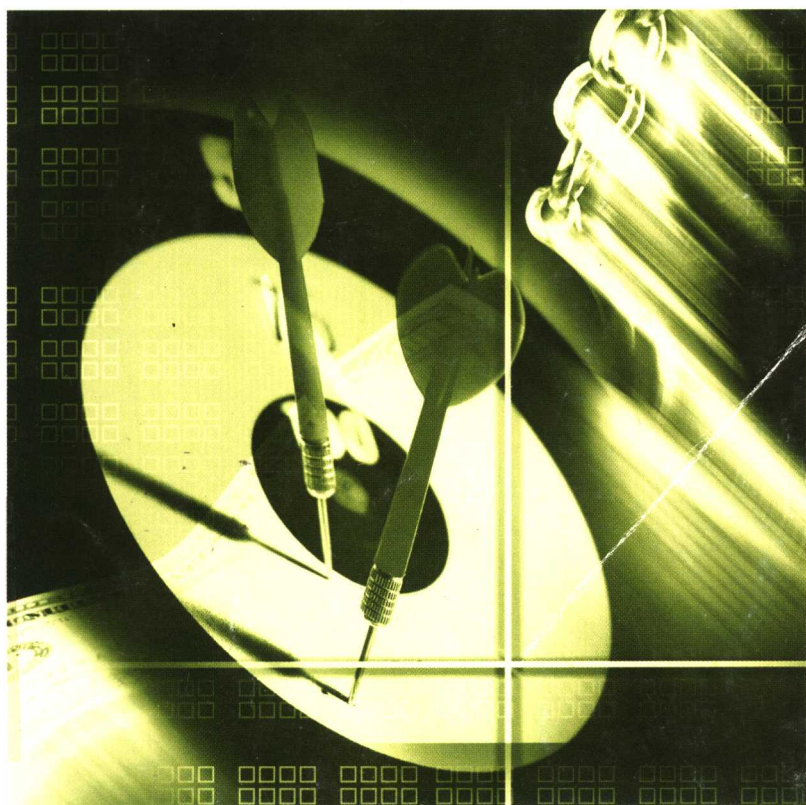


调研工具箱丛书之八 (美) 马克·S·利特维 (Mark S. Litwin) 著

周佳 高鹏 译

问卷调查 心理测验学

How to Assess and Interpret Survey Psychometrics



.7

问卷调查

心理测验学

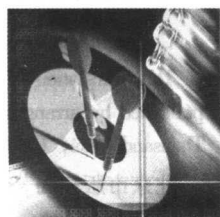


调研工具箱丛书之八 (美) 马克·S·利特维 (Mark S. Litwin) 著

周佳 高鹏 译

问卷调查 心理测验学

How to Assess and Interpret Survey Psychometrics



图书在版编目(CIP)数据

问卷调查心理测验学/(美)利特维(Litwin, M.)著;周佳,高鹏译. —北京:中国劳动社会保障出版社, 2004.2
调研工具箱丛书

书名原文: How to Assess and Interpret Survey Psychometrics
ISBN 7-5045-4391-8

I. 问… II. ①利… ②周… ③高… III. 调查研究-心理测验 IV. C31

中国版本图书馆 CIP 数据核字(2003)第 121139 号

Copyright © 2003 by Sage Publications, Inc.

This translation of How to Access and Interpret Survey Psychometrics is published by arrangement with Sage Publications, Inc., 2455 Teller Road, Thousand Oaks, California 91320 USA.

北京市版权局著作权合同登记号: 图字 01-2003-4713

中国劳动社会保障出版社出版发行

(北京市惠新东街 1 号 邮政编码: 100029)

