[瑞典] T.胡森　[德] T.N.波斯尔斯韦特　总主编

教育大百科全书

教育研究方法

（下）

INTERNATIONAL ENCYCLOPEDIA OF EDUCATION

西南师范大学出版社
国家一级出版社　全国百佳图书出版单位

国家"十五"重点出版工程项目

教育
大百科全书

教育研究方法
(下)

[澳]J.P.基夫斯　主编
石中英　译审

西南师范大学出版社

目　录

· 统计分析和数据管理

· 测验

数据分析中的贝叶斯统计(Bayesian Statistics in the Analysis of Data)

贝叶斯统计既是一门哲学,又是一套统计方法。与传统的统计方法相比,它主要的优势在于:鼓励人们用新的方式去看待数据,它把已有的信息和已知观点清晰地带入数据分析,它提供了一种更为现实的哲学视角,从而能对统计过程的结果做出评估。

贝叶斯统计主要的不足在于:数学计算有时比较复杂,已知的信息和观点有时很难在贝叶斯统计中用一种适当的方法估计出来,对已有信息和观点的考虑可能会导致误用。基于以上原因,当两种标准都具备时,这种方法自身可能更有价值。第一个标准是相当简单的统计过程,比如运用 t 检验或者二项检验;第二个标准是决定具有实践意义,它将最终影响人们做出决定(因此具有较强的动力把真实的观点和愿望分离开来)。然而,甚至一个运用传统统计技术的人,他在哲学上还可能是一个贝叶斯主义者。贝叶斯哲学影响结果的解释。在本词条中,当呈现了贝叶斯统计的基本观点以后,这些问题得到了更加充分的讨论。

1. 贝叶斯定理

"贝叶斯统计"一词起源于一个定理,这个定理最初被贝叶斯所证明,在 1763 年公布于世。后来这一定理被修订;到 20 世纪 90 年代初期,在任何初级统计或者概率论中,它很容易得到证明。让 $P(B|A)$ 代表事件 A 的可能性。假设事件 B 已经发生了,让 B' 代表 B 的完成状态(没有出现)。于是,可以把贝叶斯定理用下面的任何一种形式表示:

$$P(B|A)=\frac{P(A|B)P(B)}{P(A)} \tag{1}$$

$$P(B|A)=\frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|B')P(B')} \tag{2}$$

$$P(B|A)/P(B'|A)=[P(A|B)/P(A|B')][P(B)/P(B')] \tag{3}$$

我们把公式(3)叫作贝叶斯定理的概率形式,因为在条件 A 的基础上,它把 B 和 B' 的概率与数据以及无条件的概率联系起来。公式(3)右边括号内的第一个词通常被叫作"概率"。

贝叶斯定理的数学修正是没有疑问的。贝叶斯定理的应用使贝叶斯统计区别于传统统计。在传统统计中,它是一个很有趣但却几乎得不到应用的公式,而在贝叶斯统计中它是核心。

让我们看一下其中的原因。设想一个具有零假设的普通检验,呈正态分布。零假设为:$H_0:\mu=50$,样本的平均数为 46。假设以这些数据为基础,零假设在 0.05 水平上(单侧检验)被拒绝。这意味着什么呢?它意味着如果零假设为真,那么得到一个平均数等于或者小于 46 的样本的可能性小于 0.05。

对于刚学初级统计的学生来说,这个问题较复杂。它也包括一些严重的不足。第一,它考虑到了获得实际上得不到的数据的可能性(如它考虑到获得样本的平均数小于或者等于 46),这些本应该被认为是不相关的。第二,它只考虑到拒绝零假设的结果,而忽略了错误接受它的可能性(在统计上在假设被用于 β 水平,但实际上人们往往忽视了这些方法,因为没有简单的直接的方法去处理它们)。第三,它声称要检验几乎没有信度的零假设。通常我们认为参数值与假设值并不完全相等(如大于 20 个小数点),在大多数情况下,零假设实际上是一种人为的方法,决定参数是大于、小于假设值,还是与假设值相等。第四,它没有给出零假设为真的概率,它只告诉我们如果零假设为真,那么得到这样的结果的可能性有多大。

最后一个严重的不足是:有时,数据是不可能的,但有一个特定的前提,在不考虑这些数据的情况下接受前提假设是合理的。例如,12 岁儿童的标准成就测验的平均分数为 50,一所有名的地方大学学生样本的平均分数为 46。在这种情况下,研究者倾向于认为数据不可能低于 50 ,$H_0:\mu=50$,但是结果不可能的情况发生了。当然,处理这种问题的一种方法是选择非常小的显著性水平去拒绝。然而,在传统统计中,几乎没有选择特定的显著性水平的指导。这就是我们普遍运用 0.05 这个极其传统的值作为显著性水平的原因。

同样的问题出现在置信区间当中。用同样的方法设置95%的置信区间,这样就包括人口平均值的95%。然而,以前的与实验数据无关的信息可能更具代表性,而目前的实验也正在这不幸的5%之中(假设100～105的95%的置信区间是从哈佛大学学生IQ的平均数中得到的)。这一矛盾存在于已经建立的置信区间中,而建立的置信区间实际上没有信度。

现在考虑一下换一种方法运用贝叶斯理论。在上面的公式中,B代表任何(不一定是零假设)假设,A代表已经得到的样本数据。于是,公式表示:以样本数据为条件的假设成立的概率应该以某种特定的方式决定于无条件概率(如在数据收集之前的概率)。因此我们能直接计算出假设的合理性(如概率),避免了传统统计中的循环方法。在后面我们将通过具体的例子说明,但是通常首先我们应当指出这种方法的主要缺点。

2. 概率论

尽管所有统计学家都接受了贝叶斯定理,但是人们不会把贝叶斯定理运用于上面的例子中。在传统统计中,对数据都设定了概率,但是给假设设定概率被认为是不合理的。其原因来源于概率理论的历史。

概率理论最初应用于赌博。在哲学家得出概率理论之前,人们就把概率过程应用于数学的许多领域中。这会使人们在赌博时少出现问题,因为通常一些概率是显而易见的。然而,在科学领域中,它经常会导致误用,可能最著名的误用者要属拉普拉斯(Laplace)。他把贝叶斯定理(以及一些现在被广泛接受的假设)应用于已知的事实,如五千多年来,太阳每天都要升起,声称明天太阳升起的概率高于1 826 213:1。

