



心理
测量学

PSYCHOLOGICAL
TESTING

主讲：黎 坚 （副教授）
单 位：北京师范大学心理学院
信 箱：JianLi@bnu.edu.cn

School of Psychology
Beijing Normal University



第四章 测验的信度（reliability）

- ❶第一节 测量误差与经典测验理论
- ❷第二节 信度的概念
- ❸第三节 信度的估计方法
- ❹第四节 信度与测验分数的解释
- ❺第五节 影响信度的因素

Li, J. School of Psychology, Beijing Normal University

第一节 测量误差与经典测验理论

- ❶1.1 测量误差及其种类
- ❷1.2 测量误差的来源
- ❸1.3 经典测验理论

3

Li, J. School of Psychology, Beijing Normal University

1.1 测量误差及其种类

- ❶误差的含义
 - ↳在测量中与目的无关的因素所引起的不准确或不一致效应。

4

Li, J. School of Psychology, Beijing Normal University

测量的误差的种类

- ❶随机误差（random error）：是由与测量无关的变量引起的一种无规律的误差，不易控制。
- ❷系统误差（system error）：由与测量无关的变量引起的一种恒定而有规律的误差。

5

Li, J. School of Psychology, Beijing Normal University

❶两种误差的关系

- ↳系统误差只影响测值的准确性，与效度有关。
- ↳随机误差既影响准确性，又影响一致性。与效度和信度都有关。

6

Li, J. School of Psychology, Beijing Normal University

1.2 测量误差的来源

① 测验内部

- ① 测验项目取样的代表性
- ① 题型的猜测性：测量工具的设计及形式
- ① 表述的模糊性：语义不清、编排不良等
- ① 测题的难度过高或过低
- ① 测题的指导语用词不当
- ① 测验时限过短

7

Li, J. School of Psychology, Beijing Normal University

① 施测、评分与解释过程

- ① 测验环境
- ① 主试因素
 - ② 主试的解释、重述、语气、态度、肢体动作、反馈
- ① 意外干扰
- ① 评分计分
 - ② 计分错误、数据的分析处理不当
 - ② 主观题评分、光环效应

8

Li, J. School of Psychology, Beijing Normal University

① 受测者本身

- ① 应试动机
- ① 测验焦虑
- ① 练习效应
- ① 测验经验
- ① 反应倾向
- ① 生理变异

9

Li, J. School of Psychology, Beijing Normal University

1.3 经典测验理论 (CTT)

(Classic Test Theory)

① 也称为真分数理论 (True Score Theory)

$$\text{测验分数} = \text{真分数} + \text{误差} \quad (X = T + E)$$

- ① 真分数 (True Score, T)：受测者不受任何随机因素影响的条件下应该得到的测验分数，不可直接测量；在统计上用无数次测量的平均值代替。
- ① 误差 (measurement error, E)：即随机误差，指引起测量结果不一致的偶然误差，可正可负。

10

Li, J. School of Psychology, Beijing Normal University

CTT：真分数的数学模型

① 基本模型：测验分数是真分数与随机误差分数之和

$$X = T + E$$

① 引申定理一：若某种构想可以用平行测验反复测量足够多次，则观察分数的平均值接近于真分数

$$\varepsilon(X) = T \quad \varepsilon(E) = 0$$

① 引申定理二：真分数与测量误差的相关为0

$$\rho(T, E) = 0$$

11

Li, J. School of Psychology, Beijing Normal University

① 引申定理三：不同测量（例如平行测验）误差之间，以及测量误差与被测构想外其它变量之间的相关为0。

$$\rho(E_1, E_2) = 0 \quad \rho(E_1, T_2) = 0$$

① 引申定理四：测验分数的方差等于真分数方差和误差方差之和。

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

12

Li, J. School of Psychology, Beijing Normal University

CTT的重要概念：平行测验 (Parallel Test)

