

北京大学

心理学教材

心理与 行为科学 统计

甘怡群等 编著

2-43



北京大学出版社
PEKING UNIVERSITY PRESS

北京大学心理学教材

心理与行为科学统计

甘怡群 张轶文 邹玲 编著



北京大学出版社
PEKING UNIVERSITY PRESS

图书在版编目(CIP)数据

心理与行为科学统计/甘怡群等编著. —北京:北京大学出版社, 2005. 11

(北京大学心理学教材)

ISBN 7-301-08974-0

I. 心… II. 甘… III. 心理统计-高等学校-教材 IV. B841.2

中国版本图书馆 CIP 数据核字(2005)第 045297 号

书 名: 心理与行为科学统计

著作责任者: 甘怡群 等编著

责任编辑: 陈小红

标准书号: ISBN 7-301-08974-0/C·0330

出版发行: 北京大学出版社

地 址: 北京市海淀区成府路 205 号 100871

网 址: <http://cbs.pku.edu.cn> 电子信箱: zpup@pup.pku.edu.cn

电 话: 邮购部 62752015 市场营销中心 62750672 理科编辑部 62752021

排 版 者: 北京高新特打字服务社 82350640

印 刷 者: 北京大学印刷厂

经 销 者: 新华书店

787 毫米×980 毫米 16 开本 21 印张 430 千字

2005 年 11 月第 1 版 2005 年 11 月第 1 次印刷

定 价: 32.00 元

未经许可, 不得以任何方式复制或抄袭本书之部分或全部内容。

版权所有, 翻版必究

前 言

本书讲述行为科学(重点在心理学)中使用的基本统计知识:描述性统计、简单的假设验证和最简单多元统计。全书共分四个部分。第一部分:描述性统计,包括次数分布,集中量数和差异量数, z 分数和正态分布;第二部分:推论统计,包括样本均值的分布,假设检验初步,二项分布, t 统计量简介,两个独立样本的假设检验,两个相关样本的假设检验,总体参数的估计,单因素和重复测量方差分析,二因素方差分析,相关,回归初步, χ^2 检验,非参数检验;第三部分:最简单的多元统计,包括多元回归分析,因素分析和多元方差分析;第四部分:对初学者进行数据分析的几点建议,包括数据分析的前期准备,非常值的处理,正态转换,对效应大小的报告等,为读者提供综合性的信息。由于读者在学习过程中非常需要一些综合性的习题,我们编写了十套综合练习题,供不同阶段的读者复习之用。

本书的特色在于:

(1) 没有深奥的数学公式和推算,其对象可包括具备最初等数学知识的读者。题目尽量用简单的数据,消除读者对数学的恐惧。对于没有学习过基本概率统计知识的读者,请参阅附录中的对几个重要分布的简介,或者阅读相关书籍。

(2) 将统计与具体的研究情境相结合,向读者传递如何对所研究的问题运用适宜的统计方法的信息。

(3) 在理论方面,吸收西方心理学界新近的统计理论方法,反映学术界最新动态的内容。加进了“统计效力”、“效应大小”、“简单主效应”等内容,反映学术界最新动态。

(4) 试图将高等统计的一些重要内容,包括应用软件包 SPSS 的输出结果,以较通俗的方式介绍给读者。

本书的第一和第二部分适合高等院校心理系本科生作为初等心理统计的教材,本书的第三和第四部分可供本科或研究生作为 SPSS 和高等心理统计的辅助教材。本书既可作为高等院校心理学和社会科学专业统计课的教材,亦可用于自学和研究生入学考试复习。

