



心理统计分析:趋势、 契机与展望*

文 / 侯杰泰¹ 刘红云² Herbert W. Marsh^{3,4,5}

1 香港中文大学教育心理系 香港

2 北京师范大学心理学院 北京 100875

3 Centre for Positive Psychology and Education, University of Western
Sydney, Australia

4 Department of Education, Oxford University, U. K.

5 King Saud University, Saudi Arabia

【摘要】近年来,心理学研究领域所使用的统计方法日渐复杂,本文综述了近几十年来,心理学研究中统计方法的应用趋势,引出结构方程模型成为流行分析方法的原因。前沿复杂的统计方法,由于不易掌握,往往令实际应用者感到困惑,但方法的发展也给心理学理论的验证、拓展带来了契机,如前沿的统计技术为测验工具跨文化应用中信效度的检验提供了分析方法。展望未来,新的统计方法将持续发展,我们仍需不断学习新的统计方法,不断调整对数据分析的思考角度。在教学和培训方面,研习班、编制教材以及善用网上平台,将成为我们推动心理统计的工作重点。

【关键词】应用统计,方法学,心理统计,结构方程模型,心理学课程

DOI: 10.3969/j.issn.1000-3045.2012.Z1.022

Maller 在 1934 年回顾了 1894—1933 年主要心理学期刊的文章^[1],当时部分作者引入统计作分析,可以说是非常简单的方法(如相关系数),Maller 也详尽讨论了其优劣。在中国,辛治华在 1965 年对《心理学报》1959—1965 年的文章所采用的统计方法进行了评析^[2],讨论了“我国心理研究工作中的一些统计问题”。其内容包括:(1)取样是否随机有代表性;(2)百分数的正确使用(如,当

样本量太小,百分率没意义);(3)平均数、中数和众数的正确使用;(4)集中趋势与离中趋势的配合;(5)正确统计推论(取样代表性)。这些内容,已包含在现在的初高中数学中。

80 年后的今天,心理学研究者所面对的困难是如何应对近 10 多年来所涌现的统计方法。这些方法令他们感到困惑和彷徨,他们不知道是否应紧贴潮流,尽量使用这些新的、复杂的方法,但是熟练掌握并能正确运用这些新的方法对大多数研究者并不容易。

* 本研究得到国家自然科学基金项目(31100759)资助
修改稿收到日期:2012 年 12 月 24 日

本文综述了近几十年来,心理学领域统计方法应用的趋势,由此引出结构方程模型成为流行分析方法的原因。新的统计方法虽然艰深不易掌握,但给心理学的研究也带来了新的契机。一方面,心理学的理论不断对统计方法提出新的要求,从而促进方法的发展和完善;另一方面,统计方法的发展也不断促进心理学理论的验证和拓展。同时,新的统计方法对变量间关系的分析更加精确,对心理测验工具的发展和跨文化研究的适用性也提供了更好的验证方法。

展望未来,我们认为统计方法的发展仍然是日新月异,随着信息技术的进步,我们的思考角度也需调整。如何适应发展,快速普及新的统计分析方法,使其能更好地为研究者应用。如何组织协调培训,尤其是善用网上平台,将会成为我们的工作重点。

1 趋势及前沿方法

自心理学(或其他社会科学)学者大量在研究中采用统计方法后,不同学者都定期评估研究生需要学习什么统计方法,才能够顺利了解文献及进行研究。

1.1 统计方法的应用趋势

Skidmore 及 Thomson^[3]综合了不同时期的综述文章,包含共约 12 000 篇发表于 1948—2001 年的心理和教育领域的综述文章。依出现在文献的次序为:回归(1885 年出现)、相关、卡方分析、方差分析、t-检验、协方差分析、因子分析、聚类分析及非参数分析(1942 年)。他们共占有所有方法的 75%,这些方法也差不多在所有博士课程内教授。

