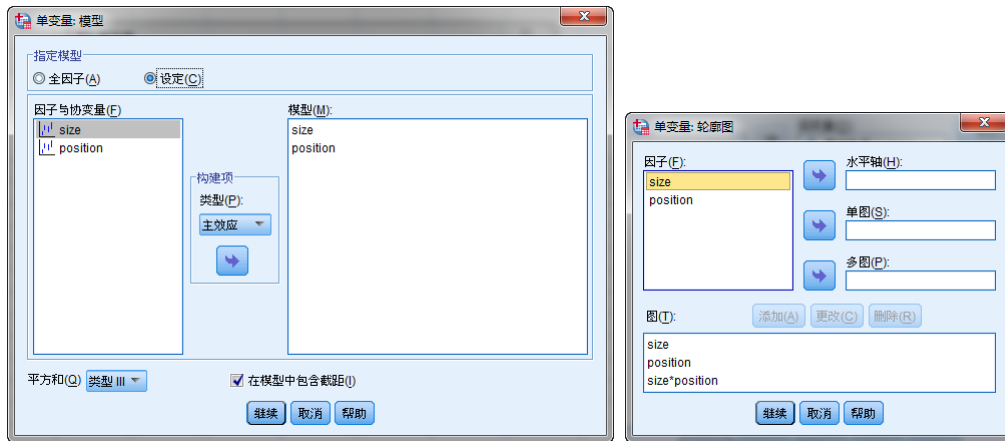


图 1.11 “模型”子对话框和“轮廓图”子对话框

因变量: 周销售量

F	df1	df2	Sig.
.	11	12	.

检验零假设, 即在所有组中因变量的误差方差均相等。

a. 设计 : 截距 + size + position + size * position

图 1.12 误差方差等同性的 Levene 检验

分析结果中首先给出的是方差齐性检验的输出, 如图 1.12 所示, 可见无法进行计算。这是因为两个因素的水平交叉, 一共会形成 12 个单元格, 这里检验的就是这 12 个单元格的方差是否齐。但如果要在考虑交互作用的模型中进行方差齐性检验, 每个单元格内至少要有 3 个元素 (数据点) 才可, 因此这里无法得到分析结果。可见多因素时方差齐性检验往往价值不大, 这也和前面提到的多因素方差分析时一般不考虑方差齐性问题的结论相呼应, 因此, 后面我们都不再涉及此话题。

图 1.13 为总的方差分析表, 有了前一个例子的基础, 该表格应当很容易阅读:

(1) 第 1 行的校正模型是对所用方差分析模型的检验, 其原假设为模型中所有的影响因素均无作用, 即超市规模、摆放位置、两者的交互作用均对销量无影响, 其系数 (全部的 α 、 β 、 γ) 均为 0。该检验的 P 值远小于 0.05, 因此所用的模型有统计学意义, 以上所提到的东西中至少有一个是有差异的, 但具体是哪些则需要阅读后面的分析结果。

(2) 第 2 行是模型中常数项是否等于 0 的检验, 显然它在本次分析中没有实际意义;

(3) 第 3、4 行分别是对超市规模、摆放位置的效应进行的检验, 其原假设分别为: 所有的 α_i 均为 0、所有的 β_j 均为 0。可见两者均有统计学意义, 即 α_i 、 β_j 中均至少有一个不为 0。

(4) 第 5 行是对超市规模和摆放位置的交互作用进行了检验, 可见 P 值无统计学意义。

因变量: 周销售量

源	III 型平方和	df	均方	F	Sig.
校正模型	3019.333 ^a	11	274.485	12.767	.000
截距	108272.667	1	108272.667	5035.938	.000
size	1828.083	2	914.042	42.514	.000
position	1102.333	3	367.444	17.090	.000
size * position	88.917	6	14.819	.689	.663
误差	258.000	12	21.500		
总计	111550.000	24			
校正的总计	3277.333	23			

a. R 方 = .921 (调整 R 方 = .849)

图 1.13 主体间效应的检验

2. 进一步简化模型

由于在本次分析中发现两个因素的交互作用无统计学意义, 为了使得模型更为简洁, 需要在模型中将其去除, 具体在“模型”(Model)子对话框中实现, 操作如下:

1. 单击“模型”按钮。
2. 将“指定模型”框组切换至“设定”单选钮。
3. 在“构建项”框组中, 将“类型”下拉列表更改为选择“主效应”。
4. 将 size 和 position 选入右侧的“模型”框。
5. 单击“继续”按钮。