本书作者谨向下列参与本书编写或收集资料的同学致谢:方华、李洁、刘松琦、刘

兴、曲晓艳、沈秀琼、吴超荣、吴昊、杨曼殊、张怡玲、张莹、庄明科、祖霁云。

本书列入北京大学 2005 年度教材建设立项,并得到北京大学教材建设项目资助经费的支持,在此表示感谢。同时,一并向北京大学出版社编辑的工作表示感谢。

限于作者的水平,本书肯定会存在着一些不足或错误,望读者不吝指正。

甘怡群

于北京大学心理学系

目 录

1 统计和度量的基本概念	1
§ 1 统计、科学和观察	1
§ 2 实验设计和科学方法.....	2
§ 3 心理统计中常用的一些概念和统计符号.....	5
2 次数分布	11
§ 1 次数分布表.....	11
§ 2 次数分布图.....	15
§ 3 次数分布的特性和计算.....	18
3 集中量数与差异量数	23
§ 1 集中量数.....	24
§ 2 差异量数.....	30
4 z 分数、正态分布和概率	40
§ 1 z 分数及其应用.....	40
§ 2 正态分布.....	43
5 概率和样本:样本均值的分布	49
§ 1 样本均值的分布.....	49
§ 2 样本均值分布与概率.....	54
6 假设检验初步	58
§ 1 假设检验的性质和种类.....	59
§ 2 假设检验的基本逻辑.....	59
§ 3 z 检验.....	62
§ 4 假设检验的两类错误.....	64
§ 5 假设检验的前提.....	65
§ 6 假设检验的效力.....	66
7 二项分布	69
§ 1 利用正态分布求二项分布概率.....	69

§ 2	百分比检验	74
8	t 统计量简介	78
§ 1	t 统计量简介	79
§ 2	t 统计量和 z 统计量的比较	80
§ 3	t 统计量的自由度	81
§ 4	t 分布	81
§ 5	t 检验	83
9	两个独立样本的假设检验	89
§ 1	独立样本均值差异的分布	89
§ 2	独立样本的 t 统计量	91
§ 3	独立样本 t 检验的统计前提	92
§ 4	独立样本的 t 检验	94
§ 5	独立样本 t 检验的效应大小和效力	98
10	两个相关样本的假设检验	101
§ 1	有关相关样本的计算	102
§ 2	相关样本 t 检验的统计前提	108
§ 3	相关样本 t 检验的效应大小	109
§ 4	相关样本设计的问题	109
11	总体参数的估计	114
§ 1	点估计的概念和优良性	115
§ 2	区间估计的概念和一般步骤	116
§ 3	和总体均值相关的估计	118
§ 4	影响置信区间宽度的因素	123
§ 5	区间估计和假设检验的联系	124
12	单因素和重复测量方差分析	126
§ 1	方差分析的基本原理	127
§ 2	独立样本方差分析	132
§ 3	事后检验	135
§ 4	重复测量的方差分析	136
§ 5	方差分析的数据前提	141
§ 6	方差分析的效应大小和统计效力	141
13	二因素方差分析	144
§ 1	相关概念及其表示方法	145
§ 2	二因素方差分析过程	150

§ 3	二因素方差分析结果的解释	156
§ 4	二因素方差分析的统计前提	162
14	相关	163
§ 1	相关的数据表和散点图	163
§ 2	相关的特点	165
§ 3	Pearson 相关	166
§ 4	Spearman 相关	175
§ 5	点二列相关	179
§ 6	Kendall 和谐系数	181
15	回归初步	183
§ 1	回归方程	184
§ 2	回归线的准确性	187
16	χ^2 检验	191
§ 1	χ^2 匹配度(拟合优度)检验	192
§ 2	χ^2 独立性检验	197
17	非参数检验	202
§ 1	顺序型数据和秩统计量	203
§ 2	曼-惠特尼(Mann-Whitney) U 检验	204
§ 3	符号检验法	208
§ 4	维尔克松(Wilcoxon) T 检验	210
§ 5	克-瓦氏(Kruskal-Wallis)单向方差分析	214
§ 6	弗里德曼(Friedman)双向方差分析	215
18	多元回归分析	218
§ 1	多元回归分析简介	219
§ 2	多元回归过程和结果输出	222
19	因素分析	233
§ 1	因素分析简介	233
§ 2	因素分析的步骤	234
§ 3	用 SPSS 进行因素分析	237
20	多元方差分析(MANOVA)	245
§ 1	多元方差分析简介	245
§ 2	相关理论问题	247
§ 3	数据要求与统计前提	247
§ 4	使用 SPSS 完成多元方差分析	248