粗略来说,近年来方差/协方差分析、相关分析和 t-检验的使用率都有下降趋势,但回归及因子分析的使用则呈上升趋势。非参数方法的使用在 20 世纪 60 年代达至顶峰后随后下降,但近年则略呈上升的态势。因子分析(近年已包含结构方程分析在内的验

证性因子分析)在研究院多独立设立一门课程教授。

方差分析及 t-检验使用的下降也是很富启发性的,一如 Skidmore 及 Thomson^[3]指出,早期很多人误以为在方差分析中,用了分类自变量便能够将研究变成实验设计,并具备实验研究的所有优点。这种误解近年逐渐得以澄清,也使得使用方差/协方差分析的人减少。另一方面,以往一些研究者将连续变量硬切割为类别变量(如将收入划分为上、中、下 3 组),以便做方差分析,这不但使变量的信度下降,也不会因为用了方差分析有额外好处,这错误观念渐改,也令方差分析使用下降。

近年来,统计方法的应用呈现出不断整合的趋势,尤其是当统计学家指出 t-检验^[4]、方差/协方差分析及回归分析都是广义线性模型的特例后,使用回归方程去替代 t-检验及方差/协方差的人愈多,也解释了为什么回归方程的使用者上升。虽然回归方程是典型相关的特例^[5],但后者在尚未盛行时已被证明是结构方程模型的特例^[6],被取而代之。

1.2 结构方程模型

自 90 年代以来,统计学家发展出多种多元统计方法, Hershberger^[7]回顾 1994—2001 年的文章,指出业内对结构方程模型的使用逐年倍增,成为多元统计最重要的工具。过往聚焦于不同形式的探索性因子分析方法已成过去,由验证性因子分析(结构方程模型内的特例)所取代。

结构方程模型所以垄断成为主流,是因为它包含其他常用技巧^[8]:多元回归、路径分析、主成份分析、探索因子分析(验证性因子分析)、多元方差分析、方差分析、判别分析、被试内(重复测量)分析、典型相关(canonical)分析、项目反应分析、时序分析等。



中国科学院

结构方程模型提供了非常灵活的框架来处理潜变量和观测变量,将心理学中的潜在构念、变量间的复杂关系整合起来,广义线性模型、路径分析以及因素分析等统计模型都是其特例。

近年来,研究者证明社会和行为科学研究中应用的许多模型都是结构方程模型的特例。Takane、de Leeuw^[9]与Kamata、Bauer^[10]证明一些项目反应理论(IRT)模型可采用结构方程模型进行分析。Bauer^[11]证明了多水平模型可以用结构方程模型的形式表示,结构方程模型也与多层回归模型结合,发展出的多层结构方程模型为同时分析多层次、多变量间复杂的关系提供了可能。Cheung^[12]提出可以将元分析的模型整合到结构方程模型的分析框架下,并指出用结构方程模型的方法更有优势。将这些模型整合在一个统一的模型框架下最大的优点就是可以用来同时分析这些模型所能描述的复杂的研究问题。Mplus^[13]将结构方程模型、多层次模型、混合模型、生存模型、潜类别模型和一些IRT模型整合到了统一的统计模型框架下。

另外,统计模型和参数估计方法的发展可以更合理地处理实际研究中遇到的问题。在心理学的研究中,由于设计或者其他原因经常会遇到缺失数据,随着极大似然估计方法和贝叶斯估计方法的发展,在结构方程模型的框架下提出了一些新的缺失数据的处理方法^[14],这些方法在大多数的结构方程模型的软件中都已包含(如LISREL、Amos、Mplus),并且不需掌握太复杂的技术就可以应用。同时,方法的发展也越来越考虑到数据的实际情况,Muthén^[15]总结并拓展了当数据非随机缺失(NMAR)情况下,常用的处理缺失数据的模型以及拓展的模型。

最后,除了能处理连续正态分布的数据外,也发展出能处理其他类型数据,如类别(Categorical)型数据的数学模型和软件包。一些操作计算机熟练的学者更喜欢用R语言及其内的共享程序,去处理各类大大小小的特殊分析。故此心理学理论所涉及的各类研究课题,都难找借口说欠缺软件