因变量: 周销售量

源	III 型平方和	df	均方	F	Sig.
校正模型	2930.417 ^a	5	586.083	30.409	.000
截距	108272.667	1	108272.667	5617.799	.000
size	1828.083	2	914.042	47.426	.000
position	1102.333	3	367.444	19.065	.000
误差	346.917	18	19.273		
总计	111550.000	24			
校正的总计	3277.333	23			

a. R 方 = .894 (调整 R 方 = .865)

图 1.14 主体间效应的检验

图 1.14 即为去除了交互项后的方差分析模型结果, 可见检验结论和前相同, 不再详述。

下面对超市规模、摆放位置的具体水平间差异使用 SNK 法进行两两比较, 操作如下:

1. 单击“两两比较”按钮。
2. “两两比较检验”框: 选入 size、position。
3. “假定方差齐性”框组: 选中“S-N-K”复选框。
4. 单击“继续”按钮。

Student-Newman-Keuls				Student-Newman-Keuls					
超市 规模	N	子集			摆放 位置	N	子集		
		1	2	3			1	2	3
小型	8	56.375			D	6	60.667		
中型	8		67.375		A	6	60.833		
大型	8			77.750	B	6		70.500	
Sig.		1.000	1.000	1.000	C	6			76.667
					Sig.		.948	1.000	1.000

图 1.15 超市规模和摆放位置水平间差异的检验

图 1.15 中的两个表格分别是对超市规模和摆放位置的两两比较, 可见超市规模越大, 销售量就越大; 而四种摆放位置也对销量有影响, C 位置的销量最大、其次为 B 位置; A 和 D 位置的销量则最小。同时, 以上差异不受另一个因素水平取值的影响, 两者间无交互作用。

1.3.2 边际均数与轮廓图

前面主要是用表格对结果进行了呈现, 实际上, 在 SPSS 中也可以使用图形对各种水平组合下均数的变化情况加以描述, 这里就来探讨此问题。首先需要了解边际均数这一概念, 它指的是基于现有模型, 当控制了其他因素的作用时, 根据样本情况计算出的用于比较的各水平的均数估计值。如果模型中有协变量, 则会按照协变量取值为均数的情况加以修正, 并进行均数估计。对单因素模型和包含全部交互项的全模型而言, 边际均数就等于样本各单元格的均数; 但是, 对于如上例这种去掉了某些交互项的模型而言, 边际均数就完全是基于模型计算而来的, 是根据当前模型设定对相应效应的估计, 并不能和样本的原始均数相对应。

在了解了边际均数后, 下面来看 SPSS 中提供的图形工具, 主对话框中有一个绘制 (Plot) 按钮, 单击后弹出的对话框用于设定轮廓图。所谓轮廓图就是一种特殊的线图, 图中的每一个点就表示某个因素水平下的边际均数值, 第二个因素的不同水平可以用来区分不同的线, 第三个因素的不同水平可以被绘制为不同的图。轮廓图可用于比较各种水平组合下均数的变化规律, 在寻找可能存在的交互作用时非常有用。如果两个因素间无交互作用, 则第一因素各水平间均数的差异不会随着第二个因素水平取值的变化而变化, 表现为绘制的各条曲线基本平行; 反之, 如果各条曲线明显不平行, 之间有剧烈的交叉, 则提示可能这两个因素的相应水平间存在着交互作用。

在设定好一张轮廓图的绘制方式后, 一定要单击下方的“添加”按钮确认, 否则相应的操作无效。以超市规模和摆放位置为例, 如果希望分别绘制这二者的轮廓图, 则操作如下:

1. 单击“绘制”按钮。
2. 将 size 选入“水平轴”框, 单击“添加”按钮。
3. 将 position 选入“水平轴”框, 单击“添加”按钮。
4. 单击“继续”按钮。