①能以相同的程度测量同一心理特质的测验。

↳ 这些测验测量同一构想

↳ 题目形式、数量、难度、区分度以及分数分布（平均数&标准差）都一致。平行测验既可以完全相同（重复施测），也可能有某种差异。

↳ 数学表达：

$$X = T + E, \quad X' = T + E'$$

以及 $\sigma^2(E) = \sigma^2(E')$

13

Li, J. School of Psychology, Beijing Normal University

CTT的三个基本假设

①真分数具有不变性

↳ 所测构想必须具有某种程度的稳定性。

①误差是完全随机的。

↳ 测量误差是平均数为零的正态随机变量。

↳ 测量误差分数与所测的构想即真分数之间相互独立。并且测量误差之间，测量误差与所测特质外其它变量间，也相互独立。

①观测分数是真分数与误差分数的和。

14

Li, J. School of Psychology, Beijing Normal University

测验分数方差/变异数 (Variance) 的分解

$$\sigma S_X^2 = S_T^2 + S_E^2$$

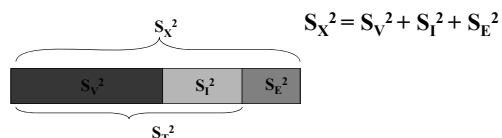
↳ 测验分数的方差 = 真分数的方差 + 随机误差的方差

↳ 系统误差呢？

$$\sigma S_T^2 = S_V^2 + S_I^2$$

↳ 系统误差方差隐含在真分数方差中

↳ 真分数的方差 = 潜在真分数的方差 + 系统误差的方差



15

Li, J. School of Psychology, Beijing Normal University

第二节 信度的概念

2.1 信度的理论定义

2.2 信度的测量定义

2.3 信度的操作定义

2.4 信度的性质

16

Li, J. School of Psychology, Beijing Normal University

2.1 信度的理论定义

①信度是指测量的可靠性 (reliability)，即用同一测验（或不同的平行测验）多次测量同一受测者，所得结果的一致性程度。

↳ 在不同的时间或情境，或有不同的评分者、施测者，所得的分数应该接近。

①信度只受随机误差影响，多次测量分数越一致，受误差的影响越小。所以信度也说明了测验分数中测量误差所占的成分。

17

Li, J. School of Psychology, Beijing Normal University

2.2 信度的测量定义

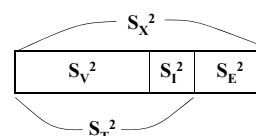
①测量理论中，信度被定义为：一组测量分数的真分数方差与总方差（实得方差）的比率，即：

$$r_{XX} = \frac{S_T^2}{S_X^2}$$

① r_{XX} ：信度系数

① S_T^2 ：真分数的方差

① S_X^2 ：测验分数的方差



18

Li, J. School of Psychology, Beijing Normal University



①由信度系数也可以推知误差造成的方差在总方差中所占百分比，例如：

↳当 $r_{XX}=0.90$ 时，可以认为实得分数中有90%的方差是真分数造成的，仅10%来自测量误差；

↳ $r_{XX}=1.00$ ，表示完全没有测量误差，所有的方差均来自真分数。

19

Li, J. School of Psychology, Beijing Normal University



①此外，信度还可以解释为测验分数与真分数之间的联系程度，以测验分数与真分数之间相关系数的平方来表示：

$$r_{XX}^2 = r_{XT}^2$$

① r_{XX} 被称为信度系数，而 r_{XT} 被称为信度指数

20

Li, J. School of Psychology, Beijing Normal University



信度与测量标准误

①测量标准误（standard error of measurement, SEM）：当某受测者完成一大批平行测验后，我们可以求得这一组测验分数的平均数和标准差（即测量误差的标准差），平均数即为受测者的真分数，而标准差则为测量误差大小的衡量指标，也就是标准误。

①标准误也能够作为测量的信度指标：

↳信度越高，标准误越小；反之亦然

21

Li, J. School of Psychology, Beijing Normal University



①由于 $S_X^2 = S_T^2 + S_E^2$

$$r_{XX} = \frac{S_T^2}{S_X^2} = \frac{S_X^2 - S_E^2}{S_X^2} = 1 - \frac{S_E^2}{S_X^2}$$

$$\therefore \text{测量标准误 } SE = S_E = S_X \sqrt{1 - r_{XX}}$$

22

Li, J. School of Psychology, Beijing Normal University



2.3 信度的操作定义

信度的测量定义逻辑清楚，易于理解，但...

信度的操作定义：两平行测验上测验分数间的相关，即用在一个平行测验上某受测者的测验分数，去正确推论另一平行测验上该受测者测验分数的能力。

$$r_{XX} = r_{XX'}$$

23

Li, J. School of Psychology, Beijing Normal University



2.4 信度的性质

①真分数的方差不能直接测量，是理论上构想的概念，因此信度只能根据一组实得分数得出估计。

①信度在0（不可信）与1（完全可信）之间。

①信度是一组测验分数的特性，而非个人分数的特性。

①信度不是一种普遍的特质，所以，测验分数不可能在任何情境下都是可靠的或一致的。

①信度越高越好，但是没有测验是完全可靠的，因为测量误差是无法避免的。

24

Li, J. School of Psychology, Beijing Normal University



①对测验信度的最低要求视使用的目的和情境而定

↳从测验类型方面考虑

①标准化能力和成就测验 ≥ 0.9

①人格、兴趣、态度、价值观 $0.80 \sim 0.85$

↳从功能方面考虑

①若用于比较两个群体平均水平的差异，则测验信度的最低要求应该 $0.6 \sim 0.7$

①若用于解释个体间的差异，则信度应该 ≥ 0.85

(Aiken, 2003)

