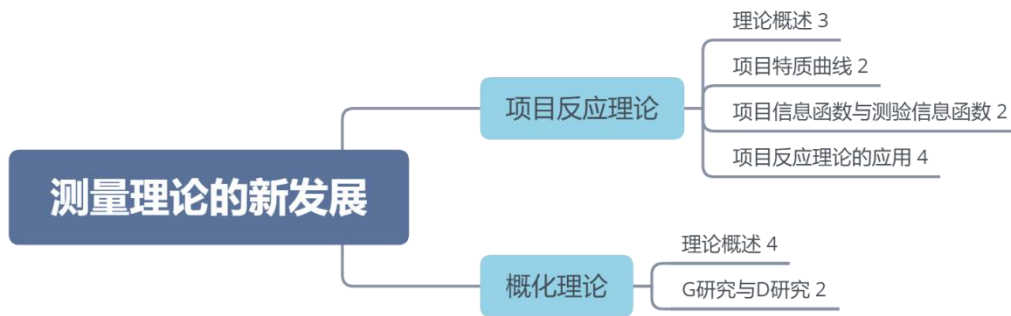


第三章 测量理论的新发展



一、项目反应理论 (IRT)

项目反应理论又称题目反应理论、潜在特质理论。

(一) 理论概述

1. 基本概念

(1) 潜在特质：被试某种相对稳定的、制约其行为的心理特征称为心理特质，但这种特质没有明确其物理或生理形态，又称潜在特质，用 θ 表示。

心理测量的任务即估计个体在潜在特质上的水平，然后预测其行为。

(2) 潜在特质空间：对某一特殊行为的发展起作用的所有潜在特质的集合，其维度越高，研究难度越大。

(3) 项目反应理论：主要研究被试在测验项目上的反应行为和与所测潜在特质之间的关系。项目反应理论认为，随着潜在特质 θ 的提高，正确反应该项目的概率 $P_{(\theta)}$ 也会提高。

2. 项目反应理论的优良性质

(1) 被试能力估计独立于测验题目的选择。

(2) 项目参数估计独立于被试样本。

(3) 被试能力参数量表与题目难度参数量表的一致性。

(4) 可以精确估计每一测试题目和测验针对每一被试施测的测量误差。对于测验测试能力为 θ 的被试，其测量标准误为： $ME(\theta) = 1 / (I(\theta) \wedge 0.5)$ 。 $ME(\theta)$ 是测量标准误（测量精度）， $I(\theta)$ 是项目信息量。如一个测验的所有题目对被试提供的项目信息量总和为 25，那么此测验对被试的测量精度（估计标准差）就是 $1 / (25) \wedge 0.5 = 0.2$ 。

3. 基本假设

(1) 单维性假设

指测验每一个项目能够只测量被试的某一种能力（如计算能力），而可以忽略其他能力对测验结果的影响（如阅读能力）。

(2) 局部独立性假设

同一特质水平的被试对不同测验的反应在统计上是独立的。即被试对一个测验项目的反应不受他们对其他测验项目反应情况的影响。

(3) 项目特征曲线假设

被试对题目所作出的反应概率遵循一定的函数关系。

(4) 非速度限制假设/无时间限制假设

因为测验的进行没有时间限制，若被试在某一项目上反应不理想，是由于能力不足，而非时间不够。

(二) 项目特征曲线

1. 项目特征曲线 (ICC)

用能稳定反映被试水平的潜在特质变量 (θ) 代替卷面总分作为回归曲线的自变量的曲线，记作 $P_{(\theta)}$ 。用来拟合项目特征曲线的函数，叫项目特征函数 (ICF)，如正态卵形函数、Logistic 函数。

该种函数中有以下几个参数：

(1) 参数 b ：题目难度。①若排除 c 的影响，潜在特征 θ 恰好等于 b 的被试，在该题目上正确作答的概率为 0.5；②若不排除 c 的影响，则同样条件下，被试在该题目上正确作答的概率为 $1/2 \cdot (1+c)$ 。横坐标 $\theta=b$ ，纵坐标 $P_{(\theta)}=1/2 \cdot (1+c)$ 的点是项目特征曲线的拐点，也是中心对称点，曲线递增的速率在此点由快转慢。因此， b 是特征曲线的定位参数，其值一旦确定，项目特征曲线在横轴上的位置也就确定了。

(2) 参数 a ：题目区分度，刻画测验题目对被试区分能力水平的高低。为曲线拐点处切线斜率的函数值，曲线在此处越陡峭， a 值越大，区分度越大意味着潜在特质 θ 在 b 值附近稍有变化，则在该题目上正确做对的概率差值就很大；相反，曲线越平缓， a 值越小， θ 的较大增减都不能引起正确作答概率的明显改变。

(3) 参数 c ：伪机遇水平，相当于猜测参数，实际测验中被试仅凭机遇作答而成功的概率，高质量的题目应有较小的 c 值。另外，直线 $P_{(\theta)}=c$ 是项目特征曲线的下渐近线，而 $P_{(\theta)}=1$ 是曲线的上渐进线。参数 c 的理论值范围 $[0, 1]$ 。

理想的区分度约在 +0.50 到 +2.00 之间。

2. 参数模型

(1) 单参数模型 (1-PLM)

足在“难度： b ”一个参数， $c=0$ ， $a=1$ 。答对题目的概率，除了跟潜在特质有关外，还和难度有关。曲线上答对题概率 50% 的一点便是曲线拐点，即 $P_{(\theta)}=0.5$ 时，横坐标 $\theta=b$ 。

(2) 双参数模型 (2-PLM)

