

信度的理论基础——经典测量理论

Classic test theory

1. 心理特质
2. 误差
3. 真分数
4. 数学模型
5. 评价

1.1 心理特质

- **心理特质 (trait) :**
 - 是表现在一个人身上所特有的相对稳定的行为方式
 - 1. 一组内部相关行为的概括，如善良、聪明
 - 2. 对不同刺激做相同反应

3. 较稳定

4. 通过特质可以预测行为

5. 可分为多个层次

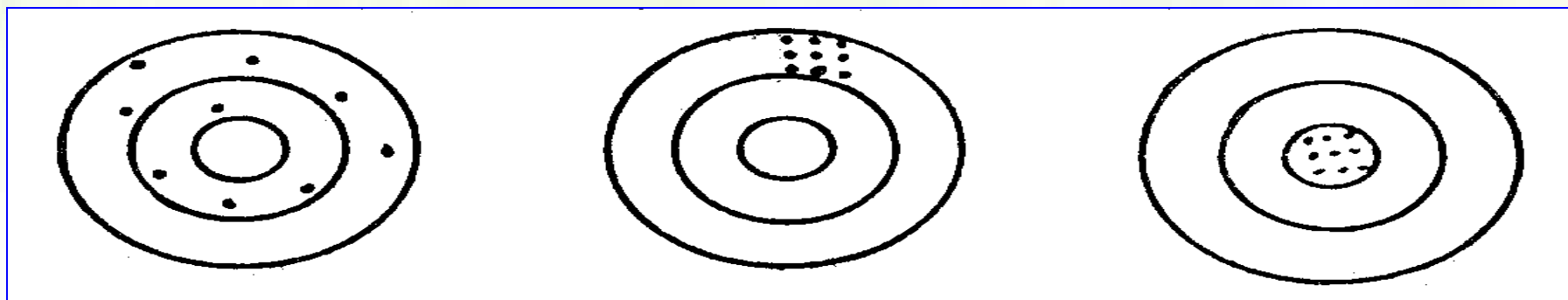
1.2 心理特质的可测性假设

- 凡客观存在的事物都有其数量。
- 凡有数量的东西都可以测量。

2.1 测量误差

- 测量误差（error）
 - 测量值与实际值的差异
 - 测量过程中产生的由那些与测量目的无关的因素所引起的一种不一致或不准确的测量效应。

2.2 测量误差的种类



- **系统误差**（system）：恒定而有规律。通常是由**测量工具**本身引起
- **随机误差**（random）：由不稳定的偶然因素引起的、又不易控制的误差
- 两种误差对测量的结果有不同的影响