反对这些误用的理论指出被赋值的概率只适用于被广泛定义为"实验"的结果。人们通常把实验定义为一种行为或过程,它能够得出单一的、明确定义的结果。因此,就像投掷普通的硬币一样,12岁儿童样本的成就测验也是一个实验。

然而,在对这些结果的概率赋值之前,实验必须有一个特定的属性;实验必须具有可重复性,至少在理论上,在相同的情境下能够重复无限次。实际上我们通常将其解释为:结果通过随机过程得到,并且结果和概率不因重复实验而变化,12岁儿童必须从一些定义明确的12岁儿童群体中随机抽取,硬币必须随机投掷,等等。

这种关于概率的新理论修正了过去对概率的误用。尼曼(Neyman)和皮尔逊(Pearson)采用了这种新理论,他们建立了当今普遍应用的统计理论和测验的基础。但是,它有很大的缺点:极强的限制;因为假设为真或假(如"在完全相同的条件下",它们不能被重复),它以最人为的方式排除了数值的设定(如通过设定概率),而相信假设;它阻止了一些简单的追逐,比如给定赛马的概率(假设赛马在相同的条件下能够重复无限次)。

而且,这种观点的哲学基础是不稳固的。没有一个实验能够真正在完全相同的条件下重复有限次;如果能够的话,将总是得出相同的结果。如果我们每次都用相同的方式投掷硬币,我们期望它们每次都以极其相同的方式着地。概率在条件不完全相同的条件下也存在。但是,在条件不完全相同时,没有客观的方法决定从一个实验到另一个实验概率是否保持恒定。实际上,这个问题可能是无意义的。

20世纪前半叶,人们提出了概率的替代方法。其中最重要的是人本主义者和决定论者对于概率的解释。根据这两派的观点,概率不应对事件赋值,而应对命题赋值——对事实的陈述。概率代表了"可信度",即陈述是正确的。人本主义者和决定论者观点的分歧在于人本主义者认为这种概率完全是个人的——特定的人在特定的前提下可能会有不同的可信度,而决定论者认为会有一个"正确的"概率(通常不知道或者是不可知的)适合于一个特定的命题。大多数运用贝叶斯定理的统计学家都持人本派观点,因为决定论者观点在探求"正确"概率时几乎没有实际的指导。

人本主义观点似乎引起了纷争,使人们对于任何命题都能设定概率。然而,事实不是这样,至少对于明智的人们来说不是这样。早期的人本主义学家拉姆齐(Ramsey 1931)、迪·菲尼蒂(de Finetti 1937)、索瓦热(Savage 1954)等认为个体的概率必

须具有特定的属性才有用。

首先,概率必须具有现实性,因为它们必须通过个体与可能的具体的行为相联系(否则概率将没有意义)。为了简化问题,如果一个特定命题的概率是3:4,那么在赌博中设定命题正确的概率为3:1,这样才是合理的。其次,这些理论家认为,概率,从一致性来说,必须满足概率理论的数学原理,违背原理必将导致在赌博中输钱(de Finetti 1937)。

这些必要条件比我们设想的有更大的限定性。例如,如果给两个明智的人呈现大量的解释清晰的数据,这些数据与特定的前提有关,这两个人应该以几乎相同的概率完成相同的命题,即使从这两个人差异很大的概率开始(Edwards et al. 1963)。相对于频率方法,概率方法在哲学上更具有正确性,它具有一定的优势——既对前提进行概率赋值,也对实验结果进行概率赋值。

然而,要注意前面的章节中用到了"should"一词。人体概率理论是一种标准化的,而不是描述性的理论。它详细说明了人们应当如何理智地去做,而不是他们如何自然地去做。像所有的统计方法一样,贝叶斯统计方法的目的是更接近理想化。

3. 举例

现在我们把这些观点应用于比较简单的具体问题中。假设一项能力测验的成绩呈正态分布,标准差为10,然后比较以下两个假设,H_0:$\mu=50$,H_1:$\mu=55$。可能这项测验中,12岁儿童的典型分数是50,而14岁儿童的典型分数是55。

样本由12岁儿童组成,但是第二个假设认为他们的能力水平实际上能代表14岁儿童。假设出于某种原因,我们认为得到14岁儿童的平均数的概率是12岁儿童的两倍,那么零假设的概率应为1:3,研究假设的概率为2:3。

以25个学生为样本,平均数为53,平均数的标准差是2,两个假设都在0.05水平上被拒绝。然而,如果这两个假设是合理的,那么不可能将二者都拒绝。

在这种情况下,运用连续的正态分布数据,贝叶斯定理可以修正如下:

$$P(B|\bar{X})=kf(\bar{X}|B)P(B)$$

$\bar{X}$代表样本的平均数,$f(\bar{X}|B)$是$\bar{X}$平均数的密度函数,k是常量。在上面的例子中,$\bar{X}$的平均数呈正态分布,标准差是2,假设零假设为真,平均数是5。因此,正态分布公式为:

$$f(\bar{X}|H_0)=\frac{1}{2\sqrt{2\pi}}\exp\left[-\frac{1}{2}\left(\frac{\bar{X}-50}{2}\right)^2\right] \quad (4)$$

同样,$f(X|H_1)$与上式中代表的意义相同,只要用55代替50就可以了。我们通过计算这些数据,能得到下面的值:

$$P(H_0|\bar{X})=k(0.0648)(1/3)=k(0.0216) \quad (5)$$

$$P(H_1|\bar{X})=k(0.1210)(2/3)=k(0.0807) \quad (6)$$

在每一个方程式中,括号内的第一个数都是$f(\bar{X}|H_i)$,第二个数是假设中的先验概率,当两个概率的和为1时,就会很容易地得到k值:

$$k=1/(0.0216+0.0807)=9.775 \quad (7)$$

$$P(H_0|\bar{X})=0.21 \quad (8)$$

$$P(H_1|\bar{X})=0.79 \quad (9)$$

分析完数据之后,H_1的概率由2:3增加到4:5。因此,可以说先验概率(实验前)是2:3,后验概率(实验后)约是4:5。

用一个较大的样本,样本的平均数更接近于50或者55(取决于哪一个假设成立),假设的后验概率会非常接近于先验概率。我们就会有充分的理由支持正确的假设,不确定性被接近确定性所取代。人们运用大量的确定的数据能得到较一致的结果,在较早的时候,人们就得到了这一结论。