25

Li, J. School of Psychology, Beijing Normal University



①一般而言，心理与教育测验的信度比自然科学的低。

↳自然科学的特征通常可以直接测量，但心理与教育测验只能间接测量而推论之。

↳自然科学的工具精确性相当高。

↳自然科学测量的特征相当稳定，不像心理特征变化很大。

26

Li, J. School of Psychology, Beijing Normal University



第三节 信度的估计方法

①注意：信度只是一个理论概念。由于真分数无法直接得到，所以，我们只能对之进行估计。

①估计信度的方法取决于你假定的误差来源是什么

误差来源	信度类型
时间	重测信度（稳定性）
内容	复本信度 / 分半信度 / Alpha系数 (等值性) (等值性) (同质性)
评价者/观察者	评定之间的相关

27

Li, J. School of Psychology, Beijing Normal University



第三节 信度的估计方法

3.1 重测信度

3.2 复本信度

3.3 内部一致性信度

3.4 评分者信度

28

Li, J. School of Psychology, Beijing Normal University



3.1 重测信度 (test-retest reliability)

①定义：同一量表，同一受测者群体，在不同时间，两次施测，求其相关。

$$r_{xx} = \frac{\sum (X_1 - \bar{X}_1)(X_2 - \bar{X}_2)}{N \cdot S_1 \cdot S_2}$$

①一般用积差相关系数表示， X_1 、 X_2 表示配对的原始分

①实质：表示测验结果的稳定性。故称之为稳定性系数 (Coefficient of Stability)

①形式：施测—适当时间—再施测

29

Li, J. School of Psychology, Beijing Normal University



重测信度的优缺点

①优点

↳只需一个测验；能提供测验结果是否随时间而变化的信息，可作为受测者将来行为表现的依据。

①缺点

↳真分数可能会发生改变

①相隔时间太长，则身心的发展与学习经验的积累等足以改变测验分数的意义，而使相关降低。

↳易受练习和记忆的影响。

①相隔时间太短，则记忆犹新，练习的影响较大。

30

Li, J. School of Psychology, Beijing Normal University

重测的测验时间间隔



- ① 间隔时间越长，稳定性系数越低。适宜时间间隔依照测验目的、性质及受测者特点而定。一般2~4周为宜。
- ① 年幼儿童，间隔要小；年长群体，间隔可大。
- ① 智力测验的间隔不能太短，成就测验的间隔不能太长。
- ① 最长间隔时间一般不超过六个月。（即不能让受测者记住上一次测验的内容，又不能让其特质发生变化，或对所学知识产生遗忘）

31

Li, J. School of Psychology, Beijing Normal University

3.2 复本信度 (alternate-forms reliability)



- ① 因为任何测验只是所有可能题目中的一份取样（行为样本），所以可编制许多平行的等值测验，叫做复本。
- ① 复本信度是指以两个测验复本来测量同一群体，然后求受测者群体在两个测验上得分的相关系数。

$$r_{xx} = \frac{\sum (X_1 - \bar{X}_1)(X_2 - \bar{X}_2)}{N \cdot S_1 \cdot S_2}$$

32

Li, J. School of Psychology, Beijing Normal University

复本等值要符合的条件



- ① 测量的是同一种心理特性。
- ① 具有相同的内容和形式，但题目不应重复。
- ① 题目数量相等，难度和区分度大体相同。
- ① 分数分布（平均数和标准差）大致相等。

33

Li, J. School of Psychology, Beijing Normal University

复本施测的形式



- ① 复本测验可同时连续实施或相距一段时间分两次实施
 - ② 同时连续实施的复本信度又称等值系数 (Coefficient of Equivalence)
 - ② 相距一段时间分两次实施的复本信度又称稳定和等值系数 (Coefficient of Stability and Equivalence)
 - ② 包含了内容和时间变异所造成的误差，更为严格

34

Li, J. School of Psychology, Beijing Normal University

复本信度的优缺点



- ① 优点：
 - ② 能够较好地避免记忆效果、学习效应等。
- ① 缺点
 - ② 有时复本只能降低练习或记忆的影响；
 - ② 从A本到B本，不同受测者的迁移能力可能不同；
 - ② 建立复本费时费力，十分困难。

35

Li, J. School of Psychology, Beijing Normal University

3.3 内部一致性信度 (internal consistency reliability)



- ① 内部一致性信度系数主要反映的是题目之间的关系，表示测验能够测量相同内容或特质的程度。包括两种：
 - ② 3.3.1 分半信度
 - ② 3.3.2 同质性信度

36

Li, J. School of Psychology, Beijing Normal University

3.3.1 分半信度 (split-half reliability)

①在测验没有复本且只能实施一次的情况下, 可将测验项目分成对等的两半, 根据受测者在这两半测验中所得的分数计算相关系数, 即为分半信度。

①计算分半信度先要对测验分半。不同的分半法可能会得到不同的信度值。

②怎么分?