包或处理方法,而不能进行分析。

2 新统计方法带来的契机

21世纪以来,计算机运算速度倍增,统计软件包的接口更清晰易用,使用复杂统计方法看似更容易,然而也带来更多错误理解与误用。但总的来说,统计方法的进展为理论发展和建设带来以下新的契机^[16,17]。

2.1 旧理论问题引发统计方法发展

很多新的统计方法都是针对存在已久的理论问题。例如,在工业心理学中,我们希望了解雇员与雇主的关系时,数据是聚类(cluster)形式,传统计算相关、回归等方法由于违背了独立性的假设,均不适宜。假设有50间公司,即50名雇主;每间公司有10名雇员,共500雇员。如果想知道雇员的满意度与雇主的管理风格之间的关系,传统上有2种方法,一是在个体层面计算500名雇员与对应50雇主(每名雇主复制10份)管理风格的相关;二是在公司层面先计算每间公司雇员的平均满意度,再计算这50个均值与50名雇主管理风格的相关。然而,这2种传统方法都不恰当。同理,在研究学习心理时,一般来说,学生个人的社经地位与成绩有正相关,但班中各同学社经地位的均值,却往往与学生成绩成负相关,这些处于不同水平数据的复杂关系,在80年代以前难以妥善解决,多层回归模式及相应的计算机软件出现后,这些问题都能一一解决。

在做文献综述时,我们常借助元分析,早期多用主观判断,因为每个研究可能提供数个互有关联的功效量(effect size),近年用多层数据分析和结构方程模型的方法进行元分析,可以更好地处理效果量不同质以及协变量存在缺失数据的问题。

传统分析中因子分析、回归分析、方差分析都是结构方程的特例(详见1.2节)。很多心理学理论涉及交互效应,如,某类效应只在某类人或其情景下发生,但传统多元回归未能很好地处理误差项,未有足够敏感度去检验这些交互项,令这些涉及交互项的心理模式未获支持及证明^[18-20]。采用潜

变量结构方程模型后,很多心理模型将会获得重新验证的机会。

2.2 新统计方法协助建立新的理论与研究法

心理学者很快地扩展新统计方法的应用,从而丰富他们的理论和研究方法。例如,上述多层回归模型可转化成增长(growth)模型,用以分析追踪研究的变量(如,儿童的成长发展特点,青少年某些特质的发展变化)。传统上这类追踪性研究,均以固定时间段去取样和观察,如在研究的第1、3、5、7、9年去测量语言能力。但采用多层回归模型框架下的增长模型后,我们可以对每个儿童不同时段、不同时间间隔的观察进行分析,新的统计方法改变了传统研究设计的限制。

基于结构方程模型基础上的增长模型,不仅可以对特质随时间的变化趋势进行分析,而且可以对个体间发展趋势的差异和原因进行分析和解释。潜类别分析模型和混合分析模型还可以对不同质发展群体进行更合理的分析和解释。这些新统计分析技术的应用,可帮助人们更深入细致地分析心理学的现象,并建立和拓展心理学理论。

2.3 处理误差、更准确描述心理建构的关系

心理学和其他社会科学内很多构念(construct)(如,人格、智力、能力、态度),都不能用单一问题或指标去量度,传统上我们将数题目得分计算均值或因子分(加权后的均值),但都不是理想的处理方法,自从90年代结构方程模型及其软件包出现后,测量上的误差都能通过潜变量获得更好的处理,构念间的关系也有了更准确的估计。

2.4 协助建立跨文化测验

跨国家、文化种族使用心理或教育(包括能力及智力)测验日渐普遍,而跨地或跨种族比较的规模也愈来愈大。如,TIMSS^[21]

涉及50个国家,PISA 2009^[22]有70多个国家或地区参与。不过Oakland调查^[23]指出,世界上80%以上的人都是住在没有本土化常模式测验的地区,也就是说,他们的测验直接取自其他地区或国家,只经过简单翻译便立刻采用。