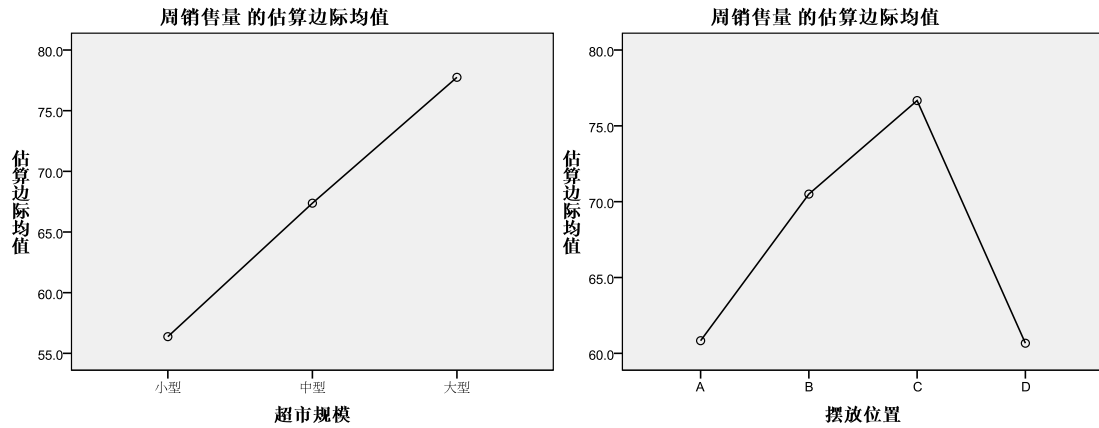


图 1.16 超市规模和摆放位置的轮廓图

相应的图形如图 1.16 所示, 实际上, 在这种情况下各散点的位置和原始样本均数并无差异。

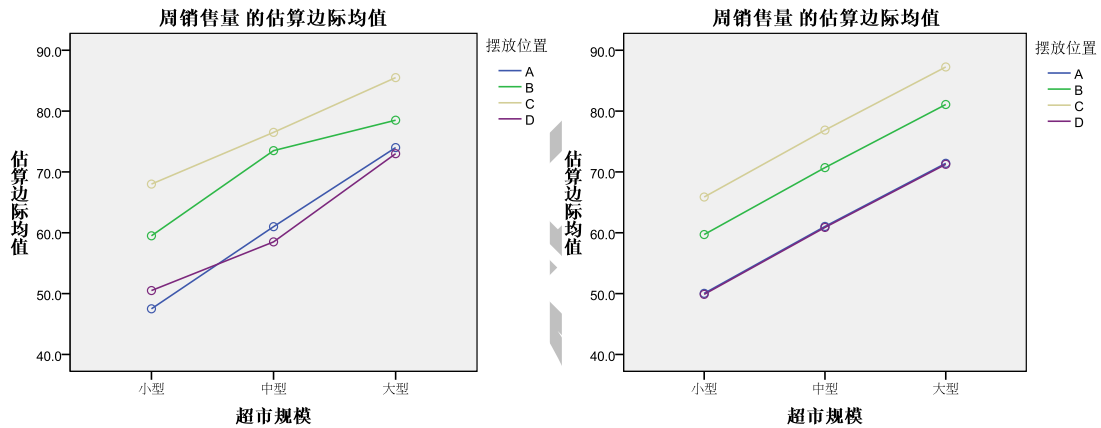


图 1.17 含交互作用和不交互作用时的轮廓图

下面来看复杂一些的情形, 如果希望绘制上述两个变量的联合轮廓图, 则将另一个变量选入“单图”(Separate lines)框中即可。对于不同的模型, 边缘均数的估计值不同, 从而联合轮廓图也不同。图 1.17 为模型中加入和不加入交互作用时的轮廓图, 可见当模型中有交互项时, 各边缘均数实际上和样本单元格均数相同, 且代表摆放位置的四条曲线大致平行, 并未出现明显的交叉, 这提示两变量间的交互作用可能不明显, 与前面检验的结论相一致; 而右侧图形反映的是不加入交互作用的模型轮廓, 由于模型中无交互项, 相当于强行规定无论摆放位置如何, 超市规模各水平的均数差异应当完全相同, 在轮廓图中四条均数线就是完全平行的。

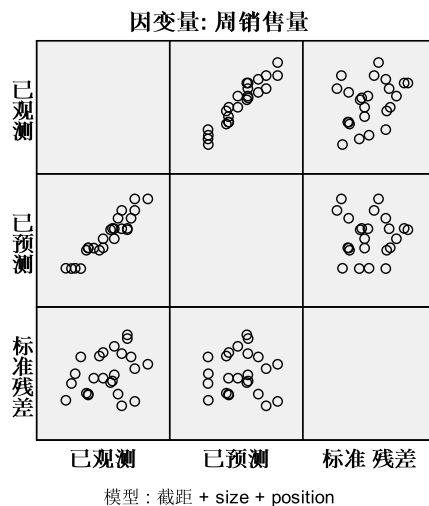


图 1.18 模型残差图

除轮廓图外，还可以使用残差图对模型的拟合效果进行观察，选择“选项”子对话框中的残差图复选框，则结果中会出现如图 1.18 所示的残差图。这幅图实际上是一个散点图矩阵，由因变量实测值、预测值和标准化残差构成，如果模型拟合效果很好，则预测值和实测值应当有明显的相关，呈现出较好的直线趋势，而标准化残差则应当随机的在 0 上下分布，其均值和离散程度均不随预测值的上升而出现变动趋势。由图中的已预测：标准残差单元格可见，残差的分布的确较好，未发现明显违反模型假设的情况，因此模型的拟合效果是令人满意的。