出 版 人: 张梦欣

*

煤炭工业出版社印刷厂印刷装订 新华书店经销

850 毫米 × 1168 毫米 32 开本 3.125 印张 56 千字

2004 年 2 月第 1 版 2004 年 3 月第 1 次印刷

定价: 9.30 元

读者服务部电话: 010-64929211

发行部电话: 010-64911190

出版社网址: <http://www.class.com.cn>

版权专有 侵权必究

举报电话: 010-64911344

目 录

学习目标	(1)
第 1 章 心理测验学	(3)
第 2 章 可靠性	(8)
可靠性的分类	(11)
1. 测试—再测试可靠性	(11)
2. 交替形式可靠性	(16)
3. 内部一致性可靠性	(26)
4. 观察者间可靠性	(31)
第 3 章 有效性	(35)
有效性的分类	(36)
1. 外观有效性	(36)
2. 内容有效性	(37)
3. 标准有效性	(39)
4. 结构有效性	(46)

目 录

第4章 测量尺构建及问卷答案预编码	(49)
第5章 设计并使用编码本	(56)
第6章 试访	(62)
试访阶段工作项目检查表	(72)
第7章 文化多样性	(73)
第8章 问卷翻译	(76)
练习题	(78)
答案	(82)
建议阅读文献	(85)
术语表	(88)
作者简介	(95)

学习目标

这本书旨在指导读者如何通过对所使用的调查问卷进行详尽的检查，从而达到评估和解释所收集数据质量的目的。这本书还详细地介绍了关于编码和问卷试访方面的内容。详细的学习目标如下所述：

- 选择和应用可靠性标准，包含以下几类：

- ✓ 测试—再测试可靠性

- ✓ 交替形式可靠性

- ✓ 内部一致性可靠性

- ✓ 观察者间可靠性

- ✓ 观察者本身可靠性

- 选择和应用有效性标准，包含以下几类：

- ✓ 外观有效性

- ✓ 内容有效性

- ✓ 标准有效性

- ✓ 结构有效性

问卷调查心理测验学

- 概要地介绍测量尺的建立及问卷答案预编码方面的基础原理

- 建立并使用编码本
- 对新设计的问卷以及成熟的问卷进行试访
- 讨论问卷调查中所存在的跨文化问题

第 1 章 心理测验学

一个以收集数据为目的的调查问卷如果要取得成功，它绝不能仅仅是在样本人群中就一系列精心设计的问题进行调查。事实上，调查问卷有成功与失败之分。失败的调查问卷只能得出一些无意义的数字，所谓无意义，即指数据不可靠、不能再生、无效或是浪费了资源而没有创造任何财富。相反，成功的调查问卷能提供重要的信息，并能真实地反映我们所感兴趣的问题。

心理测验学是调查研究的一个分支学科，它可以帮助你测定一个调查问卷的优劣。心理测验学的主要原理最初是由测量心理学领域的知识发展而来的，同时，心理测验学范畴内的知识也被许多其他学科所借鉴。心理测验学为调查研究人员提供了一个可以对定性概念测量的准确性进行量化的工具，例如，对消费者满意度、情绪低落、婚姻关系等问题。这些现象从本质上来说，都比诸如邮政编码、婚姻状况或身高等离散变量更难以测量。例 1.1 展示

了两份不同的关注酒店客人满意度的调查问卷。我们很难区分哪份问卷优于另一份问卷，但是我们可以惟一确定的一点是，这两份问卷存在明显的区别而且会产生完全不同的调查结果。本书的目的就是告诉读者怎样对问卷间所存在的差异进行评价。

例 1.1 两份酒店顾客满意度调查问卷

问卷的目的：请客人在结账时填写一份含有 3 个问题的问卷，以达到评估市中心某酒店客人满意度的目的。

问卷一：请选择一个您认为正确的答案并画圈

1. 您在本酒店下榻期间感到愉快吗？

是..... 1

否..... 2

2. 您怎样评价本酒店的服务？

好..... 1

差..... 2

3. 您会选择再次入住本酒店吗？

会..... 1

不会..... 2

问卷二：请选择一个您认为正确的答案并画圈

1. 综合考虑服务、食物及其他因素，您对在本酒店

下榻期间的经历如何评价?

- 非常愉快..... 1
- 比较愉快..... 2
- 一般..... 3
- 有些不愉快..... 4
- 非常不愉快..... 5

2. 您怎样评价本酒店的服务?