37

Li, J. School of Psychology, Beijing Normal University

①为了使两半基本等值, 可将项目按由易到难的顺序排列编号, 然后按奇数和偶数序号将项目分半。

②应该使那些性质相同、联系紧密的项目分在相同的一半, 否则会使信度值偏高。

①分半信度实际上只是半个测验的信度, 由于测验越长、项目越多, 测验的信度越高, 因此获得整个测验的信度系数需要用校正公式。

38

Li, J. School of Psychology, Beijing Normal University

如何用分半法估计整个测验的信度

1. 采用积差相关计算分半信度

2. 用斯皮尔曼-布朗公式校正

$$r_{XX} = \frac{2r_{x1x2}}{1 + r_{x1x2}}$$

r_{x1x2} 表示两个半测验分数的相关系数

39

Li, J. School of Psychology, Beijing Normal University

①如果两半测验不等值, 亦即两半测验分数没有相同的平均数和标准差, 则需采用其它公式:

②弗朗纳根公式:

$$(Flanagan) \quad r = 2\left(1 - \frac{S_a^2 + S_b^2}{S_x^2}\right)$$

③ S_a^2 和 S_b^2 : 两个半测验分数的方差

③ S_x^2 : 测验总分的方差。

40

Li, J. School of Psychology, Beijing Normal University

②卢伦 (Rulon) 公式:

$$r = 1 - \frac{S_d^2}{S_x^2}$$

S_d^2 : 两半测验分数之差的方差,

S_x^2 : 测验总分的方差。

41

Li, J. School of Psychology, Beijing Normal University

例题

学生	奇数题A	偶数题B	总分A+B	差异分数A-B
A	11	8	19	3
B	10	6	16	4
C	6	2	8	4
D	5	1	6	4
E	12	5	17	7
F	4	1	5	3
G	4	4	8	0
H	8	6	14	2
I	8	5	13	3
J	2	2	4	0
求和	70	40	110	30

42

Li, J. School of Psychology, Beijing Normal University

$$S^2 = \frac{\sum X^2}{n} - \left(\frac{\sum X}{n} \right)^2$$

已知两个半测验的Pearson积差相关为0.79，请计算校正后的全测验信度。

↳ 斯-布公式

↳ 卢伦公式： $1 - (3.8/26.6) = 0.86$

↳ 弗朗纳根公式： $2[1 - (10 + 5.2)/26.6] = 0.86$

43

Li, J. School of Psychology, Beijing Normal University

分半信度的优缺点

① 优点：

↳ 只需一份测验；不受练习和记忆的影响。

① 缺点

↳ 不同的分半方法所得结果不同；

↳ 假设两个半测验是平行测验，但有时很难实现；

↳ 有些测验本身是一个整体，很难拆解为两半；

↳ 不适用于速度测验。

44

Li, J. School of Psychology, Beijing Normal University

3.3.2 同质性信度 (homogeneity reliability)

① 同质性指测验的所有题目测量的是同一种心理特质，表现为各个题目得分之间具有较高的相关，相关越高则同质性越强。

① 同质性也是反映测验信度的指标

↳ 库德-理查逊公式（仅适用于0/1记分）

(Kuder-Richardson formula 20, 21)

↳ 克伦巴赫 α 系数（适用于任何测量尺度）

(Cronbach, 1951)

45

Li, J. School of Psychology, Beijing Normal University

同质性信度的估计方法

↳ 库德-理查逊公式（K-R20公式）

$$r_{XX'} = \frac{k}{k-1} \left(1 - \frac{\sum p_i q_i}{S_X^2} \right)$$

① K：构成测验的题目数

① p_i ：通过第*i*题的人数比例

① q_i ：未通过第*i*题的人数比例，即 $1 - p_i$

① S_X^2 ：测验总分的方差

46

Li, J. School of Psychology, Beijing Normal University

同质性信度的估计方法

↳ 库德-理查逊公式（K-R21公式）

$$r_{XX'} = \frac{k}{k-1} \left(1 - \frac{k \bar{p}_i \bar{q}_i}{S_X^2} \right) = \frac{k S_X^2 - \bar{X}(k - \bar{X})}{(k-1) S_X^2}$$

① 区别：KR21假设所有题目具有相等的难度，更保守的估计

$$\frac{k}{k-1} \left(1 - \frac{k \bar{p}_i \bar{q}_i}{S_X^2} \right) \rightarrow \bar{p}_i = \frac{\bar{X}}{k}$$