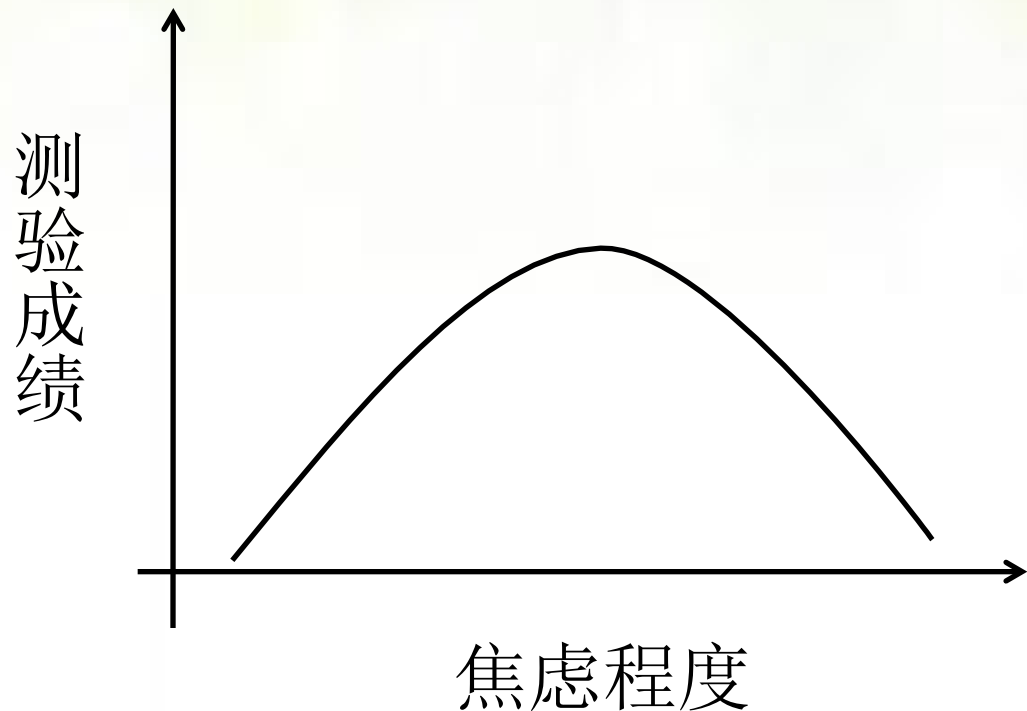
2.3 测量误差的来源-1

- **测量工具**（测验内部）引起的误差
 - 题目取样
 - 指导语
 - 难度
 - 时限
 - 测验复本不等值

2.3 测量误差的来源-2

- 由**测量对象**引起的误差

- 测验的经验
- 练习因素
- 应试动机
- 测验焦虑
- 反应定势
- 生理因素



2.3 测量误差的来源-3

- 由施测过程引起的误差
 - 物理环境
 - 主试者方面
 - 意外干扰
 - 评分计分

3 真分数

- **真分数**（True Score）：一个测量工具在没有测量误差时，所得到的纯正值。
反映被试某种心理特质真正水平的数值。
操作定义：经过无数次测量所得到的均值
- **观察分数**（Observed Score）：测某特质时得到的实测分数。

4.1 经典测验理论

- CTT数学模型

$$X = T + E$$

- X ——实得分数
- 真分数True Score ——稳定
- 测量误差Errors of measurement ——完全随机

4.2 基本假设

- 假设 (Assumptions)

- 真分数与观测分数的平均数相等。

$$\bar{T} = \bar{X}$$

- 误差分数的期望值为0

$$\varepsilon E = 0$$

4.3 真分数理论的基本定理

- 真分数与测量误差的相关为0
(uncorrelated)。

$$\rho_{TE} = 0$$

- 各次测验上的误差分数之间的相关为0。

$$\rho_{E_1E_2} = 0$$

数学模型的假设

可以从以下3个方面理解CTT数学模型的假设和公理：

其一，真分数具有不变性。实质是指反应个体某种心理特质水平的真分数是假定不会变的，研究者的任务就是估计真分数的大小。

其二，观测分数是真分数与误差分数的和。

$$\text{即 } X = T + E。$$

其三，误差是完全随机的。均值为0的正态分布；与真分数独立。

Q: 重复多次测量如何实现?

事实上，我们在实际研究中，并不是用许多平行测验来反复测查同一批被试，而是用同一个测验来同时测查很多被试。

每个人的误差都是随机的，并且服从均值为零的正态分布，所以当团体足够大时，团体内的误差也会互相抵消，整个团体的观测分数的均值会趋于该团体的真分数的平均数。

4.4 测量误差的指标： σ_{SEM}

- σ_{SEM} (Standard Error of Measurement)
- 标准误
- 样本平均数的标准差、或样本标准差的标准差、方差的标准差
- 衡量对应样本统计量抽样误差大小的尺度。

4.5 用方差对测验分数进行分解

- 测验观测分数的方差，等于真分数方差和误差方差之和

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

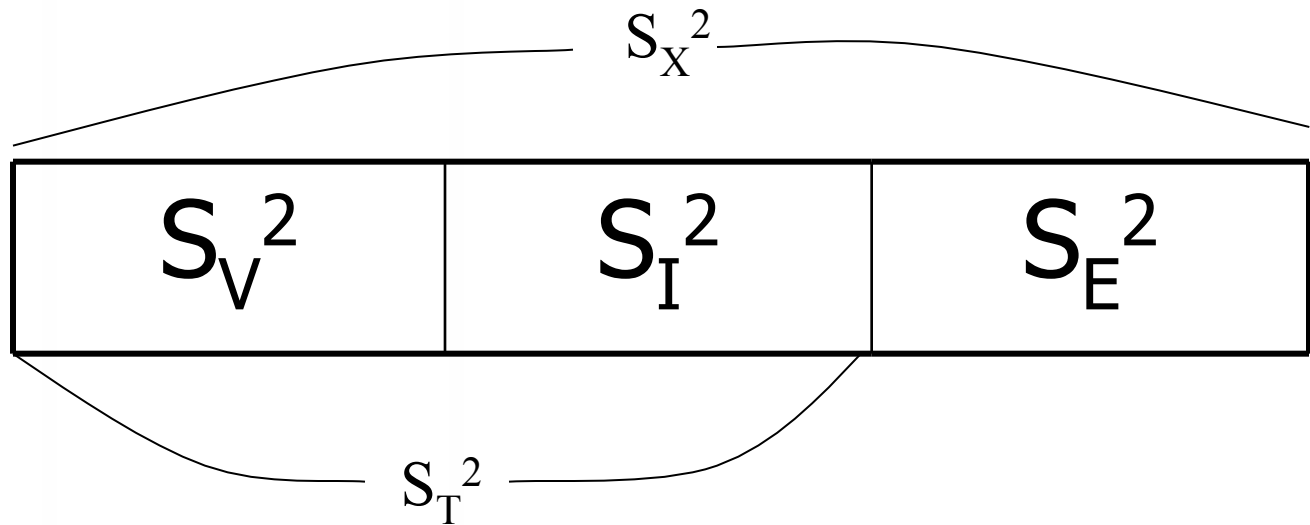
$$\begin{aligned} S^2_X &= \frac{\sum (X - \bar{X})^2}{N} = \frac{\sum (T + E - \bar{X})^2}{N} = \frac{\sum (T + E - \bar{T} - \bar{E})^2}{N} \\ &= \frac{\sum (t + e)^2}{N} = \frac{\sum t^2 + 2\sum te + \sum e^2}{N} = \frac{\sum t^2 + \sum e^2}{N} = S^2_T + S^2_E \end{aligned}$$