§ 5 重要参数及解释	249
21 对初学者进行数据分析的几点建议	257
§ 1 数据分析前的准备——数据清理	257
§ 2 非正态数据的转换	259
§ 3 假设检验的逻辑局限	259
§ 4 报告统计效力的必要性	262
§ 5 有关统计图	262
§ 6 应用多种统计方法,得到适当统计结论:一个小样本 数据分析的例子	263
参考文献	269
附录 概率基本知识	270
综合练习题	273
综合练习题 1	273
综合练习题 2	275
综合练习题 3	278
综合练习题 4	281
综合练习题 5	282
综合练习题 6	285
综合练习题 7	287
综合练习题 8	290
综合练习题 9	293
综合练习题 10	295
选择题答案	298
附表	299
附表 1 标准正态分布表	299
附表 2 t 的临界值表	302
附表 3 Cohen's d 与两个样本分布的不重叠部分百分比	303
附表 4 F 的临界值表	304
附表 5 HSD 检验中 q 的临界值	307
附表 6 F_{\max} 的临界值表	308
附表 7 Pearson 相关的临界值表	309
附表 8 相关系数 r 值的 Z_r 转换表	310
附表 9 Spearman 相关系数的临界值表(双尾)	311
附表 10 χ^2 的临界值表	312

附表 11.1 曼-惠特尼 U 检验的临界值表(双侧)	315
附表 11.2 曼-惠特尼 U 检验的临界值表(单侧)	316
附表 12 符号检验的临界值表	317
附表 13 维尔克松 T 检验的临界值	318
附表 14 克-瓦氏单向方差分析 H 临界值表	319
附表 15 弗里德曼双向等级方差分析的临界值表	321

1

统计和度量的基本概念

§ 1 统计、科学和观察

统计不仅是行为科学研究离不开的工具,我们在日常生活中也常常要和它打交道。例如,我们可以通过一个国家的GDP数值来了解它的发展水平,也可以通过一个地区的降雨量来了解近来的天气情况,还可以通过交通事故发生率来了解交通安全情况……具体到更细致的地方,我们在公司里要统计员工的出勤率,在学校里要统计学生成绩的合格率、升学率等。其实这些都是统计。利用这些统计得到的平均数、百分比等数字,我们可以获得需要的信息,不仅能对事物有一个整体的了解,而且迅速有效,因为我们只需要这一个或几个数字,而不用去关注每一个个体的情况。

以上我们举的例子是最简单的统计现象。在心理学研究当中,我们要通过发问卷、做实验等方式来收集数据,为了对研究假设进行研究,我们需要分析数据,得到支持研究假设的证据。这其实就是一个统计的过程。我们所说的统计学包括描述统计,推论统计和多元统计。本书中将介绍实用性强的描述统计和推论统计以及最简单的多元统计。

统计方法可以帮助我们有效地获取信息,得出结论。在一个研究当中,研究者使用统计方法来归纳整理数据,并推论出研究假设的正确与否。这个过程需要被其他研究者所理解,并可以在不同研究之间进行比较,因为统计学给研究者提供了一套标准化的方法,这些方法是在整个科学领域里通用和默认的。因此,我们可以清楚地理解其他研究者的统计思路以及研究结果。比如说,一个研究者报告一个班级同学的平均学业成绩是81.40分,标准差是9.56,我们就可以知道这个群体的学业成绩的大体是良好的,但离散程度还是很大的,同学之间的学业成绩并不是很接近。从这个角度来说,统计学可以帮助我们更好地学习、理解他人的研究,促进学术交流和发

展。行为科学,称之为科学,自然也是在事实的基础上进行观察研究,而要将这些观察结果转化成一目了然的可以为大家所理解的信息,就必然要使用统计方法。心理学作为行为科学的一个组成部分,其发展的历史证明,科学心理学离不开科学实验或调查,而心理实验或调查无论从最初的研究设计和控制,还是收集、录入数据,以及最后的数

据分析,从而得出有意义的研究结论,都离不开统计的帮助,这些使得统计学在心理与行为科学中已经成为一个不可或缺的科学工具。例如,如何收集资料才能最有效的反映所研究的课题;采用什么方法整理和分析所得数据才能最大限度、最客观地呈现这些数据所反映的信息;怎样才能把抽取的样本中所获得的结果推广到总体中,作出一般规律性的科学结论,等等。没有统计学的参与,这些问题的解决是不可想像的。

§ 2 实验设计和科学方法

科学的主要任务就是发现宇宙万物的规律。在古代,人们就注意到了周围世界的许多规律性,比如四季的更迭、潮涨潮落、月亮的阴晴圆缺等,并且有意识地对这些现象进行观察从而阐明这些规律性的变化。这些本身是变化的或者对不同个体有不同值的特征或条件,我们就将其称之为变量(variable)。而常数(constant)则是相反的,其本身是不变的且对不同的个体的值也相同。