在采用他国测验工具时,除关注道德(包括版权、测验使用者资格)等问题外^[24, 25];也需留意数据处理及分析的要求。前沿的统计方法,能协助我们建立这些跨文化测验的信度及效度。International Test Commission (ITC)特为此做了多项研究,并出版指引,任何大规模的跨地区语言系统研究,应到该会网站参考他们的建议。

虽然学业成就(如,数学能力)测验与心理测验(如,人格问卷),看来差异颇多,但在测量特性本质上,仍有很多共同之处。由其他地区引入某个测验后,我们难免要做各种分析,报告测验信度和效度等,应该明白这些指标受被试群体特性的影响很大,故此随便将原样本的信度和效度应用于与原报告差异甚大的群体,并不合适。

在处理跨文化或地区组别的数据时(见上述ITC网站,亦见^[23, 26-28]),我们尽可能找到工具原产生地的数据,以检验工具在跨文化(或种族、地域)的结构等同性和测量等同性。在结构等同上,我们关注的是这个概念的意思及结构在新的文化中,是否与工具的原产地有所不同^[29]。在测量等同上,我们希望保证每因子内每个题目(指标)与因子的关系(如,因子负荷)及测量尺度(metric)是跨文化一致的。

3 展望

应用于心理学的统计方法,不但在方法及技术上日新月异,在统计推论等统计思考的取向上,也可能随着时代有所改变。我们不但要在心理学理论上持续学习,对各种新



中国科学院

的统计方法,也应有一定的认识。最后在如何推动学习新的统计方法,我们亦提出一些建议^[30]。

3.1 日新月异的统计方法

如上文所显示,20世纪六七十年代前,心理学所用的统计方法,现已多在中学或本科生课程教授。如今主流心理学期刊文章所用的前沿统计方法,如,结构方程模型,在国内只有少数研究生课程独立成科教授。

不过,即便在这些前沿的方法中,它们对一些基本的分析概念,也可能在不断地演化及改变。例如:当采用心理测量工具时,经常将数题目合并处理变成量表(题目组,打包, item-parceling;即计算数题目的均值或总分)。设研究人格,共25道问卷题目,可组成5个人格因子。数十年前,要计算5个人格得分,有2种方法。一是用主成分分析(探索性因子分析),容许每题都能与每因子有点关系,计算5个人格因子分。另一种是将与某一因子相关的题目相加,求均值(或总分),这成为题目组(item-parcel)得分。结构方程模型的验证性因素模型出现后,我们在做验证性因子分析时,很整齐地将每一题目,归属于某一个因素内,一般不容许题目跨于数个因子间,也就是说,每题只从属单一因子。

近年一些软件包(如,Mplus),容许将探索性因子分析结合在结构方程模型下(称为Exploratory Structural Equation Modeling, ESEM)。一些学者(如,Marsh等^[31])甚至推崇除非能证明没有需要外,我们应习惯地先将各题目从属于所有因子,再找这些因子与其他外在因子的关系。否则,如果简单地计算题目组(数题目的均值)与外在因子的得分,会误估因子间的关系。要紧贴这些技术及处理方法的改变并不容易。

3.2 思考角度的改变

随着统计方法使用者日益增多,统计学家对应用统计方法的使用者,也提出甚为重要的批评。我们用近几十年来有关误用统计讨论最多的议题——虚无假设的显著性检查(Null Hypothesis

Significant Test, NHST)为例,来展示统计使用者因为对统计方法认知加深,思考角度也对应转移,故此在分析及报告结果时,亦有所改变。

自研究者广泛使用Fisher^[32]的显著性检查后,很多统计使用者对NHST有错误的理解或过高的期望。与其他学科(经济学、医学、化学等)出现的问题相同^[33-37],许多统计学者对研究者普遍错误理解及使用NHST,不断重复地提出忠告,指出问题,但更改进展缓慢,错误情况依旧。

