

文章编号: 1000-5862(2020)03-0292-09

标准参照测验分数报告中子分数估计方法

宋丽红¹, 汪文义², 丁树良²

(1. 江西师范大学初等教育学院, 江西 南昌 330022; 2. 江西师范大学计算机信息工程学院, 江西 南昌 330022)

摘要: 标准参照测验采用子分数衡量学生在具体内容、知识或技能上的掌握情况, 这有助于发挥考试的学习功能。在少量题目上作答数据估计子分数, 其信度较难保证。子分数估计方法可充分利用相关辅助信息获得信度较高的子分数, 这对补救教学至关重要。在简要介绍测量模型之后, 叙述了子分数的7类估计方法, 并分析了各方法的应用和表现, 需重点关注群体和个体、复杂结构、优化测验设计和其他施测方式下子分数估计。

关键词: 子分数; 目标表现指数方法; 增广分数法; 回归法; 项目反应理论模型

中图分类号: B 841 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2020.03.13

0 引言

标准参照测验(criterion-referenced tests, CRT)主要关注学生在特定内容、知识或技能上的掌握程度和表现水平。标准参照测验有助于发挥考试的诊断功能和促进学生发展, 从而对教育评价产生了深刻影响^[1]。标准参照测验已经广泛应用于水平、资格和成就考试等, 如国际学生评估项目(PISA)、国际阅读素养进步研究项目(PIRLS)、国际数学和科学成就趋势研究(TIMSS)、美国教育进步评价(NAEP)、美国研究生入学考试(GRE)、美国大学水平考试(CLEP)、中国国家基础教育质量监测(NAEQ)等^[2-4]。

标准参照测验开发涉及标准制定、项目编制、项目分析、标准设定、子分数估计等内容^[2-5-7]。测量标准往往涉及具体测量领域, 主要是请相关学科专家制定, 相关标准如美国共同核心州立标准和中国课程标准。项目编制需要严格按照标准或行为领域规范编写。项目分析既要进行内容分析又要进行统计分析, 与此同时筛选项目或确定测验长度。标准设定可采用Angoff法、Bookmark法等^[8]。如NAEQ采用了Angoff和Bookmark 2种方法相结合来确定划界分数^[4]。

标准参照测验作为教育测量的一种非常重要的测量形式, 常用于测量个体或群体在多个内容、知识

或技能领域中的掌握程度。各内容、知识或技能的掌握程度, 通常由领域分数(domain score)或子分数(subscores)来度量^[9-10]。子分数呈现学生在各内容、知识或技能上的学习状态, 这对教学和补救具有重要作用^[11-12]。子分数还可以用于预测和揭示测验结构^[13-14]。如美国大学入学考试英语和数学子分数可用于预测商业专业中统计和管理课程的学习情况, 以识别学生在专业选择中存在的风险; 对测验题目聚类, 所获得的子分数可用于揭示测验的多维性结构。

美国教育与心理测验标准强调信度、效度、可比性(comparability)和差异性(distinctiveness), 满足要求的分数或子分数才能报告。因为子分数往往基于在少量题目上作答的数据进行估计, 采用怎样的方法利用其他辅助信息用于估计子分数, 以提高子分数信度, 成为研究关注的焦点^[15-22]。本文在简要介绍测量模型之后, 详细叙述了子分数的7类估计方法, 并归纳和比较了各方法的应用和表现。

1 测量模型

在介绍子分数估计方法之前, 先约定以下记号。记被试 i ($i = 1, 2, \dots, N$) 在项目 j ($j = 1, 2, \dots, J$) 上的得分为 u_{ij} , 被试 i 在内容领域 c ($c = 1, 2, \dots, C$) 上的观察子分数列向量为 x_i , 所有被试在各内容领域上的观察子分数的均值列向量为 \bar{x} , 被试 i 在各内容

收稿日期: 2019-05-18

基金项目: 全国教育科学规划教育部重点课题“基础教育质量监测分数报告方法研究”(DHA150285)资助项目。

作者简介: 宋丽红(1981-), 女, 江西新干人, 副教授, 博士, 主要从事教育测量研究。E-mail: viviansong1981@163.com

领域上的真分数列向量为 $\tau_i = (\tau_{1i} \ \tau_{2i} \ \cdots \ \tau_{Ci})^T$, 属于内容领域 c 的项目集合为 $J(c)$, 即若项目 j 测量了领域 c 的内容, 则 $j \in J(c)$, 否则 $j \notin J(c)$. $J(c)$ 中项目数记为 J_c .