47

Li, J. School of Psychology, Beijing Normal University

① 例：一个测验由75个判断题构成，通过对某中学高一年级施测发现测验平均分为50，标准差为10，假设所有测验项目的难度相同，求测验的信度。

48

Li, J. School of Psychology, Beijing Normal University

同质性信度的估计方法

☞ 克伦巴赫 α (Alpha) 系数

☉ α 系数是经斯皮尔曼-布郎公式校正过的K个单项目测验信度的平均值。

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum S_i^2}{S_x^2} \right)$$

☞ k: 测验的题目数

☞ S_i^2 : 题目i得分的方差

☞ S_x^2 : 整个测验得分的方差

49

Li, J. School of Psychology, Beijing Normal University

☉例: 已知测验包括五个题目, 各题方差分别为: 0.80、0.81、0.79、0.78、0.82, 总分方差为15, 求 α 系数。

50

Li, J. School of Psychology, Beijing Normal University

内部一致性信度的特点

☉速度测验

☞ 所有内部一致性指标都高估速度测验的信度

☞ 最好用重测信度和复本信度

☉一般0.8以上才算好, 但应视测量性质、目的而定。

☉实际应用中, α 值至少要大于0.5, 最好能大于0.7。

(Nunnally, 1978)

51

Li, J. School of Psychology, Beijing Normal University

3.4 评分者信度 (inter-scorer reliability)

☉评分者信度是指不同评分者之间在测验结果计分上的一致性。

☉在心理测验中, 评分者信度的计算, 通常是随机抽取若干份试卷, 由至少两位受过训练的评分者按计分规则分别判分, 然后计算它们的相关。几个评分者的评分越一致, 评分者信度越高。

52

Li, J. School of Psychology, Beijing Normal University

☉适用情境:

☞ 对于无法以客观方式测评的工具, 如面试, 投射测验, 创造性思维测验、道德判断测验、论文式测验等, 评分者也是误差来源之一。

☉应注意避免评审之间的相互影响。

☉计算方式: 不同评分者间评分的相关程度

☉一般认为评分者之间平均一致性达0.90以上, 才认为评分是客观的。

53

Li, J. School of Psychology, Beijing Normal University

两位评分者时的信度计算方法

☉积差相关

☉斯皮尔曼等级相关 (Spearman rank correlation)

$$r_R = 1 - \frac{6 \sum (R_{xi} - R_{yi})^2}{n(n^2 - 1)} = 1 - \frac{6 \sum D^2}{n(n^2 - 1)}$$

公式中: R_x 为变量X的等级,

R_y 为变量Y的等级,

D为两变量的等级之差, 即 $D = R_x - R_y$

n为样本的容量

54

Li, J. School of Psychology, Beijing Normal University

例：两位面试考官对10位应聘者的面试打分

序号	R_1	R_2	$D=R_1-R_2$	D^2
1	1	2	-1	1
2	2	1	1	1
3	3	4	-1	1
4	4	5	-1	1
5	5	6	-1	1
6	6	7	-1	1
7	7	8	-1	1
8	8	3	5	25
9	9	10	-1	1
10	10	9	1	1
合计				34

55

Li, J. School of Psychology, Beijing Normal University

两位以上评分者时的信度计算方法

①如果评分者在两人以上，而且又采用等级计时，就需要用肯德尔和谐系数来求评分者信度。（Kendall coefficient of concordance）

$$r_w = \frac{\sum (R_i - \bar{R}_i)^2}{\frac{1}{12} K^2 (n^3 - n)}$$

n 是被评定的人数或作品数

$\sum R_i$ 是每个被评者的评价等级之和

56

Li, J. School of Psychology, Beijing Normal University

例：4位教师对6位学生作文竞赛的名次排列如表。问4位教师的评分者信度如何？

4位教师对6位学生作文竞赛的名次排列

学生 $n=6$	评价者 $K=4$			
	1	2	3	4
1	3	4	2	1
2	4	3	1	3
3	2	1	3	4
4	6	5	6	5
5	1	2	4	2
6	5	6	5	6

57

Li, J. School of Psychology, Beijing Normal University

学生 $n=6$	评价者 $K=4$				R_i	R_i^2
	1	2	3	4		
1	3	4	2	1	10	100
2	4	3	1	3	11	121
3	2	1	3	4	10	100
4	6	5	6	5	22	484
5	1	2	4	2	9	81
6	5	6	5	6	22	484
Σ					84	1370

58

Li, J. School of Psychology, Beijing Normal University

信度估计方法总结

- ①估计信度的方法远不止上面介绍的几种，实际上有多少误差的来源，便有多少估计信度的方法。
- ①考察测验的信度时，应根据情况采用不同的信度指标，原则上一个测验哪种误差大，便应该用哪种误差估计。
- ①有时一个测验需要有几信度系数，这样我们就能把总分数的方差分成不同的分支。

59

Li, J. School of Psychology, Beijing Normal University

实例分析

- ①假设对100个六年级学生以两个月的时间间隔先后施测一个创造力测验的A/B两个复本，所得的等值性与稳定性系数为0.70。我们还根据受测者对每个复本的反应计算出分半信度为0.80（先计算每个复本的分半相关系数。将二者平均后再用斯皮尔曼-布朗公式校正）。同时，我们让另一个评分者随机抽取50份卷子另外评分，得到评分者信度为0.92。然后，我们对这三种方法所产生的误差方差进行分析。