除“难度”(b) 外，还有“区分度”(a) 参数， $c=0$ 。答对题目的概率，除了跟潜在特质有关外，还和难度、区分度有关。

(3) 三参数模型 (3-PLM)

有“难度”(b) 和“区分度”(a) 和“猜测”(c) 参数。答对题目的概率，除了跟潜在特质有关外，还和难度、区分度及猜测有关。

$$P_{(\theta)} = C + \frac{1 - C}{1 + e^{-1.7a(\theta-b)}}$$

认知测验偏好三参数模型，是非题猜测度定为 0.5 (四选一为 0.25)。

（三）项目信息函数与测验信息函数

1. 项目信息函数 (Item Information Function)

反映了不同参数的项目对每个被试特质水平进行评价时所提供信息的大小,其值由具体参数和被试特质水平决定。

项目信息量越大,说明此项目在评价被试特质水平时越有价值,测量误差越小。在双参数模型中,特质参数与难度参数相等时,误差最小,说明题目要与能力相匹配。

2. 测验信息函数 (Test information Function)

反映了整个测验在评价不同被试特质水平时所提供信息的大小,是项目信息函数的累加和。测验提供的信息量越大,则该测验在评价被试的特质水平时就越精确。

（四）项目反应理论的应用

1. 题库建设

除传统的三个参数(难度、区分度、猜测度)外,还有项目信息函数,为拓宽题库提供了条件,提高了题库参数完备性及可管理性。

2. 计算机自适应测验 (Computerized Adaptive Testing, CAT) 编制

计算机不断估计被试潜在特质 θ ,可累加项目信息函数,提高测验精度;每轮评估出 θ 后,可挑选题库中那些难度与 θ 接近的题目进行新的测试,直到精度满足预定要求。

CAT是当今测验发展的一个重要方向,也是项目反应理论最具特色的应用。项目反应理论可满足CAT的三个条件:

- (1) 快速估计被试能力参数和能力参数估计的精度。
- (2) 能针对精度目标,选出与被试能力匹配的题目。
- (3) 使用不同题目施测的被试能估计出定义在同一参照体系上的能力值。

3. 常模参照测验和目标参照测验的编制

对于常模参照测验可预先规定最大允许误差,然后根据已知参数直接拼出符合要求的测验;对于目标参照测验可较理想地实现其编制原则(准确划定分数线、降低误判率)。

4. 认知诊断测量模型发展基础

项目反应理论在项目分析、成绩评估量表的设计、态度问卷的设计等方面也有其应用。

二、概化理论

（一）理论概述

1. 起源

经典测量理论中测验误差是一个笼统概念,误差具体是由哪些因素造成、各部分误差方差在误差总方差中占的比例都不清楚。而概化理论则应用方差分量分析的方法深入分析测验误差的来源和结构。

2. 基本概念

概化理论提出在不同测验情境关系下,测量的误差结构和误差量都不同,一个测验情境关系由一个测量目标和若干个测量侧面构成。

(1) 测量目标:通过测量描绘的实体,测量目标分数方差相当于经典测量理论中的真分数方差。

(2) 测量侧面:除了测量分数方差外,所有其余的误差分数方差。一个侧面相当于某

方面的测量条件，一个测量侧面可以有不同的水平。

(3) 测量侧面被固定得越多，测量的信效度越高，但测量目标同样受到越高的限制。

3.基本思想

任何测量都处在一定的情境关系之中，应该从测量的情境关系中具体地考察测量工作。概化理论提出了多种真分数与多种不同的信度系数的观念，并设计了套方法去系统辨明与研究多种误差方差的来源，从而就可以有针对性地来控制和提高测量精度。

4.数据采集方法

(1) 全交叉采集：测量目标在所有测量侧面的所有水平上均被测量，数据信息最丰富。

(2) 相互嵌套采集：某个侧面的各个水平包含在另一侧面的各个水平中，最简单。

(3) 混合设计：包括以上两种方法。

(二) G 研究与 D 研究

概化理论的统计分析分为两个阶段，即 G 研究和 D 研究。在此分析中，应用者必须保证样本数据具有代表性和测验侧面不宜过多。

1.G 研究

研究者设计的测验情境关系及用一定方法采集的测验数据称为测验的观察领域，G 研究在观察领域数据上进行。

目的：采用方差分量分析法定量估计观察领域中测量目标的方差以及各测量侧面所产生的测量误差方差。

2. D 研究（决策研究）

D 研究是概化理论最具特色的计量分析手段，D 研究提供最终各种测验方案下的测验误差估计值。

概化理论把采取原始数据的原测验情境关系的测验侧面全体称为可测量全域，把研究者改变了的、意欲分析比较的那此新测验情境关系的测验侧面全体称为概化全域；一般来说，概化全域是可测量全域的子集。全域分数相当于经典测量理论中的真分数。

(1) 目的：利用 G 研究的结果数据，在原设计的测验情境关系范围之内，分析比较各种可能的测验方案，由所求的各种新测验情境关系下的误差方差和信度系数，选择实际的测验方案。

(2) 误差指标

①相对误差方差：所有与测量目标有关的交互效应方差之和，用于常模参照测验编制。

②绝对误差方差：除测量目标效应方差之外的所有方差之和，用于目标参照测验编制。

(3) 综合指标

①概化系数：即 G 系数，用相对误差估计出来的信度系数，测量目标效应/（测量目标效应+相对误差方差），是对常模参照测验的稳定性程度的度量。

②可靠性指数：即 ϕ 系数，用绝对误差估计出来的信度系数，测量目标效应（测量目标效应+总效应方差），是对目标参照测验的稳定性和一致性两种程度的度量。

两指标的开放等于实得分数与真实水平之间的相关，其值越高，测验信度就越高。