先给出用于分析 0-1 评分的单维和多维项目反应理论模型^[23]:

$$P_j(\theta_{1i}) = P(u_{ij} = 1 | \theta_{1i}) = 1 / (1 + \exp(-D \cdot a_{ji}(\theta_{1i} - b_j))) ,$$

$$P_j(\theta_i) = P(u_{ij} = 1 | \theta_i) = 1 / (1 + \exp(-D \cdot (a_j^T \theta_i - b_j))) ,$$

其中 $P_j(\theta_i)$ 表示能力向量 $\theta_i = (\theta_{1i} \ \theta_{2i} \ \cdots \ \theta_{Ki})^T$ 的被试在项目 j 上的正确作答概率, a_j 、 b_j 分别为项目 j 的区分度向量、难度, 维度数 K 可取为 C , 常数 D 可取 1.7 或 1.0, 单维模型中的符号含义类似.

2 子分数估计方法

标准参照测验子分数估计方法主要包括: 目标表现指数 (objective performance index, OPI) 方法^[19]、单维项目反应理论模型下子分数估计 (unidimensional model, UM) 方法^[9, 15]、增广分数 (augmented scores, AS) 方法^[20]、基于均方误差比例减量 (proportional reduction in mean squared error, PRMSE) 标准的回归法^[21]、高阶项目反应模型记分 (higher order item response model scoring, HO) 方法^[24]、多维项目反应理论模型下能力子分数估计 (multidimensional scoring, MS) 方法^[25]、群组子分数估计方法^[10]、基于认知诊断模型属性分数估计法 (在此不叙述)^[26-29] 等.

2.1 目标表现指数方法

目标表现指数方法假设子分数先验分布服从贝塔分布, 并且假设给定正确率条件下观察子分数服从二项分布^[19]. 利用贝塔分布为二项分布的共轭先验, 得到正确率的后验分布仍为贝塔分布, 由此以正确率的后验分布的期望作为子分数估计. 因为先验分布的均值和方差是根据被试在测验上能力估计并计算得到, 而子分数的经验贝叶斯估计正是利用了融入测验信息的先验分布, 因此目标表现指数方法采用测验上被试能力信息用来提高特定内容领域上子分数对应的真分数的估计精度. 目标表现指数方法估计子分数的主要步骤如下:

(i) 假设能力为 θ_{1i} 的被试 i 在内容 c 对应的子测验上正确得分率 T_{ci} 的先验分布服从贝塔分布 $\beta(r_{ci} \ s_{ci})$, 计算 T_{ci} 的期望 $\mu(T_{ci} | \theta_{1i})$ 和方差 $\sigma^2(T_{ci} |$

$\theta_{1i})$:

$$\mu(T_{ci} | \theta_{1i}) = r_{ci} / (r_{ci} + s_{ci}) , \quad (1)$$

$$\sigma^2(T_{ci} | \theta_{1i}) = r_{ci}s_{ci} / ((r_{ci} + s_{ci})^2(r_{ci} + s_{ci} + 1)) . \quad (2)$$

(ii) 根据被试 i 在整个测验上的得分向量 $u_i = (u_{i1} \ u_{i2} \ \cdots \ u_{ij})$, 基于项目反应理论模型可由极大似然方法估计被试能力 $\hat{\theta}_{1i}$, 然后可由得分向量 u_i 计算观察的 T_{ci} , 并由估计能力计算 T_{ci} 的期望:

$$T_{ci} = \sum_{j \in J(c)} u_{ij} / J_c ,$$

$$\hat{\mu}(T_{ci} | \theta_{1i}) = \sum_{j \in J(c)} P_j(\hat{\theta}_{1i}) / J_c ,$$

观察正确得分率的期望直接作为 $\mu(T_{ci} | \theta_{1i})$ 的估计.

(iii) 再计算 $\sigma^2(T_{ci} | \theta_{1i})$ 的估计值:

$$\hat{\sigma}^2(T_{ci} | \theta_{1i}) = \left(\frac{1}{J_c} \sum_{j \in J(c)} P_j(\hat{\theta}_{1i}) \right)^2 / I(\hat{\theta}_{1i}) ,$$

其中 $P_j(\hat{\theta}_{1i}) = \partial P_j(\hat{\theta}_{1i}) / \partial \theta_{1i}$, $I(\hat{\theta}_{1i})$ 为能力估计处的费希尔信息量. 在 2 参数逻辑斯蒂克模型下, 可得 $P_j(\hat{\theta}_{1i}) = Da_j P_j(\hat{\theta}_{1i}) (1 - P_j(\hat{\theta}_{1i}))$ 和 $I(\hat{\theta}_{1i}) = \sum_{j=1}^J D^2 a_j^2 P_j(\hat{\theta}_{1i}) (1 - P_j(\hat{\theta}_{1i}))$.

