

扩展知识:

一. 测量标准误的优缺点:

测量标准误的优点是:

①测量标准误的单位和测验分数相同,故在解释个人的测验结果时,可以直接指出其误差范围。

②测量标准误不受群体变异性的影响,即不同形式的测量标准误基本上是相似的。而信度系数的大小是和群体得分之离差(即分数的分布范围)有关,因此,测验手册所提供的信度总是和所要测量的群体有些差异。与信度系数相比较,测量标准差的稳定性较高,较有实用价值。

测量标准误的缺点是:

①无法对不同测验间的信度进行比较。如果要评价测验结果的质量或比较不同测验信度的高低,就一定要采用信度系数。只有在对个别分数进行解释时,才采用测量标准误。

②因为测量标准误是根据信度系数计算得到的,而信度系数有很多种。所以,测量标准误也有很多种。在具体运用时,要根据要求,选用不同的测量标准误。例如,我们对一个人在三个月期间分数的稳定性感兴趣,我们就可以拿三个月为时距所求出的稳定性系数作为信度系数,以此估计相应的测量标准误。

③上面所介绍的估计测量标准误的公式,是建立在“不论高分还是低分,其标准误都相同的”假设基础上的。而事实上,这一假设并不完全成立。因此,所求得的是整个分数范围内的平均测量标准误。

二. 几种类型的心理测验的信度系数

几种类型的心理测验的信度系数,供参考。

表 1 几种心理测验的信度系数

测验类型	信度		
	低	中	高
成套成就测验	0.66	0.92	0.98
学术能力测验	0.56	0.90	0.97
成套倾向性测验	0.26	0.88	0.96
客观人格测验	0.46	0.85	0.97
兴趣测验	0.42	0.84	0.93
态度量表	0.47	0.79	0.98

(根据 Aiken, 1985, P91)

所以,我们可将某测验的信度与已知的同一类型的测验信度相比较,以分析其信度的高低。通过对测验结果的信度分析,可以发现编制测验中的缺点,并帮助教师提高编制测验的技能。但是必须指出的是,获得较高的信度系数并不是心理测验追求的最终目标,它只是迈向目标的一步,是使测验有效的一个必要条件。

三. 利用 SPSS 进行信度分析

为了解心理与教育测量工具的可靠性与有效性,要用 SPSS 做测量结果的信度检验。信度有外在信度(external reliability)和内在信度(internal reliability)之分。外在信度通常指不同时间测量时,量表一致性的程度,再测信度(test-retest reliability)即是外在信度最常用的检验法。在多选项量表(multipleitem scales)中,内在效度特别重要,所谓内在效度指的是每一个量表是否测量单一概念(idea),同时组成量表题项的内在一致性程度如何,他最常使用的是 Cronbach a 系数。

在 SPSS 中,专门用来进行测验信度分析的模块为 Scale 下的 Reliability Analysis; 使用 Data Reduction 之下的 Factor 模块。Reliability Analysis 模块主要功能是检验测验的信度,主要用来检验折半信度、库利及 a 系数以及荷伊特信度系数值。至于重测信度和复本信度,只需将样本在二次(份)测验的分数的数据合并到同一数据文件之后,利用 Correlate 之下的 Bivariate 求其相关系数,即为重测或复本信度;而评分者信度则就是使用的 Spearman 等级相关及 Kendall 和谐系数。

实现信度分析具体执行方式如下:

例 1 求出工作倦怠量表各因素层面(已知共四个层面)及总量表的内部一致性系数

在数据编辑页面“Data Editor”中按“统计分析/量表法/信度分析(Statistics/Scale/Reliability Analysis)”操作顺序,打开信度分析的对话框。

操作 1

将问卷中归为第一个层面的题项选入右边的“项目(Items)”框内,在“模式(Model)”方框选项选取“Alpha”(内部一致性 a 系数检验),下面表 4-6 呈现出了 Model 每一个操作的中文术语。

表 4-6 Reliability Analysis 模块的 Model 选项的参数及对应中文术语

关键字	功 能
Alpha	Cronbach a 系数
Split-half	折半信度, n 是第二分量表的题数
Guttman	Guttman 最低下限真实信度法
Parallel	各题目变异数同质时的最大概率(maximum-likelihood)信度
Strict parallel	各题目平均数与变异数均同质时的最大概率信度

操作 2

按“统计量”(Statistics)钮,出现“Reliability Analysis: Statistics”(信度分析:统计量)的次对话框。下面表 4-7 呈现出了 Statistics 每一个操作的中文术语。