当然,在上面简化的例子中,其他一些假设的概率(如平均数为52.5)可能被修正,也可能被忽略了。实际上,我们只要对程序进行较小的修正,就能够同时考虑所有假设的数据,下面我们将简要介绍这一点。

当考虑所有可能的假设时,贝叶斯统计就变得非常有趣、非常有用。在上面的例子中,我们可以把实线上所有的点看成人口平均数的可能的值。通常在这种情况下,我们不再考虑平均数与某一特定的值完全相等的可能性(概率一般为0),而是考虑平均数介于两个值之间的可能性。于是,我们可以用密度分布描述概率——曲线:在曲线下两个数值之间的面积代表了平均数介于两个数值之间的

概率。

重新考虑上面的例子，既然我们希望儿童的表现像14岁孩子那样，设想我们相信（收集数据之前）平均数的值更可能是55。更近一步设想，平均数在50和55之间这一假设成立的概率约为0.34，平均数在55和60之间这一假设成立的概率也约为0.34。并且，假设平均数在45和65之间的确定性为95%。能够用一条正态分布曲线很好地代表先验概率，其平均数为55，标准差为5。

于是，贝叶斯定理可以被重新表示为：

$$f(\mu|\bar{X}) = kf(\bar{X}|\mu)f(\mu) \qquad (10)$$

用密度分布代替了概率。这里$f(\mu)$代表后验概率的密度分布，选择了k，以使由$f(u|\bar{X})$所决定的曲线下的面积为1。

现在计算就显得更困难了，包括两个指数函数，必须处理二者的乘积。然而，实际的计算却非常简单。如果，实验前可以用正态分布代表μ，其平均数是m，标准差是σ，$\bar{X}$的取样分布呈正态（平均数为μ，标准差为$s/\sqrt{N}$），那么，实验后，μ也可以用正态分布表示，其平均数是m^*，标准差为σ^*。

$$m^* = \frac{(s^2/N)m + \sigma^2\bar{X}}{(s^2/N) + \sigma^2} \qquad (11)$$

$$\sigma^* = \sqrt{\frac{(s^2/N)\sigma^2}{(s^2/N) + \sigma^2}} \qquad (12)$$

本例中，分析数据后的后验概率呈正态分布（平均数$m^* = 53.3$，标准差$\sigma^* = 1.9$）。正态分布变得更狭长（如标准差更小），这表明我们更有把握相信这是学生平均数的真值。例如，μ在50和55之间的概率由收集数据前的0.34，增长到收集数据后的0.85。

一个发人深思且非常重要的问题是：即使先验概率的分布不是标准的正态或者先验概率的平均数不是55，但是后验概率的分布却接近正态，其平均数接近53，标准差接近2。因此，先验概率起的作用比较小，而数据所起的作用相当大，它决定了后验概率。当数据的代表性很强而前提假设相对不清晰时，这种情况经常出现。从以上的方程式可以看出，如果样本量很大，那么后验平均数与样本平均数几乎相等，后验标准差与样本平均数的标准误差几乎也相等。因此，再次指出，由于数据的代表性很强，概率的主观性实际上消失了，明智的研究者将较能达成一致。

现在已经算出后验概率的分布，我们可以把它运用于各种计算。如进行点估计和置信区间估计，做出关于μ的测验假设。一般意义上，最好的点估计是后验分布的平均数53.3。设定95%的置信区间，除去样本后验分布上下2.5%（如在$z = 1.96$上），那么μ在49.6和56.9的置信水平为95%。

当检验假设时，贝叶斯统计会做出选择。通常当检验零假设时，其兴趣并不存在于零假设本身。例如，如果检验H_0:$\mu = 50$，那么想得到的结论是μ是否会大于或者小于50。样本平均数表示或大或小，显著性水平表示结论的置信度。在贝叶斯统计中，可以直接设定假设$\mu < 50$。它只是一个正态分布的随机变量，其平均数是53.3，标准差是1.9，值小于50。对于上面的例子，概率约为0.04。要注意是“接近”，而不是“完全是”，可能通过传统的显著性测验（单测检验，在0.05水平上）拒绝H_0:$\mu = 50$。用大样本可以得到$\mu < 50$的后验概率与传统单测检验的p值几乎相等。许多贝叶斯统计方法的支持者声称传统统计的实际价值在于用传统方法所得结果和贝叶斯统计结果的一致性。

有时候，意义实际上在于零假设本身。可能12岁儿童测得的结果真的像典型的12岁儿童（如平均数为50）。在这种情况下，贝叶斯定理的概率形式就非常方便了。然而，有必要知道在零假设为假设为真的条件下，数据的概率。这样，传统的统计方法就不能发挥作用了，因为，实际上，我们必须知道β的水平，而通常我们不知道β的水平或者β的水平是不可知的。

爱德华兹等人（Edwards et al. 1963）详细描述了贝叶斯法。对于上面的例子，零假设不成立的概率约为1.14，其意思是零假设不成立的后验概率等于先验概率乘1.14。如果先前我们认为零假设不成立的概率约为2:1，那么后验概率约为2.28:1。传统测验几乎要在0.05水平上拒绝零假设（单测检验），而数据对支持另一种假设只能起微弱的作用。原因很简单。尽管这些数据在零假设前提下成立的可能性不是很大，它们在另一种假设前提下

成立的可能性也不是很大。因此,对于任何一种假设它们都不能给予强有力的证据。

以上只是贝叶斯定理的两个简单例子。它们也能运用于其他分布中,如二项分布、贝尔松分布几乎都能得到简单的结果。它们还能用于平均数与方差未知的正态分布,以及更复杂的问题中,如对强和协方差的分析、多元回归等等。

4. 问题

最初贝叶斯统计方法之后的哲学思想遭到了统计学家的反对。现在,这些反对意见正逐渐消失。但是,目前有四大困难正在阻止贝叶斯统计方法的广泛应用。

第一大困难是计算——贝叶斯统计方法中的计算比传统统计中的计算要复杂。一个很明显的解决方法就是发展计算机程序。然而,通常使用的统计软件包,如 SAS 或者 SPSS,都不包括贝叶斯程序。这可能是一个"恶性循环"。这些程序不常用,所以商业程序也就不把其包括在内。但是,在商业程序中找不到它,那么人们也就不可能常常用到它。