(iv) 为了检查被试 i 的不同子分数的维度一致性, 需要计算自由度为 C 的卡方统计量:

$$Q_i = \sum_{c=1}^C \frac{(T_{ci} - \hat{T}_{ci})^2}{\hat{T}_{ci}(1 - \hat{T}_{ci}) / J_c} .$$

(v) 视 $\hat{\mu}(T_{ci} | \theta_{1i})$ 和 $\hat{\sigma}^2(T_{ci} | \theta_{1i})$ 分别为 $\mu(T_{ci} | \theta_{1i})$ 和方差 $\sigma^2(T_{ci} | \theta_{1i})$ 的估计值, 由 (1) ~ (2) 式可解出 r_{ci} 和 s_{ci} 的估计值分别为

$$\hat{r}_{ci}^* = \hat{\mu}(T_{ci} | \theta_{1i}) \hat{n}_{ci}^* ,$$

$$\hat{s}_{ci}^* = (1 - \hat{\mu}(T_{ci} | \theta_{1i})) \hat{n}_{ci}^* ,$$

给定检验水平 α 的自由度为 C 的卡方统计量的上分位数 $\chi_{\alpha}^2(C)$, 若 $Q_i > \chi_{\alpha}^2(C)$, 则被试 i 的各个子分数的一致性不高, 不适合使用整个测验上被试能力信息作为先验信息, 此时取 \hat{n}_{ci}^* ; 否则, 取 $\hat{n}_{ci}^* = \hat{\mu}(T_{ci} | \theta_{1i}) (1 - \hat{\mu}(T_{ci} | \theta_{1i})) / \hat{\sigma}^2(T_{ci} | \theta_{1i}) - 1$.

(vi) 假设在给定 T_{ci} 条件下 $u_{ci} = \sum_{j \in J(c)} u_{ij}$ 的条件分布 $P(u_{ci} | T_{ci})$ 为二项分布 $b(J_c, T_{ci})$, 因为贝塔分布 $\beta(\hat{r}_{ci}^*, \hat{s}_{ci}^*)$ 为似然函数 $P(u_{ci} | T_{ci})$ (二项分布) 的共轭先验, 所以 T_{ci} 的后验分布仍服从贝塔分布

$\beta(\hat{r}_{ci}^* + u_{ci} \hat{s}_{ci}^* + J_c - u_{ci})$, 由 (1) 式可得 T_{ci} 的后验期望为

$$\hat{T}_{ci} = (\hat{r}_{ci}^* + u_{ci}) / (\hat{n}_{ci}^* + J_c).$$

值得注意的是, 当 $Q_i > \chi^2_\alpha(C)$ 时, 由 $\hat{n}_{ci}^* = 0$ 知 \hat{r}_{ci}^* 和 \hat{s}_{ci}^* 均为 0, 得 $\hat{T}_{ci} = T_{ci}$.

(vii) 若需要报告标准分数, 则可使用标准正态分布函数的反函数 Φ^{-1} 计算 \hat{T}_{ci} 对应的能力子分数^[16] 为

$$\hat{\theta}_{ci}^{\text{OPI}} = \Phi^{-1}(\hat{T}_{ci}).$$

2.2 单维项目反应理论模型下子分数估计

该方法采用被试在测验上所有题目上的作答反应估计整体能力, 而各能力子分数则采用测验上各内容领域上的题目作答反应分别估计^[15]. 在内容领域 c 下, 设 $\hat{\theta}_{ic}$ 表示被试 i 在试题集 $J(c)$ 上的能力估计值, 内容领域 c 共有 M_c 个试题, 这 M_c 个试题集合记为 $M(c)$. 设已经获得所有题目在统一量尺上的项目参数, 并给定题目 j 的抽样权重 w_j . 领域期望正确分数 $d(\hat{\theta}_{ic})$ 和百分数形式的领域分数 $D(\hat{\theta}_{ic})$ 计算公式如下^[9]:

$$d(\hat{\theta}_{ic}) = \sum_{j \in M(c)} w_j P_j(\hat{\theta}_{ic}) / \sum_{j \in M(c)} w_j,$$

$$D(\hat{\theta}_{ic}) = 100d(\hat{\theta}_{ic}) / M_c.$$

根据能力估计的标准误差 $S_E(\hat{\theta}_{ic}) = 1 / \sqrt{I(\hat{\theta}_{ic})}$, 可以计算出 $D(\hat{\theta}_{ic})$ 的标准误差为

$$S_E(D(\hat{\theta}_{ic})) = \frac{100S_E(\hat{\theta}_{ic})}{M_c} \sum_{j \in M(c)} w_j \frac{\partial P_j(\hat{\theta}_{ic})}{\partial \theta_{ic}},$$

能力估计 $\hat{\theta}_{ic}$ 可采用极大似然估计方法和期望后验估计方法等得出, 研究显示采用期望后验估计方法所得的领域分数估计的均方误差更小^[9].