60

Li, J. School of Psychology, Beijing Normal University

测验的误差方差来源分析

信度类型	误差方差	误差方差来源
复本信度 (间隔施测)	$1-0.70=0.30$	时间与内容取样
分半信度	$1-0.80=0.20$	内容取样
上述二者差异	$.30-.20=0.10$	时间取样
评分者信度	$1-0.92=0.08$	评分者差异
误差方差总和	$0.20+0.10+0.08=0.38$	
真实方差	$1-0.38=0.62$	

61

Li, J. School of Psychology, Beijing Normal University

信度系数的类型	误差方差的来源	计算公式	施测次数	所需复本数目
重测信度	时间取样	积差相关系数	2	1
复本信度 (连续施测)	内容取样	积差相关系数	1	2
复本信度 (间隔施测)	时间取样、内容取样	积差相关系数	2	2
分半信度	内容取样	积差相关系数 (需校正, 常用斯-布公式)	1	1
同质性信度	内容的异质性	α 系数 (非二分法)	1	1
评分者信度	评分者间的差异	积差相关系数 斯皮尔曼等级相关系数, 肯德尔和谐系数	1	1

第四节 信度与测验分数的解释

04.1 解释个人分数的意义

04.2 比较不同测验分数的差异

04.3 信度与相关校正

63

Li, J. School of Psychology, Beijing Normal University

4.1 解释个人分数的意义

0 信度系数表明一个测验有多可信

0 测量标准误 (Standard Error of Measurement) 告诉我们一个人的分数究竟有多可信 (信度的另一种表达方式)

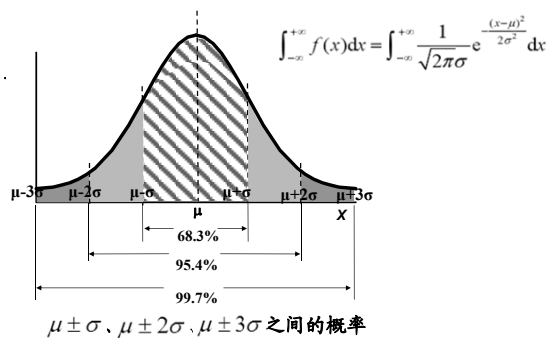
$$SE = S_x \sqrt{1 - r_{xx}}$$

0 测量标准误的应用: 估计真分数的范围

64

Li, J. School of Psychology, Beijing Normal University

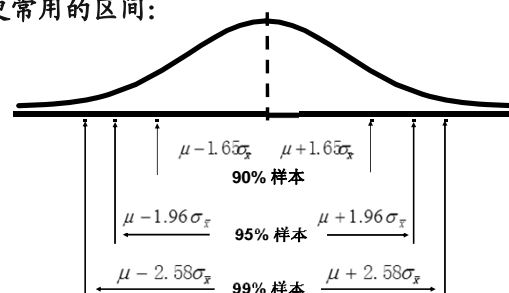
正态分布的特殊概率值



Li, J. School of Psychology, Beijing Normal University

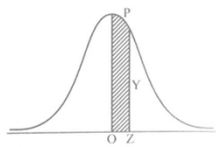
正态分布的特殊概率值

• 更常用的区间:



Li, J. School of Psychology, Beijing Normal University

附录

附表1 正态分布表
(曲线下的面积与纵高)


Z	Y	P	Z	Y	P	Z	Y	P
0.00	0.39894	0.00000	0.30	0.38139	0.11791	0.60	0.33322	0.22575
0.01	0.39892	0.00399	0.31	0.38023	0.12172	0.61	0.33121	0.22907
0.02	0.39886	0.00798	0.32	0.37903	0.12552	0.62	0.32918	0.23237
0.03	0.39876	0.01197	0.33	0.37780	0.12930	0.63	0.32713	0.23565
0.04	0.39862	0.01595	0.34	0.37654	0.13307	0.64	0.32506	0.23891
0.05	0.39844	0.01994	0.35	0.37524	0.13683	0.65	0.32297	0.24215
0.06	0.39822	0.02392	0.36	0.37391	0.14058	0.66	0.32086	0.24537
0.07	0.39797	0.02790	0.37	0.37255	0.14431	0.67	0.31874	0.24857
0.08	0.39767	0.03188	0.38	0.37115	0.14803	0.68	0.31659	0.25175
0.09	0.39733	0.03586	0.39	0.36973	0.15173	0.69	0.31443	0.25490
0.10	0.39695	0.03983	0.40	0.36827	0.15542	0.70	0.31225	0.25804
0.11	0.39654	0.04380	0.41	0.36678	0.15910	0.71	0.31006	0.26115
0.12	0.39608	0.04776	0.42	0.36526	0.16276	0.72	0.30785	0.26424
0.13	0.39559	0.05172	0.43	0.36371	0.16640	0.73	0.30563	0.26730
0.14	0.39505	0.05567	0.44	0.36213	0.17003	0.74	0.30339	0.27035
0.15	0.39448	0.05962	0.45	0.36053	0.17364	0.75	0.30114	0.27337