- 比我所下榻过的其他酒店更有效率..... 1
- 在办事效率方面,与我所下榻过的其他酒店一样... 2
- 不如我所下榻过的其他酒店更有效率..... 3

3. 您在多大程度上会选择再次入住本酒店?

- 非常有可能..... 1
- 有可能..... 2
- 不确定..... 3
- 不太可能..... 4
- 非常不可能..... 5

例 1.1 中的哪份问卷更好呢? 问卷二看起来会得出更加有价值的信息,因为它提供的不仅仅是我们在问卷一中所看到的“是或否”类型的简单答案。但问卷一的简单明了可能会使更多的酒店客人愿意填写。如果选用问卷二,也许只有那些在入住期间感到非常不满意的客人会花时间填写,从而会忽略了那些在入住期间感到满意的客人的意

见而造成有误导性的结果。商务客人也许是酒店最重要的顾客群，但因为他们一般会比那些以休闲旅游为目的的客人更加忙于赶飞机，所以更不可能在结账时花时间填写问卷二。综上所述，我们实际上根本无法量化地评价哪个问卷更优。那么，我们又如何对调查问卷的“优劣”进行确切的定义呢？我们称之为“更优”的问卷应该是那些能更加准确地测量客人的满意度，从而得出更加有价值的信息，使研究人员可以依此对酒店的表现做出结论性评价的问卷。

严格地说，评价我们所收集的数据的质量是件十分困难的事情。相比较而言，对用来收集信息的调查问卷的准确性进行评价反而更加容易一些。这样的评价主要关注的是调查问卷的可靠性及有效性。例 1.2 中，我们看到是两个不同的用来测量电阻的工具可以产生完全不同结果的例子。如果想确定这两种测量工具中哪一个，或者是否存在一个可以达到准确测量的目的，我们惟一可以采用的方法就是直接观察测量工具的准确性以及有可能导致测量误差产生的原因，诸如答案差异及无应答的情况。

例 1.2 告诉了我们一个非常重要的概念，即在运用某个调查问卷去收集有价值的数据之前，我们必须对这个调查问卷进行测试，以保证其调查的准确性。不管我们关注的是电阻测量仪、客户满意度问卷、市场调查问卷、犯罪评估、情绪低落尺度或者其他各类调查问卷，我们都必须

这么做。我们所特别关注的不是数据有多么量化，而是调查问卷是否能令人满意地履行其调查责任。在本书的第2章及第3章中，将会更细致深入地讨论调查问卷的可靠性及有效性的问题。

例 1.2 电阻测量仪

在一个新产品测试中，两位持照上岗的电工用两个不同的电阻测量仪测量四个新电线回路的电阻值。帕特选择了一个她已经使用了15年的测量仪。杰瑞选择了一个他刚刚从厚姆兰（Homeline）公司——一家非常有名的通过邮寄方式销售产品的公司——购买的新型测量仪。在两个人测量结束以后，帕特的测量数据分别为6，16，38和119欧姆，而杰瑞的测量数据分别为19，28，73和184欧姆。因为无从判断两套数值中哪一套，或者是否有一套是正确的，所以我们要求帕特和杰瑞去测量一系列已知电阻值的、用于质量控制的电路，以达到评估这两个测量仪器本身的准确性的目的。在帕特和杰瑞分别对用于质量控制的电路测量了三次以后，我们发现帕特的测量仪更加准确。因此，我们进行新产品分析时采用了帕特的测量结果。

第 2 章 可靠性

在我们所收集的任何一组数据中，都会存在一些误差。一般来说，我们都希望能够将误差最小化，从而使我们所得到的数据能够尽可能地反映真实情况。

在问卷调查中，误差由两部分组成：随机误差和测量误差。随机误差是在任何调查中都会出现的不可预知的误差。随机误差可能由许多不同的因素造成，但主要还是由抽样方法造成的。为了降低随机误差的发生，我们可以选用一个样本量足够大的、更加有代表性的样本。但是采用大样本量通常会增加调查成本，而且经常是既不实用也不可行的。所以常用的方法是用统计学方法计算出某个调查结果产生随机误差的发生率。如果误差发生率在预期范围之内，我们就能够“否定无效假设 (H_0)”，并对研究样本下相应的结论。在统计学分析中，我们事先设定一个保守假设，被命名为无效假设，即“我们所研究的两组样本在某项研究变量上没有存在不同”。例如，在例 1.1 中，