$$r_{TE} = \frac{\sum te}{NS_T S_X}$$

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

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

$$S_X^2 = S_V^2 + S_I^2 + S_E^2$$

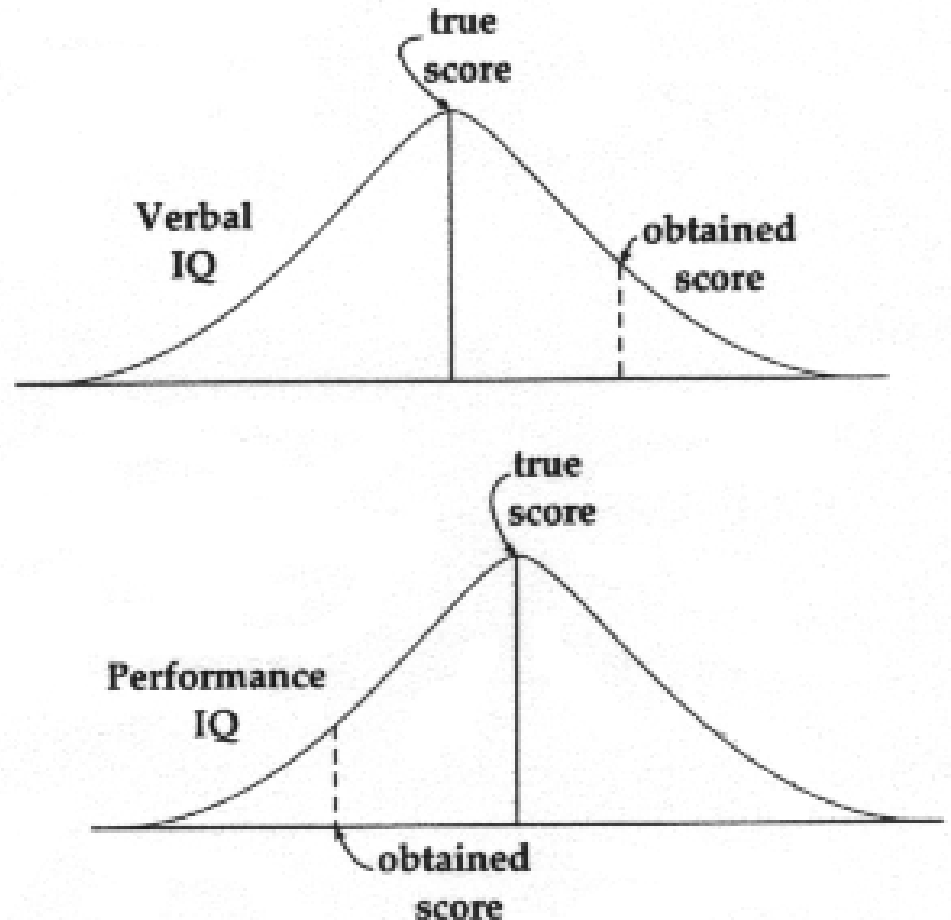


5.1 经典测量理论的地位

- 心理测量的理论与实践一直建立在真分数模型基础上。
- 以弱假设为基础，容易验证并易于被绝大多数测验数据资料所满足。
- 建立了一整套理论及统计分析方法。

5.2CTT的局限

- 观察分 X （以及 T ）不是等距量表。因而，无法直接比较两组测验的得分。



5.2 CTT的局限

- 信度、效度、项目特征等的统计分析，依分析时所使用的样本而异
- 已经指出测量误差的存在。但测验结果精度以信度和标准误来表示。忽略了单个被试在题项上的得分。

信度I

一、定义

二、类别和估计

一、 定义

- 信度： 测验结果的可靠性
(trustworthiness)、 一致性
(consistency)、 或稳定性
(stability)
 - 一致性—测题间是否相互符合
 - 稳定性—不同测验时间下， 测验分数的一致程度