第二大困难涉及用概率确定先前观点的问题。如果数据代表性很强,这一问题就不存在了——数据的力量大大超过了运用者对先前概率的任何一种明智估计。然而,对于这样的数据,传统统计方法通常能得出恰当的结果。

对于更加复杂的问题,例如方差分析,就变得较困难了。必须首先估计相关参数的先验概率。让我们来举个简单的例子,一项能力测验要估计 10 岁、12 岁、14 岁儿童组的分数。我们不能确定 14 岁儿童组的成绩高于平均成绩,但我们比较确信 14 岁儿童组的成绩高于 12 岁儿童组的成绩。在本例中分别估计每组的先验概率是不够的,必须得到三组平均数的联合分布情况。如果标准差已知,那么就会使问题更复杂化了。

对于这个问题有两种解决方法。第一种,尽力消除主观性,而保留贝叶斯数学方法。对参数预先进行合理设定是一种通用的方法,它消除了各个研究者分别进行预先估计的困难。有时,合理的预先设定来自先前出现的数据。产生结果的过程叫作"经验型贝叶斯过程"。然而,一些人认为这些过程违背了贝叶斯精神,至少是贝叶斯方法,它使贝叶斯方法的许多优势得不到发挥。第二,化复杂问题为简单问题。不对变异进行总体分析,而是进行正交比较,对比较结果进行选择以使一个比较值同其他比较值彼此独立。这种方法在理论上很好,但是选择比较结果在实际中很难。

阻止贝叶斯统计方法应用的第三个、第四个问题有更多的心理原因。

第三个问题是:使用者不愿意放弃"舒适",即由统计分析结果决定他们的观点。如果数据明确显示要接受或者拒绝零假设,研究者就不需要为做困难的决定负责任。当然,这种"舒适"是靠不住的,当你选择了显著性水平,你就接受了责任。但是,许多使用者不愿意放弃幻想。贝叶斯统计方法在商业中应用非常广泛,很可能基于以上原因,因为在商业中这种幻想能带来价值。

第四个问题是简单的惰性。通常人们运用传统的统计方法。因此,学生必须学习传统的方法以使他们能理解现存的知识。于是,他们运用学过的统计,始终保持传统的方法。

5. 其他的应用

也应该讨论一下贝叶斯思想的两种其他的应用。第一种包括经验的贝叶斯方法,它是一种贝叶斯理论的技术应用。有时,特别是在检验数学模型时,估计参数并不简单,所以,有必要设计 adhoc 法。但是,去判断是否这些方法都做出了合理的估计是很困难的。如果我们发现通过运用贝叶斯程序的合理的前提,估计值升高了,那么估计本身是合理的。

第二种是"哲学贝叶斯主义"。从上面我们可以看出,如果用大样本,贝叶斯测验的结果与传统测验结果几乎相同。因此,我们可以用贝叶斯方式即根据后验概率去代替和解释传统测验。如果是小样本,也可以根据贝叶斯思想去解释结果。例如,即使样本的代表意义很强,如果它支持了一个不合理的假设,那么它的代表性也会大打折扣。但是,没有清晰的贝叶斯方法,解释通常不会很清楚,并且如果没有清晰的贝叶斯计算方法,我们也不能

确定解释的恰当性。

H. R. 林德曼(H. R. Lindman) 著
张宏学 译

附录

de Finetti B 1937 La Prévision: Ses lois logiques, ses sources subjectives. *Annales del'Institut Henri Poincaré* 7:1—68[1964 Foresight: Its logical laws, its subjective sources. In: Kyburg H E, Smokler H E (eds.) 1964]

Edwards W, Lindman H, Savage L J 1963 Bayesian statistical inference for psychological research. *Psychol. Rev.* 70(3):193—242

Ramsey F P 1931 *The Foundations of Mathematics and Other Logical Essays.* Kegan Paul, London

Savage L J 1954 *The Foundations of Statistics.* Wiley, New York

其他参考文献

Kyburg H E, Smokler H E (eds.) 1964 *Studies in Subjective Probability.* Wiley, New York

Lindley D V 1965a *Introduction to Probability and Statistics from a Bayesian Viewpoint. Part 1: Probability.* Cambridge University Press, Cambridge

Lindley D V 1965b *Introduction to Probability and Statistics from a Bayesian Viewpoint. Part 2: Inference.* Cambridge University Press, Cambridge

Novick M R, Jackson P H 1974 *Statistical Methods for Educational and Psychological Research.* McGraw-Hill, New York

Phillips L D 1974 *Bayesian Statistics for Social Scientists.* Crowell, New York

Savage L J 1981 *The Writings of Leonard Jimmie Savage: A Memorial Selection.* American Statistical Association/Institute of Mathematical Statistics, Washington, DC

Schmitt S A 1969 *Measuring Uncertainty: An Elementary Introduction to Bayesian Statistics.* Addison-Wesley, Reading, Massachusetts

Winkler R L 1972 *An Introduction to Bayesian Inference and Decision.* Holt, Rinehart and Winston, New York

典型分析(Canonical Analysis)

20 世纪七八十年代,人们逐渐接受和使用多元分析过程去检验潜变量的相关,这些潜变量以观测变量的形式出现。早在 1935 年霍特林(Hotelling)就开始使用典型变量分析,但是当时只能用人工计算,由于这种方法计算的复杂性,致使它没有被广泛应用。在引进电子计算机之后,越来越容易得到支持它使用的程序。然而,人们没有普遍认可分析过程的力量。典型变量分析实际上是大多数参数统计过程中一种常见的分析方法,t 检验、多元方差分析、主成分分析、因素分析、回归分析和判别分析只是它的特例。不管从历史上还是从观念上,多元分析中运用的基本分析过程都来自典型变量分析。因此,对典型变量分析中观念和原理的充分理解为理解其他分析过程奠定了良好的基础,如偏最小平方路径分析(PLS)、线性结构相关分析(LISREl)、多元方差分析(MANCOVA)。本词条主要介绍典型变量分析的分析过程。根据巴特利特所说(Bartlett 1941),"variate"一词指引用到分析中的观测值,而"variable"是为潜在的结构服务的,这些潜在的结构以观测值或者观测元组合的形式存在。