一、实验设计

科学涉及到对不同变量之间关系的探索。比如说在广告的投放量和产品的销量之间存在着一种关系。在最开始的阶段,广告的投放量不断增加,产品的销量也随之增加;待广告投放了一段时间以后,产品的销量开始稳定,不再随着广告投放量的增多而提高。而且如果此时减少广告的投放的话,有可能销量还会回落。如果我们想要弄清楚这两个变量之间的关系究竟如何,就必须对这两个变量进行观察,即计算广告的投放量和产品的销量。在心理学研究中,我们常常用以下几种方法来研究这些变量间的关系:相关研究、实验研究、准实验研究和非实验研究。

相关研究(correlational method)也可以看做是观察研究(observational method),即观察在自然情境中存在的两个变量,是寻求变量之间关系的最简单的方法。比如说我们想知道中学生的学业成绩高低和每天学习时间是否有关联,可能我们会假定学习时间越长学业成绩越高。为了验证这种假定是否正确,我们可能就要同时观察这两个变量,最后得到二者之间的一个关系。但是这种相关研究只能够提供两个变量之间相关程度的结果,却不能提供因果关系的证据,我们不能够说学业成绩高是因为学习时间长。如果想要进一步研究因果关系,确定学习时间长是学业成绩高的原因,就要控制许多无关变量,设定自变量和因变量,这些就是实验研究的任务。

实验研究的目的是要确定两个变量之间的因果关系,即一个变量的变化是不是由另外一个变量的变化引起的。在一些比较复杂的实验研究中,可能需要同时控制几个变量的变化来观察另外几个变量的变化,我们这里只举最简单的例子,即控制一个变量的变化,来看另外一个变量的情况。实验研究一般来说有两个特征:(1) 研究者需要操纵一个变量,然后观察另外一个变量,看这种操纵是否带来了变化;(2) 要对研究中的

其他一些无关变量进行控制,以确保这些变量不会对研究的结果产生影响。研究者需要操纵的变量在实验研究中称为自变量(independent variable);需要观察的变量就是因变量(dependent variable)。研究者的目的就是要看自变量的变化能否引起因变量的变化。但是,能够引起因变量变化的因素有很多,我们可以把这些能够对因变量造成影响的变量统称为有关变量,其他的就是无关变量。在上面那个例子中,如果我们想要考查学习时间长短和学业成绩高低之间是否具有因果关系,那么学习时间就是自变量。在行为科学研究中,自变量常常包括两个或更多的处理条件。比如我们可以将学习时间的长度分为平均每天6,8,10个小时这样三个水平,我们也可以处理成每天5,9个小时这样两个水平,这要随着研究的需要而定。而学业成绩是因变量,我们可以用阶段考试的分数作为学业成绩的指标。所有能对学业成绩产生影响的因素都可称为有关变量,其他如天气、学生的身高等则是无关变量。而在有关变量当中,除了自变量之外,其他的变量是研究者不感兴趣的,即额外变量,但是必须控制这些变量才能真正了解自变量对因变量的影响,因此额外变量又叫做控制变量。在上述例子中,必须控制的变量之一就是智商,即必须保证所有被研究的中学生智商在同一水平。当然可能还有许多其他会影响实验结果的变量需要控制。

顾名思义,准实验研究(quasi-experimental design)就是介于真实验研究和非实验研究之间的一种研究,它对无关变量的控制好于非实验研究,但它又不能像真实验研究控制得那么充分和严格。在真实验研究中,我们往往是控制自变量,然后对因变量的结果进行比较。而在准实验研究中,自变量是一些我们无法控制的、自然存在的因素,如性别、时间序列等,我们考查已有的各组被试间的差别(如性别差异)或在不同时间内采集的数据的差异(如处理前和处理后),这里的分组变量称为准自变量,每个被试的分数则是因变量。例如,我们想要研究20岁和40岁不同年龄的人记忆能力的差异,我们可以先给两个年龄组的被试同样的学习材料,让他们不断记忆,直到记住为止。一周以后,让这两个组的被试进行自由回忆,从而比较记忆能力的差异。那么这里被试的年龄就是分组变量即准自变量,而回忆材料的数量则是因变量。被试其他的一些特征,如教育背景、性别等都是额外变量,有可能与准自变量产生混淆,应该加以控制。那么我们最好在选择20岁和40岁的被试时,在教育背景、性别、种族或者其他所有的特征上让两个年龄组的被试是完全等同的,然而这在实际中基本上是不可能的,所以完全的实验研究在这个问题中是不可能实现的,我们为了获得一些有用的信息,不得不退而求其次,采取折中的办法——准实验研究。