我们不在这里详细介绍错误应用NHST所涉及的问题(详见Cohen^[34]等经典文章;亦见温忠麟,吴艳^[38, 39]的综论及Cortina及Dunlap^[40]、Hagen^[41]、Wainer^[42]的不同意见),简单来说,是误解 p 值的意思和功能,Cohen指出 p 值是在检查 $P(D|H_0)$ (在已知假设下数据的机会),而非大部分研究者希望知道的 $P(H_0|D)$ (已知数据下假设的机会)。

导致这种问题的原因很多,因为教授、教科书、主要期刊仍继续着错误传统;这对 p 的错误解读,是更易理解,看来更合理;再者,可以简易取代 p 的方法也不多,Fidler和Cumming^[35]指出^[43],实验心理学自从1940年迅速发展,在20世纪50年后期(即10年后),美国心理学会(APA)86%的主要期刊文章均使用NHST。

可是,影响千本以上相关期刊写作要求的APA出版手册,对该问题少有提及,在1994年出版的手册^[44]只建议作者报告功效量,在2001版^[45]亦只是强力建议作者报告置信区间(亦见Wilkinson et al.^[46])。一些期刊编辑(如,1994—1998 Memory & Cognition 编辑 Geoff Loftus^[47];亦见 see also Fidler 等^[48])强烈要求报告标准误及误差范围等,但当他们任期满后,新任编辑便不再坚持这些要求。

Fidler和Cumming^[35]认为,期刊编辑应要求而非建议所有投稿者使用新的标准(不再单单依靠NHST)。期刊之间也应采取一致行动(非个别期刊的要求),教科书及教学内容也应相应改动,让新的思维成为心理学研究者的思考角度及方法。

心理学界对NHST仍存误解,对实质性差异(substantial difference)等概念完全未有听闻者甚多,改革路途甚为遥远。

由此可见,对各种新旧统计方法,心理学家需抱着一个不断学习的态度,即便是看似简单的统计方法,在深入认识后,可能会有全新的角度及处理方法。

3.3 统计课程与训练

与其他大部分学科的学习相同,要学习前沿的统计方法仍然主要依靠上课、自修以及跟随导师边做边学、师徒制形式进行^[23],当然三者并行不悖。

在培训研究生方面,中国与西方国家面对相同困难,虽然统计方法愈趋复杂,但一般心理学课程设计,能给予统计方法的课时仍然不变^[49, 50],大学对心理博士生只提供1.2个统计测量课程(中位数^[51])。研究生也急于争取在学期间发表文章,以支持升学或就业的需要。故此额外用于学习统计的时间甚为有限。

心理学相对传统物理、化学等学科,在中国是年轻学科,加上熟悉新近复杂统计法的教授,在中外也不算太多。在不同国家,要推广这些前沿的统计方法,多依靠寒暑假或主要学术会议前后,举办研习班。这类学习班多以二三十人边学边练习的形式进行。这类研习班的成本太高,中国地广人多,即使倍加每班人数,还是杯水车薪。集合国内不同大学的相关学者,编制更多实际例子的教材,应是可行且十分重要有效的方向。

近年越来越普遍的网上学习平台,提供一个很好的契机^[30]。海内外以文字或讨论小组形式的网站、博客、微博及统计讨论小组甚多。但国内有系统的前沿心理学网上课程不多(如,结构方程模型课程KTHau.COM)。较高水平且有系统地结合课本与

练习的课程,应该富有成效。但我们仍需一些大学、专业学会或政府教育部门的协调、资助及推动,才能制作而成。

在编辑各类前沿统计课程或书本时,我们需留意内容编排及用语。如Hancock和Mueller^[52]指出,大部分结构方程模型的中、高级参考书,都包含太多数学和统计学知识,非一般心理专业读者消化和使用。我们亟需的是,用通俗易懂的语言去介绍这些较复杂的统计方法的教材。

一些专业性学术期刊,也特地用专刊,以整期的篇幅,介绍这些复杂统计方法应用于主要心理学理论的具体例子(如,Contemporary Educational Psychology, 2007, 32期专刊,亦见美国心理学会的大量专刊),中国学术期刊也可以仿效。