1. 引言

典型分析或典型变量分析(CVA)是一种用于考察两组变量(X、Y)间关系的统计方法。每组变量至少包括一个变量。达林顿等人(Darlington et al. 1973)正视了三种不同类型的问题,我们可以把它们叫作有关变量($n_x + n_y$)的相关矩阵问题。

(a) 两组变量间独立相关的数量和性质问题。

(b) 两组变量的交叠或冗余的程度问题。这意味着一组变量在多大程度上可以预测另一组变量,反之亦然。

(c) 两组组内相关或者协方差矩阵的相似性问题。

CVA 经常适合于解决(a)类问题,有时对(b)类问题也适用,但不适用于解决(c)类问题。

为了更清楚地理解 CVA 的作用,达林顿等人

(1973 P. 439)列出了一表格来表示 CVA 和方差分析技术、多元回归分析以及多元方差分析间的关系。他们还指出通过适当地运用虚拟变量,CVA 也能进行多元辨别分析和简单的列联表分析。因此,当 X 组或 Y 组包括一个或者多个变量,并且变量可以是连续变量、类别变量或者混合变量时,我们都可以运用典型变量分析技术。

在典型变量分析中,第一个最高的典型相关,它是被加权的 X 变量组合和被加权的 Y 变量组合之间的相关(组合作为典型变量的第一对)。我们能在 X 和 Y 被加权的组合中发现第二个最高的典型相关,它与第一对典型变量不相关(直交的)。通过巴特利特卡方近似性检验、威尔克斯 lambda 分布以及 Fratio 考察第一、第二相关和其他典型相关(存在于所有的 $n_{\min}$ 相关中,这里 $n_{\min}$ 是 n_x 和 n_y 的最小值)的显著性。

两组变量间的典型相关系数(Rci)表示两组变量间关系的强度。相关系数的平方 Rci^2 表示在同一组典型变量中,一个变量对另一变量的方差解释率。

2. CVA 与回归分析和主成分分析的关系

以上对第一个和其他典型相关的描述可以同另两类分析相比较。

第一,多元回归分析。在这里,Y 组只有一个变量,通过取 Y 变量(通常叫作标准变量)和 X 变量组(称作预测变量)相关系数(Ry)的最大值得到了 X 变量的加权组合。我们可以把 Ry^2 看作标准变量的变异比率,这些标准变量能被 X 变量组,特别是预测变量"预测"(或"解释")。

第二,在每组中,第一典型相关、第二典型相关以及其他变量的构成在因素分析中都具有单变量的相似性,每一个因素都是 X 变量的加权组合。理论上的因素分析,其第一个因素的方差最大。于是第一个因素的影响从原先的相关矩阵中排除,第二个变量与第一个变量不相关。每一个成分(变量或因素)的负载,当被乘方或相加时,能够决定从因素中萃取的方差值。范德格尔(Van de Geer)论述了回归分析与 CVA、因素分析和 CVA 间的关系,他提出了关于多元模型的两种观点:图示观和矩阵操作观,着重指出"多元分析技术不是为特定情境所准备的技术,它不是进行单独分类,而是组成更紧密的集体"(Van de Geer 1971 P. 91)。

3. 典型变量分析的图示表示法

为了检验和表示典型变量分析结果(Van de Geer 1971),通常较恰当的方法是提供结果图示。一般情况下,方格里是观测变量,圆圈里是潜变量,前者由标准的字母表示,后者由希腊字母表示。如果变量 χ 只与变量 η 相关,那么联系的双向性就由一条双箭头的曲线表示。典型变量分析的分析过程本身并不与可以用于解释分析结果的不同的概念化相区别。这些概念化的东西是从假设的理论中得出的模型,通过理论预测模型的性质。如果在 X 变量组和 Y 变量组之间没有偶然的相关,那么我们可以把典型相关看作一种双因素分析。如图 1 所示。

然而,如果我们认为 X 组中的变量决定 Y 组中的变量,模型在观念上会有所不同,变量间的关系以及各变量会以不同的方式出现。如图 2 所示。

为了简化图示,图 2 中的潜变量 χ 和 η 分别对应于 X 组中的变量和 Y 组中的变量,我们把它们综合起来,就得到模型的简化形式或等价式,如图 3 所示。

4. 结构系数和转换权重

有两类系数有助于对典型变量分析做出解释,即"转换权重"和"结构系数"。有时会用"函数"和"典型权重"代替"转换权重",用"标准化的典型负载"一词代替"结构系数"。对典型变量分析中因

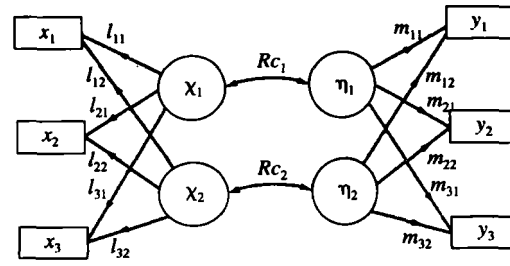

图 1 典型变量分析可以被看作是具有两个垂直潜变量的双因素分析

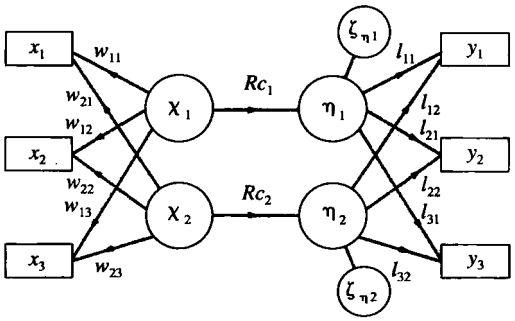

图2 典型变量分析可以被看作是具有两对垂直潜变量的回归模型

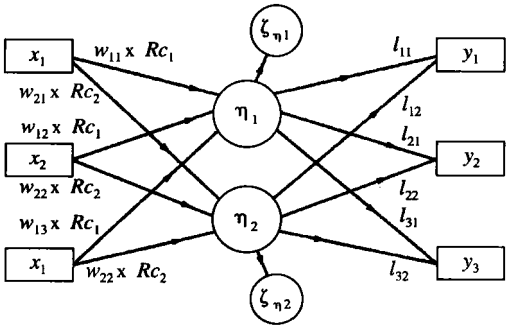

图3 典型变量分析可能被看作是具有两个垂直潜变量的简化的回归模型

素的解释传统上以标准化的转换权重为基础，它类似于多元回归分析中的β权重，在形成每一个潜变量时，它作为变量的线性组合指向原始变量。另外，我们可以运用结构系数，它是每一个典型的导出变量和原始变量的相关系数。从多因素分析的角度看，这些结构系数或者负载能够鉴别典型变量对，并以其特殊的方式与对应的变量相关。而且，结构系数的平方和能决定通过因素分析而得到的各组的方差。