比起准实验研究,在控制的严格性上更弱一些的是非实验研究,它一般用于考查自然存在的变量之间的关系,是一种对现象的描述。常见的方法有观察法、问卷法等。例如考查一种认知治疗方法对于社交焦虑患者的疗效,可以在治疗前和治疗后分别测量患者的焦虑分数。但是这种非实验研究能够控制的因素更少,相比准实验研究也没有控制组进行比较。从以上准实验和非实验研究的例子中,我们也可以看到,心理学的研

究不同于完全的自然科学研究,它的许多现象是不能用绝对的真实实验研究去进行的,如果非要完全的采用实验研究的方法,就会损失大量的有用的信息,从而不能更全面和深刻的探究心理学现象,因此,在我们追求科学性的同时,也要考虑到心理学领域的特殊性,使用一些准实验、非实验的研究方法,来探讨社会文化中的心理学问题。

二、科学方法

在心理学乃至大部分科学研究中,理论和假设是两个不可或缺的概念。心理学理论是对行为的潜在机制的一系列陈述,它可以用来解释心理学各个领域中的问题,我们所做的心理学研究也常常是在理论背景的指导下进行的。而在进行每个心理学研究之前,我们都会提出假设,在统计分析时,也需要在提出虚无假设的基础上进行。假设是针对每次研究更加具体的一种预测,它会提出不同变量之间可能的关系。比如说,在有些实验研究中,一个假设往往就是预测如何操控自变量能够影响到因变量,在有些研究中,一个假设就是预测某个变量在不同的环境或者在时间发展的不同点上会有什么不同的表现。事实上,一个研究就是要来验证某假设正确与否。当研究的结果与我们的假设相符,那么,这个从一系列理论中得到的假设成立,原来的理论得到加强或补充;如果研究的结果与我们的假设背道而驰,那么假设被推翻,我们需要将这个新的发现填充到理论中,对理论进行修订,然后从中得到一个新的假设,继续进行研究。

理论中常常会包含一些假设的概念,这些概念可以帮助描述许多行为表象下潜藏的机制,但正因为它们是假设的潜藏机制,因此是无法观察到的对象。我们将这些概念称为构念,心理学中常见的构念包括智力、人格、动机等,这些构念无法实际观察,也不像学业成绩那样可以明确量化。那么,如何研究它们呢?为了解决这个问题,我们就必须去定义这些概念,从而使它们可以被观察和研究。这些定义我们就称之为操作性定义。同样的一个构念,可以有几种不同的操作性定义,经过操作性定义的解释后,将原本无法观察和研究的构念转化成可以观察的过程和操作。比如说,我们可以将智力定义为《韦氏成人智力量表》的得分。看到这里,读者也许会产生这样的疑问:既然一个构念可以拥有不同的操作性定义,那么到底哪个可以真正代表构念呢?其实心理学的研究最独特的地方方便于此,因为许多心理和行为现象提升到构念这个层次后,是无法触摸到的,因此也很难有人告诉我们哪个操作性定义所得到的结果是更接近其实质的,也正因此才会产生信效度的问题。我们希望所使用的操作性定义是具有较高的信效度以及易操作性的。在心理学领域中,会有许多操作性定义是得到普遍认同的,这些定义可能经过了反复的验证,并在最大程度上接近构念的理想化状态。当我们在进行研究的时候,对待任何一个要研究的构念,对于如何产生操作性定义都要谨慎一些,要不断地推敲,并且借鉴前人的经验,从而使得到的操作性定义更少被质疑,更多被认同,这样研究的结果才能被认可和普遍接受。