三种等价的信度定义：

- 1. 一个被测团体的真分数与实得分数的变异数之比。

$$r = S^2_T / S^2_X$$

- 2. 一个被试团体的真分数与实得分数的相关系数的平方

$$r = \rho^2_{TX}$$

- 3. 一个测验（A卷）与它的任意一个“平行测验”（B卷）的相关系数。

$$r = \rho_{XX'}$$

2、误差与信度

- 随机误差：由概率因素所支配
- 系统误差：由测量工具本身引起的，恒定不变的误差。

信度受其影响吗？

- 信度受测量的随机误差影响，不受系统误差的影响。

3、变异与信度 (Variance and Reliability)

- $X=T+E$

- $S^2_X = S^2_T + S^2_E$

4、信度的理论模型

$$1 = S^2_T / S^2_X + S^2_E / S^2_X$$

$$r = S^2_T / S^2_X$$

or

$$r = 1 - S^2_E / S^2_X$$

二、测查信度的途径

- 重测信度 (Test-retest reliability)
- 复本信度 (Parallel forms reliability)
- 评分者信度 (Inter-scorer reliability)
- 分半信度 (Split-half reliability)
- 内部一致性信度 (Internal consistency reliability)

1、重测信度

• 定义

- 以同一种测量工具，对同一群被试，前后测验两次所得结果的一致性程度
- 又称稳定性系数

• 使用条件

- 特质是稳定的
- 练习和遗忘的效果基本上相互抵消
- 施测间隔内，所测特质没有获得更多的学习和训练

皮尔逊积差相关系数

Product-Moment correlation coefficient formula

$$r = \frac{N\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{\left[N\Sigma X^2 - (\Sigma X)^2 \right] \left[N\Sigma Y^2 - (\Sigma Y)^2 \right]}}$$

被试	第一次测验	第二次测验
A	70	77
B	80	72
C	75	83
...
...

重测信度评价

- 优点
 - 简单
 - 适用于稳定特质
- 不足
 - 成熟 (Maturation)
 - 练习效应 (Carryover effects)
 - 反应行为 (Reactivity)

3、复本信度

- 又称平行测量技术
- 同一群受测试者接受平行测验得分之相关系数——等值性系数

被试	测验1	测验 2
A	70	77
B	80	72
C	75	83
...
...

重测复本信度

（稳定性与等值性系数）

- 程序：同群体在不同时间接受平行测验
- 信度系数：两个分数间之相关
- 误差来源：时间取样与内容取样

复本信度评价

- 优点

- 可以克服成熟、反应、练习效应

- 缺点

- 编制不易

- 非平行量表问题 (lack of equivalence)

- 疲劳

- 成本较高

2 、评分者信度

- 定义

- 不同评分者对于评定物评分的一致性
- 是将评分转化为等级以后进行估算信度的方法。比较适合于主观题的评分。
(eg:作文)

评分者信度计算方法

- 2 raters, 评分者间分数的相关系数（积差相关或等级相关）
- more than 2 raters, 肯德尔和谐系数W
 - 若 $3 < K < 20$, $3 < N < 7$ 时, 直接查表, 实际计算值大于表中相应值时, 说明信度较高。
 - 若 $N > 7$, 用 χ^2 检验

4. 分半信度

- 将一个测验分成对等的两半后，受测者在这两半上所得分数的一致性程度
- **How to divide the test?**
 - Random (随机法)
 - odd-even (奇偶法)
 - 按内容或题目难度分半
 - 1st half of items & last half of items (顺序法) (难度、题型)

分半信度计算方法

[return](#)

- Spearman—Brown prophecy formula

$$r_{SB} = \frac{2r_{hh}}{1 + r_{hh}}$$

单	偶
35	36
40	38
...	...
...	...

在两半测验分数的变异数相等时才能使用。

分半信度计算方法

- 弗朗那根公式:

$$r = 2 [1 - (S_a^2 + S_b^2) / S_x^2]$$

- S_a^2 S_b^2 分别为所有被试在两半测验上得分的变异； S_x^2 为全体被试在整个测验上的总得分的变异。
- 卢仑公式:

$$r = 1 - S_d^2 / S_x^2$$

- S_d^2 是全体被试在两半测验上得分之差的变异。

分半信度评价

- 优点
 - 无须复本、无须多次测量
- 缺点
 - 不同折半方式影响信度
 - 题目缩短，信度下降

4 、内部一致性信度

- 内部一致性信度又叫同质性信度，是指测验内部所有题目间的一致性程度。
 - 分半信度（Split—half）——粗略估计
 - Kuder—Richardson
 - Coefficient alpha

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/315340023320011130>