表 4-7 Reliability Analysis 模块的 Statistics 部分选项的参数及对应中文术语

关键字	功 能
F test	Hoyt 信度系数
Friedman Chi	Friedman 等级变异数分析及 Kendall 和谐系数
Cochran Chi	Cochran' s Q 检验,适用于答案为二分(如是非题)的量表
Hotelling' s T	Hotelling' s T2 检验
Tukey' s	Tukey 的可加性检验
Intraclass	量表内各题目平均数相关系数

在“Descriptives for”(描述统计量对象)框中选取“Scale if item deleted”(删除项目后之量表摘要),按“continue”(继续)钮,回到“Reliability Analysis”(信度分析)对话框。

这些操作之后,则在“Output-SPSS Viewer”结果窗口,会出现第一层面的信度检验结果。

操作 3

回到数据文件窗口，重复前述步骤，出现“Reliability Analysis”（信度分析）对话框时。此时右边“项目（Items）”框内仍保留的是刚才选入的第一层面的变量项，这时，我们要把这些变量还原至左边的空框内（如要清除所有设定，按最右边之“Reset”一重设钮，可将所有设定还原），再将第二个层面的题项全部选入右边的“项目（Items）”框内。选取完成之后，如按了按“Reset”钮后，刚才的 Model 方框及 Statistics 按钮内的选项均要重新选择。选取完成后按“OK”（确定）钮，在“Output-SPSS Viewer”结果窗口，会出现第二层面的信度检验结果。

分析第三个层面及第四个层面的信度，其操作方法与前面相同。

例 2 求量表的重测信度，即求其两次的积差相关。

在双变量数据相关中，如果两个变量均为连续变量，应以皮尔逊积差相关方法求二者之间相关程度，积差相关系数可作为两个连续变量间线性相关的指针。

操作 1

“Statistics/Correlate/Bivariate”（统计分析/相关/双变量）

操作 2

出现“Bivariate: Correlations”（双变量相关分析）对话框。

将左边变量清单之需考察因素题项选入右边“Variables:”（变量）下的空框内；

在“Correlation Coefficients”（相关系数）框内选取（积差相关系数）；

在“Test of Significance”（显著性检验）框内选取（双尾检验，用于探索性的数据分析）；

勾选最下面的“Flag significant correlations”（相关显著性讯号）积差相关检验结果如达到 0.05 显著水平，在相关系数旁边会以一个“*”符号表示，达到 0.01 显著水平，在相关系数旁边会以两个“**”符号表示。

四. 信度的特殊问题

1. 速度测验的信度

对于速度测验，不存在评分者信度，也无法计算同质性信度，而重测信度和复本信度均可按传统的方法求得，只有分半信度不能按传统方法估计。

要估计速度测验的分半信度，不能按题目的奇偶项来计划分两半测验，而应该测验时间划分相等的两部分，再求出两部分测验间的相关，才是分半信度。具体实施时是把测验题目分成数量相等的两部分，将奇数题和偶数题分别印在两份卷子上，以总测验的一半的时间作为时限分别对被试施测，然后计算两部分得分之间的相关系数。这个方法相当于连续施测两个等值的型测验，但每一部分的长度只是整个测验的一半，而被试的分数却是根据整个测验来计算的，因此求得的信度亦需要用斯皮尔曼-布朗公式校正。

如果我们无法独立地施测测验的两部分，一种替代的方法是把整个时限分为四等分，并求出在每个时限内的分数。如整个测验规定一小时的时限，主试则每隔 15 分钟发出一个预定的信号，让被试在听到信号时在做题目上划一个记号。测验完成之后，把被试在第一、四段时间里得到的分数相加，第二、三段时间里得到的分数也相加，然后计算这两半的相关，并用斯皮尔曼-布朗公式校正，这也是分半信度的一种估计方法。

2. 分测验的信度

有些测验包括几个分测验，这些分测验可以合成一个总分，也可以分别处理。例如韦氏成人智力量表（WAIS）是由言语量表和操作量表两部分构成，其中言语量表包括 6 个分测验，操作量表包括 5 个分测验，每个分测验有一个分数，据此可将它们组合为言语量表分、操作量表分和全量表分，以估计被试的言语智商 (VIQ)、操作智商 (PIQ) 和总智商 (FIQ)。

当一个测验有几个分测验时，如果整个测验只有一个总的信度估计，绝不能认为分测验的分数将与合成分数一样地可靠。因为信度与测验长度有关，分测验分数几乎可以肯定不如合成分数可靠。因此，测验使用者在使用分测验分数时必须查看每一个分测验是否有信度估计，若没有这方面的资料，从分测验的得分作推论就会出现误断。