§ 3 心理统计中常用的一些概念和统计符号

一、总体、样本和随机取样

在进行心理学研究时,我们首先要做的是确定研究对象,这就涉及到总体的确立,以及样本的选取。

总体(population)是指具有某些共同的、可观测特征的一类事物的全体,构成总体的每个基本单元称为个体。在心理学研究中,总体是特定研究所关注的所有个体的集合,我们往往根据研究的兴趣和目的规定研究的总体,其特征和范围也随目的和要求的变化有所不同。比如某个人格障碍研究所关心的是所有12~18岁青少年这个总体,而一个记忆衰退研究可能关注的就是40~60岁的中老年总体。总体既可以是有限的也可以是无限的,全靠如何定义和推理这个总体。例如,要研究一天中经过某个地点A的人数是多少,如果设定要研究的是2003年3月份的每一天经过A的人数,那么就可以得到一个总和为31的有限的总体,而如果研究从古至今再到未来的每一天经过地点A的人数,那么这个总体就是无限的。

研究一个问题,最好的情况是对总体中的每个个体加以测量,但在实际研究中,往往无法对整个总体进行研究,有的是无法办到,有的是人力财力的限制,有的是根本没有必要。因此我们只能从总体中抽取一些个体作为真正的研究对象。从总体中选择出的个体的集合,我们称之为样本(sample)。换句话说,样本是总体的子集。在研究中,为了解决上述问题,我们常常以样本为基础,通过统计推论,得到关于总体的结论。从总体中抽取的样本有大有小,一般来说要依据研究的目的而定。例如,进行美国总统选举的民意调查,它所选定的样本数量就不可能是10个人或者100个人,而可能要几千人。而如果我们要研究单盲病人的脑结构损伤情况,两三个个体组成的样本就已经是难能可贵的了。一般来说,样本越大就越接近总体,对总体的代表性就越强,所反映的结果就越接近总体的真实情况;样本小时,有可能产生取样的偏差,个别数值的变化就会造成整个统计结果的变化,误差较大,对总体的代表性也较差。

从总体中抽取一个小样本的优点是节约整个总体研究的时间与开支,假如抽样恰当,即样本具有较强的代表性,就能将样本信息以最小的误差(近似正确)推断到整个目标总体。这里就涉及随机取样的问题。随机取样(random sampling)是从总体抽取样本的一种策略,要求总体中的每一个个体被抽到的机会均等,用随机取样法得到的样本叫做随机样本。随机取样的方法是多种多样的,然而不论采用什么方法,在进行随机取样之前,对总体的特征有个全面地了解都是非常必要的。只有对总体的分布、特征等有了全面地认识,才可能选取恰当的随机取样方法,保证所抽取的样本在最大可能上具有同总体一致的分布和特征,达到采取随机方法所希望的效果。比如,要研究某大学一年级

学生目前学业成绩与高考成绩是否相关,那么就需要了解这个大学各个院系一年级学生的人数和男女比例情况。假如物理学院男女比例是 10:1,那么在抽取的时候就应该注意保证进入样本的物理学院的学生男女比例也是 10:1,这样才能使得分析的样本能够在最大程度上代表总体。具体的随机取样方法及适用情境将会在后面的章节中介绍。

二、描述统计和推论统计

当心理学研究得到了所有被试的反应数据之后,我们首先想知道的无非是数据的总体情况:这些被试的分数位于哪个区间?平均水平是多少?离散情况如何(被试的分数是较为接近还是差异较大)?……要了解这些问题,我们就需要对数据进行描述统计。描述统计(descriptive statistics)是指用来整理、概括、简化数据的统计方法,侧重于描述一组数据的全貌,表达一件事物的性质。最常用的描述统计莫过于求平均值,不论一组数据中包括多少个值,100 个还是 1000 个,最终都可以用一个平均值给全部的数据提供有关集中趋势的简单的描述。而要了解数据的离散情况,我们常用标准差来提供这方面的信息。描述统计常常利用图、表的方式来表示,给我们一个直观的对数据的整体认识。

但是在心理学研究当中,我们常常不满足于描述统计得来的结果,因为心理学研究的目的常常是要对一个总体的某个问题得出特定的结论来验证研究假设。但由于现实条件的限制,我们无法逐个考查总体当中的每一个个体,只能利用有代表性的样本来推测总体的情况。这就是推论统计的理念。推论统计(inferential statistics)就是指运用一系列的数学方法,将从样本数据中获得的结果推广到样本所在的总体。进行推论统计的关键在于所抽取的样本要能够尽量接近所要研究的总体,能够充分代表总体,使得用样本中的信息来推论总体时产生的误差最小。我们可以根据具体的研究情境选择适合的随机取样方法。