塔苏罗卡(Tatsuoka)曾经论述过结构系数的应用：

> 当我们意识到标准化权重(典型或判别的)是其他变量的影响被消除或控制后的偏系数时，很显然这是一种更合理的解释典型因素(反对估计每一个原始变量对因素的相对贡献)的方法。当我们的目的是考察在众多的变量中每一变量的贡献时，这种方法很适合，但是当我们希望做出大量的解释时，这种方法不适合。(Tatsuoka 1973 P. 280)

然而，我们能从与模型性质有关的系数中得到不同的意义，根据模型来解释典型变量分析。

在图2显示的路径模型中，我们假定X变量完全决定潜变量χ_1和χ_2，这些潜变量部分决定潜变量η_1和η_2，残差为$\zeta_{\eta 1}$和$\zeta_{\eta 2}$。转变权重(w)给出了从变量X到变量χ的路径，而结构系数l给出了从变量η到变量Y的路径。如图2中所示，典型相关系数Rc_1和Rc_2显示了潜变量χ对潜变量η的影响。

在图3中，潜变量χ被忽略了，箭头由变量X直接指向两个潜变量η，由对应的转换权重与典型相关系数的乘积得到路径系数的值。如图3所示，残差也影响潜变量η。

在图1中，模型运用典型变量分析测查两组变量，即进行双因素分析，适当的结构系数(l)提供了从潜变量到观测变量的路径。因此，结构系数和转换权重在解释数据中具有不同的作用，其作用取决于由理论所假设的模型的性质。

应该注意不是所有的计算机程序都能提供结构系数和转换权重，因为所用的模型可能要求对两套系数都要进行充分理解，必须精心选择使用哪一种计算机程序进行典型变量分析，用一套特定的数据去检测由相关理论观点导出的特定的模型。

5. 对方差的考虑

通常一些用来进行典型分析的计算机程序，也包括打印输出和冗余测量(Cooley and Lohnes 1971)。这些测量方法包括由一种典型变量可以预测的另一典型变量的方差比例Rc_i^2，以及由一测验组可以预测的另一测验组的方差比例(Vi)。每一个典型变量都是这样决定的，我们称它为因素冗余：$Rc_i^2 \times Vi$。把因素冗余相加就得到由另一组决定的一组因素的总冗余值。因此，是因素冗余(而不是Rc_i^2)描述了能被另一组预测的变异的多少。而对于单变量回归分析模型，Ry^2表示可预测的方差的比例。这些冗余检验方法说的要比做的多，而

有关典型分析结果的典型相关系数做的要比说的多。但是,皮尤和胡(Pugh and Hu 1991)认为要避免对冗余系数的解释,因为它们缺乏多元属性。

6. 冗余的分离

梅斯克(Mayeske)等人(1969)在对数据的重新测量(这些数据是从一项关于教育机会均等的全国学校调查中得到的)时通过一种程序把受到联合影响的方差估计分离开,威斯勒(Wisler 1969)和穆德(Mood 1971)对这一程序进行了更加充分的论述。这种技术最早由牛顿和斯珀雷尔(Newton and Spurrell 1967)提出,皮克(Peaker 1971)在他的Plowden成长报告中使用了这种技术。穆德(1971)指出了这项测量技术的不足——在特定的情境下,去解释方差的联合影响可能是不正确的,因为这些影响是通过相互干扰造成的。尽管在解释分离方差结果方面存在不足,但它对于引起人们对联合影响的注意还是一项有用的技术,这些联合影响可能发生在两个单独的预测变量之间,也可能发生在预测变量组(以标准变量形式对变异进行解释)间。

后来,库利和洛内斯(Cooley and Lohnes 1976)设计了典型分析中的方差分离程序。一组预测源对经过标准测量的多元预测值的唯一贡献是由整个模型所解释的标准的部分变异,如果不用预测变量的特定部分,就不能得到部分变异。由特定预测源预测的部分冗余将解释典型分析中这些预测源何时被加入到其他的预测源中。预测源的联合影响是由整个模型解释的冗余,它小于单独影响之和。因此,总冗余可以分离成单独的和联合的影响,联合的影响可以分离成部分影响。

分离方差和冗余的过程类似于把相互交叠的数据分开的过程。如果有两组以上的预测源,那么就可能根据不同的预测组分离冗余;如果有四组以上的预测源,那么可能的分组数就会很大,并且不是很有意义。

库利和洛内斯(1976)为找到分离冗余的规则感到骄傲。在这里,$V(j,k,l)$是典型回归模型的冗余,预测组为J、K、L。这些规则运用了维恩图,如图4所示:

两个预测变量组:

变量组1 = $U(1) = -V(2) + V(1,2)$

变量组2 = $U(2) = -V(1) + V(1,2)$

变量组1和变量组2的共同部分 = $C(1,2) = V(1) + V(2) - V(1,2)$

三个预测变量组:

变量组1 = $U(1) = -V(2,3) + V(1,2,3)$

变量组2 = $U(2) = -V(1,3) + V(1,2,3)$

变量组3 = $U(3) = -V(1,2) + V(1,2,3)$

共同部分 $C(1,2) = -V(3) + V(1,3) + V(2,3) - V(1,2,3)$

共同部分 $C(1,3) = -V(2) + V(1,2) + V(2,3) - V(1,2,3)$

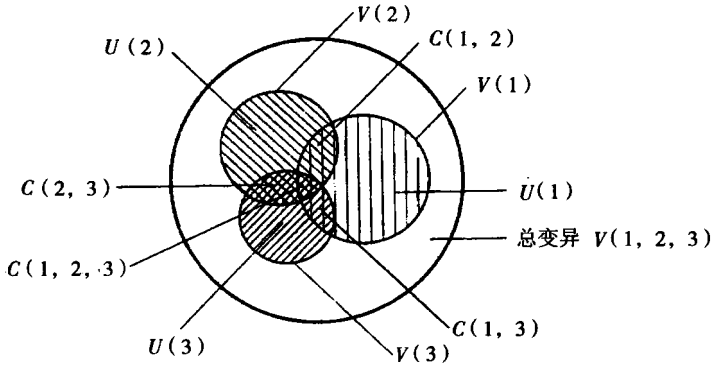

图4 三组预测变量的维恩图

共同部分 $C(2,3) = -V(1) + V(1,2) + V(1,3) - V(1,2,3)$

共同部分 $C(1,2,3) = V(1) + V(2) + V(3) - V(1,2) - V(1,3) - V(2,3) + V(1,2,3)$

基夫斯(Keeves 1975,1986)举了典型分析的例子,里面既包括图示表示法也包括冗余分离法。在冗余分离中,运用这种分析过程所得的数据分析结果可以同运用其他过程所得的结果相比较。

7. 显著性测验

在典型变量分析中,当能对从整个数据提取的潜在数据进行显著性检验时,它也提供了对所得到的典型相关数据进行显著性水平检验的方法,这些检验需要简单随机抽取的样本。然而,简单随机抽取的样本是特例,教育测验中的数据并不都是简单随机抽取得到的。因此,所用的显著性检验都是不合适的。而且,对于转换权重、结构系数以及冗余系数的显著性检验没有普遍适用的程序。因此,研究者必须精心地检验不同系数的大小,以决定在对数据进行解释时加入哪个系数是合适的。

8. 一些发展

皮尤和胡(1991)论述了典型分析的用途和解释意义。最有趣的是他们做出了与分析中变量的数量相关的被试量的论述。他们认为没有既简单又合适的规则,建议尽可能使用大样本,样本量和变量的比例至少为20:1,这对于进行精确的解释是必要的。而且,他们强烈支持运用交叉效度检验分析的恒定性。

用典型分析进行预测和描述,很可能用逐步进入的方法对存在的变量进行分析。然而,如果使用逐步进入的方法,就需要有交叉效度。另外,通过计算剩余相关的合适的矩阵,有可能运用部分(偏)典型分析,但是这种分析过程还没有被广泛应用于大量研究(Thorndike 1978)。

另外,范德格尔(1971 P. 169 ~ 170)认为当典型变量中的第一对和第二对变量相关或者非直交时,应该运用典型分析过程。奥尔斯特(Horst 1965)和范德格尔(1968)提出了解决问题情境的方法,在这里有3个或3个以上数据矩阵,运用典型变量分析同时将它们匹配。但是,LISREL的发展使得典型变量分析在检验这些以及更为复杂的模型时有被代替的趋势。

J. P. 基夫斯(J. P. Keeves)
J. D. 汤普森(J. D. Thompsen) 著
张宏学 译

附录

Bartlett M S 1941 The significance of canonical correlations. *Biometrika* 32:29—38

Cooley W W, Lohnes P R 1971 *Multivariate Data Analysis.* Wiley, New York

Cooley W W, Lohnes P R 1976 *Evaluation Research in Education.* Irvington, New York

Darlington R B, Weinberg S L, Walberg H J 1973 Canonical variate analysis and related techniques. *Rev. Educ. Res.* 43: 433—454

Horst P 1965 *Factorial Analysis of Data Matrices.* Holt, Rinehart and Winston, New York

Hotelling H 1935 The most predictable criterion. *J. ExP. Psychol.* 26:139—142

Keeves J P 1975 The home, the school and achievement in mathematics and science. *Sci. Educ.* 59(4): 207—218

Keeves J P 1986 Canonical variate analysis. *Int. J. Educ. Res.* 10(2):164—173

Mayeske G W et al. 1969 *A Study of Our Nation's Schools.* United States Government Printing Office, Washington, DC

Mood A M 1971 Partitioning variance in multiple regression analyses as a tool for developing learning models. *Am. Educ. Res. J.* 8:191—202

Newton R G, Spurrell D J 1967 A development of multiple regression for the analysis of routine data. *Appl. Stat.* 16:51—64

Peaker G F 1971 *The Plowden Children Four Years Later.* National Foundation for Educational Research, Slough

Pugh R C and Hu Y 1991 Use and interpretation of canonical correlation analysis. *J. Educ. Res.* Arti-

cles:1978—1989; *J. Educ. Res.* 84(3):147—152
Tatsuoka M M 1973 Multivariate analysis in educational research. In:Kerlinger F N (ed.) 1973 *Review of Research in Education 1*. Peacock, Itasca, Illinois
Thorndike R M 1978 *Correlation Procedures for Research*. Gardner, New York
Van de Geer J P 1968 *Matching k Sets of Configurations*. University of London, Department of Data Theory, London (Report RN-005—68)
Van de Geer J P 1971 *Introduction to Multivariate Analysis for the Social Sciences*. Freeman, San Francisco, California
Wisler C F 1969 Partitioning the explained variation in a regression analysis. In:Mayeske G W et al. 1969

其他参考文献

Pedhazur E J 1982 *Multiple Regression in Behavioral Research*. Holt, Rinehart and Winston, New York
Thompson B 1984 *Canonical Correlation Analysis: Uses and Interpretation*. Sage University Paper Series on Quantitative Applications in the Social Sciences, 07—001. Sage, Beverly Hills, California

聚类分析(Cluster Analysis)

对相似物体、相似的人以及其他相似事物进行分组,从而划分出类别,是人类最基本的能力之一。例如,史前人类应该能够分辨许多具有共同属性的物体,以确定哪些能吃,哪些有毒,哪些是凶猛的动物,等等。非常明显,把相似的东西分成不同类别的观念是很原始的,因为从广义上说,分类是语言发展的必要条件,语言由词组成,这些词帮助人们辨认和讨论不同类型的事件、物体和所遇到的人。如语言中的每一个名词都是用来描述具有显著特征的事物的标签。所以,动物才有猫、狗、马等等诸如此类的名字,这样的名字把个体集中起来形成组。命名就是分类。