如Marsh及Hau^[16]指出,在复杂的统计方法影响下,研究生的能力也倾向两极化。一些学生只懂理论和质性研究,不懂得做也看不懂较复杂统计的文章,另一些学生则用甚多时间去学习复杂统计,但只懂统计,心理学基础甚弱。在一些国家(如,中国),本科并非心理学专业的学生,也能进入心理研究生的课程,本科为数理的学生,在博士课程专研统计方法,本科为文科的学生,欠缺心理本科的统计训练,则更专注理论性课题,令该问题两极化更为严重。

在高水平的统计研究领域,不难见到很多统计功底很好的年青人,他们的训练及兴趣集中在先进统计方法,而心理学理论基础较弱^[16]。他们所能处理的多是模拟数据研究,集中于数据分布(非正态、多层特性等)及误设(misspecified)模型对拟合的影响等。颇多课题与心理学核心理论关系不大。该现象随着统计的复杂性增加而日渐普遍,中外皆是。

理论与先进统计方法紧密结合,才对两



中国科学院

者的发展有所贡献。但这类有助推广前沿统计方法的文章,难以在国内外方法类或理论性期刊发表^[9]。

方法类期刊认为这些用实际数据的文章不够严谨,对方法学的贡献不大,故拒登。理论性期刊的主编则通常邀请熟悉方法的审稿员,从方法角度评审文章,但显而易见,真实数据罕有完全满足所用统计方法的各项假设及要求(如数据正态分布,各组人数大致相同)。故此,即使这些文章相对其他用简单统计的文章,对统计方法的使用更加恰当,但总是有千万个理由拒登。也即是说,理论性期刊对使用复杂统计方法的文章,可能有更高的要求。

尤其是当读者群对复杂统计方法认识较浅,甚至有所抗拒时,结合复杂统计与理论的文章更难打入主流理论期刊。统计与理论割裂的问题,不容易解决,除请统计老师多举心理学各种主要理论有关的实际例子外,也应多编撰这些例子的教材。

4 结语

近10多年来的复杂统计,如结构方程模型,确实使得估计变量间关系更为准确,也改变了一些传统心理研究的方法。但统计必须结合心理学的理论,要培养两者兼备的研究生或学者并不容易。本文指出一些改善策略和方向,希望对心理学的发展,有所帮助。

参考文献

- 1 Maller J B. Forty years of psychology: A statistical analysis of American and European publications, 1894-1933. *Psychological Bulletin*, 1934, 31: 533-559.
- 2 辛治华.我国心理学研究工作中的一些统计学问题. *心理学报*, 1965, (2): 178-184.
- 3 Skidmore T S, Thompson B. Statistical techniques used in published articles: A historical review of reviews. *Educational and Psychological Measurement*, 2010, 70(5): 777-795.
- 4 Cohen J. Multiple regression as a general data-analytic system. *Psychological Bulletin*, 1968, 70: 426-443.
- 5 Knapp T R. Canonical correlation analysis: A general parametric significance testing system. *Psychological Bulletin*, 1978, 85: 410-416.
- 6 Bagozzi R P, Fornell C, Larcker D F. Canonical correlation analysis as a special case of a structural relations model. *Multivariate Behavioral Research*, 1981, 16: 437-454.
- 7 Hershberger S L. The growth of structural equation modeling: 1994-2001. *Structural Equation Modeling*, 2003, 10: 35-46.
- 8 Hau K T, Wen Z L, Cheng Z J. 结构方程模型及其应用(Structural equation model and its applications). Beijing, China: Educational Science Publishing, 2004.
- 9 Takane and de Leeuw. On the relationship between item response theory and factor analysis of discretized variables. *Psychometrika*, 1987, 52(3): 393-408.
- 10 Kamata and Bauer. A Note on the Relation Between Factor Analytic and Item Response Theory Models. *Structural Equation Modeling*, 2008, 15: 136-153.
- 11 Bauer D J, Preacher K J, Gil K M. Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, 2006, 11: 142-163.
- 12 Cheung M W. L. A model for integrating fixed-, random-, and mixed-effects meta-analysis into structural equation modeling. *Psychological Methods*, 2008, 3: 182-202.
- 13 Muthén L K, Muthén B O. *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén, 2010.
- 14 Allison P D. *Missing data*. Thousand Oaks, CA: Sage, 2001.
- 15 Muthén B, Asparouhov T, Hunter A et al. Growth modeling with non-ignorable dropout: Alternative analyses of the STAR*D antidepressant trial. *Psychological Methods*, 2011, 16: 17-33.
- 16 Marsh H W, Hau K T. Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology*, 2007, 32: 151-170.
- 17 Marsh H W, Martin A J, Hau K T. A multiple method perspective on self-concept research in educational psychology: A construct validity approach. In M. Eid & E. Diener (Eds.), *Hand-*