三、参数、统计量和取样误差

对于在研究中所获得的数据,我们有必要来区分它们是来自总体还是样本。总体的任何一个特征,我们称之为总体参数。参数(parameter)是描述总体的数值,它可以从一次测量中获得,也可以从总体的一系列测量中推论得到。而样本的特征则用统计量表示。统计量(statistic)是描述样本的数值,它可以从一次测量中获得,或者从样本的一系列测量中推论得到。一旦取了样本,我们就可以知道统计量,但是我们在总体范围内换一个样本,统计量可能就改变了。也就是说,参数是一个固定的数值,而统计量的值是不定的,它会随着我们所取的样本而变化。在心理学的很多研究中,总体参数常常是我们无法知道的,因此我们也常用样本统计量来估计总体参数。

为了区分一个特征是描述总体的还是样本的,我们往往用不同的数学符号来表示

参数和统计量。如总体平均数用 μ 来表示,而样本平均数用 \bar{X} 来表示。

既然样本不能完全等同于总体,那么在样本统计量和总体参数之间自然就会多多少少地存在着一些差异,这些差异就是取样误差。取样误差(sampling error)是指样本统计量与相应的总体参数之间的差距。比如,你第一次见到一个陌生人,并且只和他交谈了五分钟,之后你会对这个人有一些评价:聪明的或是愚蠢的,开朗的或是腼腆的等。你的这些判断都是基于你和他的一面之缘以及那五分钟的谈话,这就相当于统计中所说的,这仅仅是一个样本,你会用这个样本来推论总体,即这个人的所有的人格品行等。换句话说,你用第一印象建立对一个人的认识就好比用一个样本来推断总体一样,在后面越来越多的接触中,你会发现,里面存在着许多偏差,不够准确,即基于有限的信息所作出的推论或结论有可能不是那么准确的。影响取样误差的因素有很多,包括样本容量、总体的变异情况、取样方式等。当然,取样误差是难以完全避免的,我们所要做的就是尽可能减少取样中的误差,并使之保持在研究所允许的范围内。具体的方法(如取样方法等)我们会在后面陆续介绍,这里就不再赘述。

四、离散型变量和连续型变量

在心理学研究中,变量的类型通常可以分为离散型变量和连续型变量,变量的类型影响到统计分析方法的选取。

离散型变量(discrete variable)是由分离的、不可分割的范畴组成,在邻近范畴之间没有值存在。以掷骰子为例,掷出的1点和2点之间是没有任何其他值存在的,我们不会说我掷出了1.2点。通常我们说到离散型变量,比较有代表性的是计数数字,比如说一个年级组织出去旅游,需要各个班级统计一下人数,有的班是40人,有的班是41人,在40人和41人之间是没有其他值的。同样的,那些某方面性质存在差异的不同类别也可以是离散型变量,比如说一个心理学家研究人格障碍问题,根据一些标准将人格障碍病人分为反社会型人格障碍、强迫性人格障碍、表演性人格障碍等,这些人格障碍类型就是离散型变量,因为我们可以观察到相互区别而又有限的类别。

当然,还有许多变量不是离散型的,比如说时间、长度、质量等,它们都不可能被限定成一组分离的、不可分割的类别。只要我们愿意,可以用米、分米、厘米、毫米,甚至是微米作为度量单位来测量长度,我们称这样的变量为连续型变量。连续型变量(continuous variable)在任何两个观测值之间都存在无限多个可能值,它可以分割成无限多个组成部分。例如我们测量一组被试的反应时,有的被试的反应时是0.3秒,有的被试是0.4秒,只要测量工具足够的精确,我们完全可以测到0.34秒,0.35秒,甚至0.345秒这样的反应时。一般来说,一个连续型变量可以用一条连续的实数直线表示,在实数直线上存在着无限的点,在任何两个相邻的点之间依然可以找到无数个点。

但是,我们在说到一个连续变量的某个观测值时,往往不是指实数直线上的某个固定的点,而是代表着实数直线上的一个区间。例如,23所代表的区间是从22.5到