分类不仅是一种基本的人类思维活动,它还在科学的大多数分支中起基础性作用,它具有两个基本的科学功能:(a)对感兴趣的或者正在调查的事物进行描述;(b)建立概括的规则或理论,通过这些规则或理论可以解释或预测特定的事件。分类在生物学中起着非常重要的作用,在生物学中,试图对生物体进行分类可以追溯到亚里士多德(Aristotle)时期,分类是达尔文(Darwin)进化论发展的前提;在化学中,门捷列夫(Mendeleyev)的元素周期表中对元素的分类对于揭示原子结构具有深远的影响;在医学中,对疾病进行适当的分类是寻找病因和治疗疾病的必要前提;在教育中,研究者经常热衷于对学生和教师进行分类。

1. 数值分类技术

20 世纪 60 年代以来,数值分类方法发展迅速,显然它是随着电子计算机的发展而发展的,因为数值分类技术需要计算机进行大量的数学计算。大多数数值分类技术处理两种矩阵:一种是由变量所得分数矩阵,它直接对这一矩阵进行处理,使每一事物或个体被分类;另一种是由原始数据矩阵导出的距离或者相似性矩阵。

我们可以通过许多标准软件包学习和使用聚类方法,这些软件包有 BMDP、SAS、SPSS,还有其他更加专业的软件包,如 CLUSTAN、GENSTAT、S-PLUS。从广义上,聚类可划分为三种最通用的方法:聚集层次法、K-均值迭代法、综合法。无论如何,我们记住下面一点非常重要,就是在一些情况下,用简单的图表法鉴别组或聚类就足够了。例如,这里有一组样本的体重和身高的记录。这组数据的散点图如图 1 所示。我们可以清晰地看到由点组成的截然不同的两"类"。在某种意义上,大多数聚类技术的目的是模拟人类研究者或者想用机器代替人,使分类过程自动化,当数据只包括两个变量时,人类能做得很好,而聚类技术试图把它运用到更普遍的情境,这些情境是每个被调查的个体有两个以上的变量值。埃弗里特(Everitt 1993)提出了一些其他有助于聚类的绘图法。

在介绍下面的内容之前,我们应当注意的一点是:聚类与聚类之前的数据有关。在研究开始时,我们不知道类的数目和组成,正是这一点使聚类分析技术和判别分析技术(分组数目已知是其前提)区分开。埃弗里特和邓恩(Everitt and Dunn 1992)

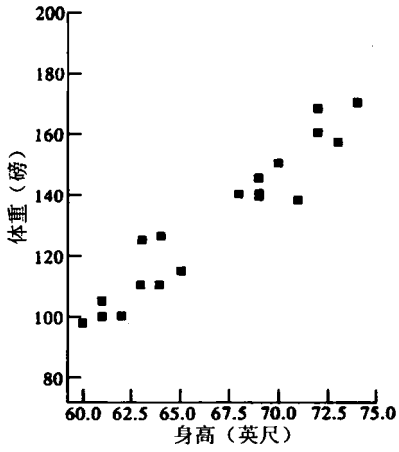

图 1　身高与体重的散点图

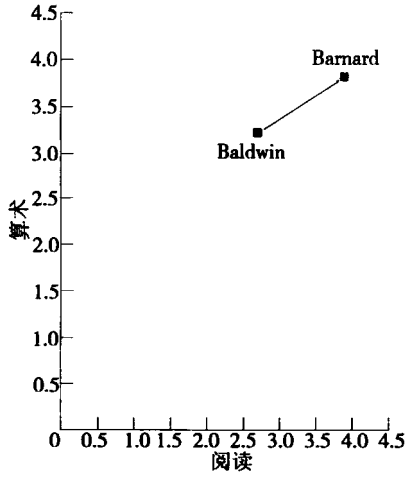

图 2　两组测度间的欧氏距离

详细描述了这两种方法。

2. 距离测度和相似测度

聚类分析的原始数据通常是由许多被试所得的一组分数，如表 1 所示。这些数据来自 25 所学校 4 年级和 6 年级学生阅读和算术的成就测验分数。对学校的聚类可能既可以区分出学校的不同水平，又能估计出从 4 年级到 6 年级不同学校变化模式的异同。

许多聚类分析方法（尽管不是所有的方法）都不直接对每一个个体的变量值进行分析，但它们会对由原始数据导出的个体间距离或者相似值的矩阵进行分析。例如，最通用的距离测度是欧氏距离，图 2 显示了用这种方法计算的表 1 中前两个学校 4 年级的成绩。

当两个个体都包括两个以上变量值时，用欧氏距离计算的公式为：

$$d_{ij} = \sqrt{\sum_{k=1}^{p} (x_{ik} - x_{jk})^2} \qquad (1)$$

p 表示变量的个数，x_{ik} 和 x_{jk} 中 $k = 1, 2, \cdots, p$，而 k 是有关的两个个体的变量值。

表 2 显示了表 1 中前 5 个学校的欧氏距离矩阵。埃弗里特（1993）指出了使用欧氏距离法的问题，特别是变量的规模问题。

相似测度是聚类分析的常用方法，特别是对于

表 1　25 个学校的成就测验分数

学校	4 年级		6 年级	
	阅读	算术	阅读	算术
Baldwin	2.7	3.2	4.5	4.8
Barnard	3.9	3.8	5.9	6.2
Beecher	4.8	4.1	6.8	5.5
Brennan	3.1	3.5	4.3	4.6
Clinton	3.4	3.7	5.1	5.6
Conte	3.1	3.4	4.1	4.7
Davis	4.6	4.4	6.6	6.1
Day	3.1	3.3	4.0	4.9
Dwight	3.8	3.7	4.7	4.9
Edgewood	5.2	3.9	8.2	6.9
Edwards	3.9	3.8	5.2	5.4
Hale	4.1	4.0	5.6	5.6
Hooker	5.7	5.1	7.0	6.3
Ivy	3.0	3.2	4.5	5.0
Kimberley	2.9	3.3	4.5	5.1
Lincoln	3.4	3.3	4.4	5.0
Lovell	4.0	4.2	5.2	5.4
Prince	3.0	4.2	5.2	5.4
Ross	4.0	4.1	5.9	5.8
Scranton	3.0	3.2	4.4	5.1
Sherman	3.6	3.6	5.3	5.4
Truman	3.1	3.2	4.6	5.0
West Hills	3.2	3.3	5.4	5.3
Winchester	3.0	3.4	4.2	4.7
Woodward	3.8	4.0	6.9	6.7

数据来源：Hartigan 1975