如果希望对残差图作进一步的观察，则可以双击该图形，进入编辑状态，以添加相应的参考线，相应操作请大家参考基础教程中的绘图章节，这里不再详述。

1.3.3 拟合劣度检验

在本例的分析中，由于交互项没有统计学意义，我们直接将其剔除出模型重新进行了数据的拟合。当交互项较少时，这样做并不困难。但是，如果需要分析的因素很多，则交互项也会非常繁杂，对它们一一进行筛选不仅十分麻烦，还很有可能出错，毕竟方差分析模型没有提供变量筛选的方法，而且这样多个检验并行，又涉及到了一类错误是否被正确控制的问题。事实上，对交互项是否应当纳入模型的分析可以被理解为和纳入全部主效应和交互效应的模型（全模型）相比，当前模型对样本信息的解释程度是否充分，两者间的差异是否有统计学意义？如果当前模型和全模型的解释程度无统计学差异，则表明模型已经包含了数据的主要信息，不需要再另行纳入更多交互项了。反之，则意味着还有交互项需要纳入，以改善模型对数据的解释。结合上面有、无交互项时的轮廓图，大家可能对此会有更直观的了解。

选项子对话框中提供了一个缺乏拟合优度（Lack of fit）复选框，就是用于进行当前模型和全模型效果相比的拟合劣度检验（也被称为失拟检验）。如果无效假设被拒绝，则说明现有模型尚不能充分刻画因/自变量间的关系，可能还有交互作用未被发现，或尚有其他因素需要被引入模型。对上面无交互项的模型，拟合劣度检验结果如图 1.19 所示。注意表格中的纯误差（Pure Error）一栏实际上就是全模型的误差项输出，而主效应模型的 Error 其 SS 为 346.917，和全模

型的误差项的 $SS=258$ 相减，恰好就等于失拟 (Lack of Fit) 一栏 SS 的输出 88.917! 事实上自由度的计算也是如此。也就是说，该检验比较的就是两个模型的方差解释量，随后进行的假设检验就是标准的 F 检验，这里不再详述。最终的 P 值为 0.663，正好等于全模型中对交互项的检验 P 值。当只有一个交互项时，这两个检验应当是等价的。但在交互项多于一个的时候，拟合劣度检验的优势就很明显了。

因变量: 周销售量

源	平方和	df	均方	F	Sig.
失拟	88.917	6	14.819	.689	.663
纯误差	258.000	12	21.500		

图 1.19 失拟检验



有趣的是，根据不同的模型起源，这种考察当前模型对数据解释是否充分的检验有时被称为拟合劣度 (Lack of fit)，有时却又被称为拟合优度 (Goodness of fit)，例如在 Logistic 模型中就是如此。这两种不同称呼的方法看似对立，但实质上是完全等价的，就像存活率和死亡率一样。对拟合优度的进一步讲解请参见 logistic 模型一章，这里不再展开讨论。

(1.4 节-1.5 节本样章略去)

1.4 因素各水平间的精细比较

1.4.1 POSTHOC 子句

1.4.2 EMMEANS 子句

1.4.3 LMATRIX 和 KMATRIX 子句

1.4.4 CONTRAST 子句

1.5 方差分析模型进阶

1.5.1 随机因素的方差分析模型

1.5.2 自定义效应检验使用的误差项

1.5.3 四类方差分解方法

思考与练习

1. 本章在模型入门部分使用月收入作为因变量进行了模型架构的描述，但收入往往是服从正态分布的，请思考：

(1) 在建模中如何能够对该数据是否符合模型的适用条件进行判断？

(2) 如果发现数据不符合模型的适用条件, 可以考虑进行怎样的处理?

2. 使用轮廓图、残差图等工具对例 1.1 的数据进行观察, 帮助理解模型的拟合情况。

3. 如果将随机效应错误的指定为固定效应进行分析, 则模型的分析结果是什么含义? 请思索在自己所熟悉的领域有没有这种误用的例子, 这样误用的影响是什么?

参考文献

1. IBM SPSS Advanced Statistics 20. SPSS Inc, an IBM Company. Chicago, Illinois. 2011
2. Neter, Kutner, et al., Applied Linear Statistical Models, Fourth Edition, McGraw-Hill, 1996
3. 张文彤, 钟云飞. IBM SPSS 数据分析与挖掘实战案例精粹. 北京: 清华大学出版社. 2013
4. 张文彤, 邝春伟. SPSS 统计分析基础教程(第 2 版). 北京: 高等教育出版社. 2011
5. 吴明隆. 问卷统计分析实务:SPSS 操作与应用. 重庆: 重庆大学出版社. 2010
6. 胡良平. 现代统计学与 SAS 应用. 北京: 军事医学科学出版社. 1996