67

①测量误差 (X-T) 被假设呈正态分布, 因此可以根据测验分数, 结合测量标准误与正态分布表 (Z值表), 对真分数进行区间估计。

①一般将置信度设定为0.95 (即犯错误的概率—显著性为0.05), 则置信区间 (credibility interval)

$$-1.96SE \leq X - T \leq 1.96SE$$



$$X - 1.96SE \leq T \leq X + 1.96SE$$

68

Li, J. School of Psychology, Beijing Normal University

①例: 一个测验平均值=50, $S_x=10$, $r_{xx}=0.84$, 那么

$$SE = 10\sqrt{1-0.84} = 4$$

②若某生在该测验上得了35分, 那么, 有95%的可能其真分数落在下列范围中

$$27.16 \sim 42.84$$

69

Li, J. School of Psychology, Beijing Normal University

练一练

①在一个智力测验中, 某受测者的测验得分为120分, 请估计其真分数范围? (置信度设为0.95, 已知该智力测验的标准差为15, 信度系数为0.95)

②解:

70

Li, J. School of Psychology, Beijing Normal University

wisc-IV
韦氏儿童智力量表-第四版 (中文版)

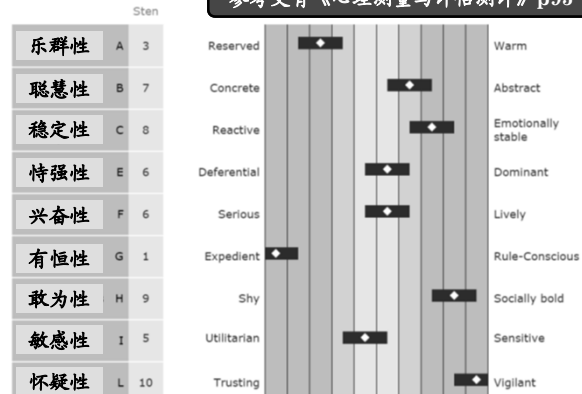
量表分数总和与合成分数转换表

量表	量表分数总和	合成分数	百分等级	95% 置信区间
言语理解	51	言语理解指数: 144	99.8	135-148
知觉推理	50	知觉推理指数: 143	99.8	132-147
工作记忆	22	工作记忆指数: 106	66	99-112
加工速度	26	加工速度指数: 116	86	105-123
全量表	149	总智商: 139	99.5	133-143

71

Li, J. School of Psychology, Beijing Normal University

参考艾肯《心理测量与评估测评》p95



如果.....

❶如果你有一个测验,

$S_X=10, r_{XX}=1$, 那么

$SE=10\sqrt{1-1}=0$

❷即没有误差!

❸所以, 如果某生在这个测验上得了45分, 那么他的真分数就是45分。

❶如果你有一个测验,

$S_X=10, r_{XX}=0$, 那么

$SE=10\sqrt{1-0}=10$

❷整个世界只剩误差!

❸完全不可信的度量, 随机分数

❹测验的信度越高, 一个人的真分数估计越接近测验分数!

73

Li, J. School of Psychology, Beijing Normal University

4.4 比较不同测验分数的差异

❶测量标准误和测验信度在评价两个测验的分数是否有明显差异时也非常重要。

❷两个人在同一测验上分数的差异

❸两个人在不同测验上分数的差别

❹同一受测者在不同测验上的差别

❺测验的可信性将影响我们对分数差异的解释

74

Li, J. School of Psychology, Beijing Normal University

差异分数的信度

❶差异分数的信度估计(Gulliksen, 1950)

$$r_{diff} = \frac{\frac{r_{XX} + r_{YY} - r_{XY}}{2} - r_{XY}}{1 - r_{XY}}$$

❷ r_{XX} 和 r_{YY} : 测验X、Y的信度

❸ r_{XY} : 测验X和Y的相关

❹例: 两个测验X和Y的信度分别为.84与.76, 两个测验间的相关为.70, 则 r_{diff} 为.33。

❺差异分数的信度远低于各测验的信度!

75

Li, J. School of Psychology, Beijing Normal University

差异分数的标准误 (Standard Error of Difference)

❶公式

$$SE_d = \sqrt{SE_X^2 + SE_Y^2} = \sqrt{(s\sqrt{1-r_{XX}})^2 + (s\sqrt{1-r_{YY}})^2} = s\sqrt{2-r_{XX}-r_{YY}}$$

❷ r_{XX} 和 r_{YY} : 测验X、Y的信度

❸S: 将X和Y测验分数转换到相同尺度后的标准差

❹注意: 来自不同测验的原始分数无法直接比较, 必须转换为相同尺度的标准分, 例如T分数, 才能比较

76

Li, J. School of Psychology, Beijing Normal University

例

❶你做了一个职业兴趣测验, 在六个度量(维度)上的得分如下:

❷Realistic=65

❸Investigative=67

❹Artistic=70

❺Social=75

❻Enterprising=68

❼Conventional=61

❶问: 你的S分数与A分数是否有显著差异?