- book of multimethod measurement in psychology (441-456). Washington, DC: American Psychological Association, 2006.
- 18 Marsh H W, Hau K T, Wen Z et al. Moderation. In Little, T. D. (Ed.), Oxford handbook of quantitative methods (Vol. 2 Statistical analyses). New York: Oxford University Press, 2012.
- 19 Nagengast B, Marsh H W, Scalas F et al. Who took the "x" out of expectancy-value theory? A psychological mystery, a substantive-methodological synergy, and a cross-national generalization. *Psychological science*, 2011, 22: 1 058-1 066.
- 20 Marsh H W, Wen Z, Nagengast B et al. Structural equation models of latent interaction. In Hoyle, R. H. (Ed.), Handbook of structural equation modeling. New York: Guilford Press, 2012.
- 21 Mullis I V S, Martin M O, Beaton A E et al. Mathematics achievement in the primary school years: IEA's Third International Mathematics and Science Study (TIMSS). Chestnut Hill, MA: Boston College. 1997.
- 22 Organisation for Economic Co-operation and Development (OECD). PISA 2009 Results: What Students Know and Can Do - Student Performance in Reading, Mathematics and Science (Volume I), OECD Publishing, 2010.
- 23 Oakland T. How universal are test development and use? In E. Grigorenko (Ed.), Assessment of abilities and competencies in an era of globalization (1-40). New York: Springer, 2009.
- 24 Leach M, Oakland T. Ethics standards impacting test development and use: A review of 31 ethics codes impacting practices in 35 countries. *International Journal of Testing*, 2007, 7: 71-88.
- 25 Leach M M, Oakland T. Displaying ethical behaviors by psychologists when standards are unclear. *Ethics and Behavior*, 2010, 20: 197-206.
- 26 Byrne B M, Oakland T, Leong F T L et al. A critical analysis of cross-cultural research and testing practices: Implications for improved education and training in psychology. *Training and Education in Professional Psychology*, 2009, 3: 94-105.
- 27 Hambleton R, Bartram D, Oakland, T. Technical advances and guidelines for improving testing practices. In P. Martin, F. Cheung, M. Kyrios, L. Littlefield, M. Knowles, B. Overmier & J. M. Prieto (Eds). The IAAP Handbook of Applied Psychology (338-361). London: Blackwell, 2011.
- 28 Oakland T. Principles, standards, and guidelines that impact test development and use and sources of information. In M. M. Leach, M. J. Stevens, A. Ferrero, Y. Korkut, & G. Lindsay (Eds.), Oxford international handbook of psychological ethics (201-215). New York: Oxford University Press, 2012.
- 29 Yan G, Saklofske D H, Oakland T. Views of Chinese psychologists toward intelligence. *School Psychology International*, 2009, 30: 456-473.
- 30 Huck S W. Reform in statistical education. *Psychology in the Schools*, 2003, 44(5): 2 007.
- 31 Marsh H W, Ludtke O, Nagengast B et al. Why item parcels are (almost) never appropriate: Two wrongs do not make a right - Camouflaging misspecification with item-parcels in CFA Models. *Psychological Methods*, in press.
- 32 Fisher R A. Statistical methods for research workers. Edinburgh, Scotland: Oliver & Boyd, 1 925/1 951.
- 33 Belia S, Fidler F, Williams J et al. Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, 2005, 10: 389-396.
- 34 Cohen J. The earth is round ($p < .05$). *American Psychologist*, 1994, 49: 997-1 003.
- 35 Fidler F, Cumming G. Lessons learnt from statistical reform efforts in other disciplines. *Psychology in the Schools*, 2007, 44: 441-449.
- 36 Hubbard R, Ryan P A. The historical growth of statistical significance testing in psychology—and its future prospects.