如果我们要比较男女顾客对酒店满意度的差异，无效假设为两组人对酒店的满意度不存在差异。我们要设计一个调查方案来检验无效假设，看看我们是否能够有确凿证据来否定无效假设，然后我们可以下一个相应的关于男女客人对酒店满意度的结论。（如果需要获知更多关于假设论证方面的信息，请参考本系列丛书第九册《如何管理、分析和解读调查数据》）

测量误差是指某调查问卷在某调查样本范围内调查能力的优劣。应该说没有一份调查问卷是完美的，所以在调查过程当中出现一些误差是在我们预期范围之内的。例如，有一个跑步计时器，它只有分针没有秒针，所以就无法以秒来计算运动员的5 000米跑成绩，这个计时器最精确的计算就是到分。所有运动员成绩之间少于60秒的差异就不可能从这个计时器上表现出来，这就是所谓的跑步计时器的测量误差。测量误差越小，数据就越接近事实。然而，即使当随机误差被假定为零，测量误差还是会存在。测量误差能够反映出调查问卷本身准确与否。

可靠性是一个用来评价某调查问卷所收集数据的可再生性和稳定性的统计测量标准。例2.1用一个例子给我们介绍了可靠性的原理，这个例子是比较两个用来测量布料的工具。在这个例子中，荣誉卫士（Honor Guard）公司中的金属尺比诺特索（Notso）公司中的橡皮尺更加可靠，出现测量误差的机会也更低。因此，荣誉卫士公司的测量

结果能够更加准确地反映实际情况。即使这样，还是会存在一定程度的测量误差。如果荣誉卫士公司的销售人员没有仔细看或者没能准确辨认 35 英寸和 36 英寸，那么雷恩买的布就仍然有被量短了的可能。不管是金属尺还是橡皮尺，它们都不是完全可靠的，在测量中总会出现一定程度的测量误差。

例 2.1 布料测量

荣誉卫士公司在销售布料时，选用的是较贵的铝尺来测量每次客户需要购买的布料的尺寸。诺特索公司销售布料时也用尺子来测量，但选用的是非常便宜的橡皮尺。雷恩想买 3 码（1 码为 36 英寸）蓝色的棉布给他的双胞胎的小床制作防护栏。在诺特索公司，雷恩通过过去在此店的购买经验判断这家公司对所卖布料的测量经常不一致，因为橡皮尺可以被不同程度地拉伸。由于销售人员的力气大小的差别，诺特索公司销售的 1 码布的长度有可能是在 35~41 英寸之间。而在荣誉卫士公司，所有测量的布料长度基本一致，所存在不同也是由于销售人员对测量长度观察的差异所造成的。雷恩选择去荣誉卫士公司购买布料，因为他知道在那儿，如果想买 1 码布，得到的就是 1 码。

可靠性的分类

没有一份调查问卷或一个测试是完全可靠的，但是某些调查问卷肯定较其他的更为可靠。当我们评估一组数据的价值时，应从评估调查问卷的可靠性开始。一般来说，可以通过三种形式对可靠性进行评估：测试一再测试、交替形式及内部一致性。在下文中，还将更详细地讨论观察者本身可靠性及观察者间可靠性的相关问题。

1. 测试一再测试可靠性

测试一再测试可靠性是最常被用来衡量调查问卷可靠性的指标。它主要测量的是一组数据在多大程度上可以再生。调查人员可以在两个不同时间点，让同一群被访者回答同一份问卷，然后分析答案的稳定性。研究人员可以计算相关系数，即 r 值来比较被访者的两组答案（请参考本系列丛书第九册《如何管理、分析和解读调查数据》）。这些相关系数被整体地认为是调查问卷可靠性的指标。一般来说，如果相关系数不小于 0.70 那么就被认为是理想的。相关系数可以告诉我们在多大程度上，在不同时间点做的