77

4.5 相关的校正

(Correcting Correlations for Attenuation)

❶我们往往更想估计两个构想的真正相关, 而不是由于不可信的测量而得到的低估了的相关。

❷可以运用校正公式估计在完美测量的情况下两个构想的相关性(理论关系)

$$r_c = \frac{r_{XY}}{\sqrt{r_{XX} \times r_{YY}}}$$

78

Li, J. School of Psychology, Beijing Normal University



相关校正的实例:

✎ Spellman博士声称, 他的“街头智慧” (street-smart) 测验测量了某些与传统IQ测验所测量的不同的东西, 理由是在一个大而广泛的高中学生样本中 (N=132,214), SS与IQ测验分数的相关相当低, 只有.47。

✎ 这一结论合理吗?

79

Li, J. School of Psychology, Beijing Normal University



	SS	IQ
SS	.37	
IQ	.47	.85

$$r_c = \frac{r_{XY}}{\sqrt{r_{XX} \times r_{YY}}} = \frac{0.47}{\sqrt{0.37 \times 0.85}} = 0.84$$

✎ 估计真正相关的最大值 (.84) 远高于观测的相关 (.47), 即SS与IQ在理论上最高相关值远大于最初的设想!

80

Li, J. School of Psychology, Beijing Normal University



第五节 影响信度的因素

05.1 测验长度

05.2 样本特征

05.3 测验难度

05.4 时间间隔

81

Li, J. School of Psychology, Beijing Normal University



5.1 测验长度

✎ 测验长度, 即测验的题目数量, 也是影响信度系数的重要因素。一般来说, 在一个测验中增加同质的题目, 可以使信度提高, 原因在于:

✎ 测验越长, 测验的测题取样或内容取样就越有代表性。

✎ 测验越长, 受测者的猜测因素影响就越小。

✎ 注意, 过度增长测验反而会引起受测者的疲劳和反感, 降低可靠性。

82

Li, J. School of Psychology, Beijing Normal University



测验长度与测验信度的关系

✎ 可由斯皮尔曼-布朗公式: $r_{kk} = \frac{K \times r_{XX}}{1 + (K-1)r_{XX}}$

✎ 导出: $K = \frac{r_{kk}(1-r_{XX})}{r_{XX}(1-r_{kk})}$

✎ K : 改变后的测验与原测验长度之比

✎ r_{XX} : 原始测验信度

✎ r_{kk} : 长度变为K倍后的测验信度

83

Li, J. School of Psychology, Beijing Normal University

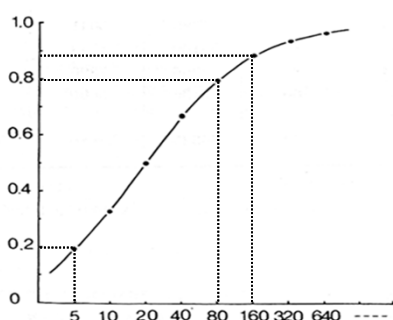


✎ 例: 一个包含40个题目的测验信度为0.80, 欲将信度提高到0.90, 问至少需要增加多少题目?

84

Li, J. School of Psychology, Beijing Normal University

边际效应



85

Li, J. School of Psychology, Beijing Normal University

5.2 样本特征

① 信度常用信度系数来表示，信度系数是相关系数，相关系数受样本的异质性及样本团体平均能力水平的影响。

② 相关系数受样本分数分布范围的影响，而分数范围与样本团体的异质程度有关。若获得信度的取样团体较为异质的话，往往会高估测验的信度，相反则会低估测验的信度。

③ 对于不同能力水平的团体，同样的题目有不同的难度。

86

Li, J. School of Psychology, Beijing Normal University

5.3 测验难度

① 一般的能力测验最适合用来测量中等能力水平的受测者，而对较高/较低水平的受测者可能较易/难，使得分数分布范围缩小，信度水平降低。

② 对于过难的题目，受测者只能通过猜测作答；对于过易的题目，受测者很可能由于粗心大意而失分。

③ 一个标准化的测验应根据不同能力水平报告测验的信度，以作为选择测验的参考！

87

Li, J. School of Psychology, Beijing Normal University

5.4 时间间隔

① 时间间隔只对重测信度和不同时测量时的复本信度（重测复本信度）有影响，对其余的信度来说不存在时间间隔问题。

② 两次测验间隔时间过短：

▮ 练习效应大，记忆效果，信度提高

③ 两次测验间隔时间过长：

▮ 自然发展，教育经验，信度降低

88

Li, J. School of Psychology, Beijing Normal University



电 邮: JianLi@bnu.edu.cn

办公室: 后主楼1516



Thank You !

School of Psychology
Beijing Normal University