- pects. Educational and Psychological Measurement, 2000, 60: 661-681.
- 37 Kline R B. Beyond significance testing: Reforming data analysis methods in behavioral research. Washington, DC: APA Books, 2004.
- 38 温忠麟, 吴艳. 屡遭误用和错批的心理统计. 华南师范大学学报: 社会科学版, 2010, (1): 47-54.
- 39 Balluerkan G, Mez J, Hidalgo D. The Controversy over Null Hypothesis Significance Testing Revisited. Methodology, 2005, 1(2): 55-70.
- 40 Cortina J M, Dunlap W P. On the logic and purpose of significance testing. Psychological Methods, 1997, 2: 161-172.
- 41 Hagen R L. In praise of the null hypothesis statistical test. American Psychologist, 1997, 52(1): 15-24.
- 42 Wainer H. One cheer for null hypothesis significance testing. Psychological Methods, 1999, 4: 212-213.
- 43 Fidler F, Cumming G, Burgman M et al. Statistical reform in medicine, psychology and ecology. Journal of Socio-Economics, 2004, 33: 615-630.
- 44 American Psychological Association. Publication manual of the American Psychological Association (4th ed.). Washington, DC: Author, 1994.
- 45 American Psychological Association. Publication manual of the American Psychological Association (5th ed.). Washington, DC: Author, 2001.
- 46 Wilkinson L, Task Force on Statistical Inference. Statistical methods in psychology journals: Guidelines and explanations. American Psychologist, 54: 594-604.
- 47 Loftus G. Editorial comment. Memory & Cognition, 1999, 21: 1-3.
- 48 Fidler F, Thomason N, Cumming G et al. Editors can lead researchers to confidence intervals but they can't make them think: Statistical reform lessons from medicine. Psychological Science, 2004, 15: 119-126.
- 49 Gorard S, Rushforth K, Taylor C. Is there a shortage of quantitative work in education research? Oxford Review of Education, 2004, 30: 371-395.
- 50 Hutchinson S R, Lovell C D. A review of methodological characteristics of research published in key journals in higher education: Implications for graduate research training. Research in Higher Education, 2004, 45: 383-403.
- 51 Aiken L S, West S G, Millsap R E. Doctoral training in statistics, measurement and methodology in psychology. American Psychologist, 2008, 63: 32-50.
- 52 Hancock G R, Mueller R O. (Eds.). Structural equation modeling: A second course. Greenwich, CT: Information Age Publishing, Inc., 2006.

Statistical Analyses in Psychology: Trends, Crises, and Prospect

Kit-Tai Hau¹ Liu Hongyun² Herbert W. Marsh^{3,4,5}

(1 Department of Educational Psychology Chinese University of Hong Kong

2 School of Psychology, Beijing Normal University 100875 Beijing

3 Centre for Positive Psychology and Education, University of Western Sydney Australia

4 Department of Education, Oxford University U. K.

5 King Saud University Saudi Arabia)

Abstract There has been an increasing use of sophisticated statistics in psychological research in recent years. We reviewed the trends in the statistics used in the last several decades and explained the popular use of structural equation modeling. Despite the potential confusion on researchers, the new complicated statistics provides chances for changes and theoretical im-

(转至 208 页)