2.3 增广分数方法

增广分数方法利用不同内容领域上子分数之间的相关性, 以提高各子分数的真分数估计的信度^[12, 30]. 在经典测验理论下, 该方法基于子测验的信度建立子分数的真分数对观察子分数的回归方程, 用来估计子分数的真分数. 增广分数的计算步骤如下:

(i) 计算被试 i 在内容领域 c 下的观察子分数向量 $\mathbf{x}_i = (x_{1i} \ x_{2i} \ \cdots \ x_{Ci})^T$, 元素 x_{ci} 可由得分向量 \mathbf{u}_i 计算, 即 $x_{ci} = \sum_{j \in J(c)} u_{ij}$. 同时计算观察子分数的均值

$$\bar{\mathbf{x}} = \sum_{i=1}^N \mathbf{x}_i / N.$$

(ii) 由观察子分数向量及其均值向量, 计算观察子分数的协方差矩阵 $\mathbf{S}^{(\text{obs})} = (s_{cc}^{(\text{obs})})$:

$$\mathbf{S}^{(\text{obs})} = \frac{1}{N-1} \sum_{i=1}^N (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T,$$

值得注意的是 $\mathbf{S}^{(\text{obs})}$ 概率为 1 的正定阵的充要条件是 $N > C$ ^[31].

(iii) 计算内容领域 c 上所有项目所组成的子测验的信度, 可采用克隆巴赫 α 系数计算子测验信度 ρ_c^2 作为 x_c 与 τ_{x_c} 相关系数 $\rho(x_c, \tau_{x_c})$ 的平方的估计:

$$\rho_c^2 = \rho^2(x_c, \tau_{x_c}) = \frac{J_c}{J_c - 1} (1 - \sum_{j \in J(c)} s_{cj}^2 / s_{x_c}^2),$$

其中 $s_{cj}^2 = \sum_{i=1}^N (u_{ij} - \bar{u}_j)^2 / (N-1)$, $s_{x_c}^2 = \sum_{i=1}^N (x_{ci} - \bar{x}_c)^2 / (N-1)$, $\bar{u}_j = \sum_{i=1}^N u_{ij} / N$ 和 $\bar{x}_c = \sum_{i=1}^N x_{ci} / N$.

(iv) 由子测验的信度和观察子分数的协方差矩阵计算子分数的真分数的协方差矩阵:

$$\mathbf{S}^{(\text{true})} = \mathbf{S}^{(\text{obs})} - \mathbf{D},$$

其中误差矩阵 \mathbf{D} 为对角阵, 即 $\mathbf{D} = \text{diag}((1 - \rho_1^2) s_{11}^{(\text{obs})}, (1 - \rho_2^2) s_{22}^{(\text{obs})}, \cdots, (1 - \rho_C^2) s_{CC}^{(\text{obs})})$.

(v) 建立子分数的真分数对观察子分数的回归方程:

$$\hat{\tau}_i = \bar{\mathbf{x}} + \mathbf{S}^{(\text{true})} (\mathbf{S}^{(\text{obs})})^{-1} (\mathbf{x}_i - \bar{\mathbf{x}}).$$

经过上面 5 步, 已经完成子分数的估计. 另外, 可通过下面的公式计算子分数的真分数估计的信度 $r_c = a_{cc} / b_{cc}$, 其中元素 a_{cc} 和 b_{cc} 分别来自于下面无条件真分数协方差矩阵 \mathbf{A} 和无条件真分数估计的协方差矩阵 \mathbf{B} :

$$\mathbf{A} = \mathbf{S}^{(\text{true})} (\mathbf{S}^{(\text{obs})})^{-1} \mathbf{S}^{(\text{true})} (\mathbf{S}^{(\text{obs})})^{-1} \mathbf{S}^{(\text{true})},$$

$$\mathbf{B} = \mathbf{S}^{(\text{true})} (\mathbf{S}^{(\text{obs})})^{-1} \mathbf{S}^{(\text{true})}.$$

对于上面的回归方法, 若只考虑特定内容领域上的真分数对其观察子分数的回归, 则回归方程变为 $\hat{\tau}_{ci} = \bar{x}_c + \rho_c^2(x_c, \tau_{x_c})(x_{ci} - \bar{x}_c)$, 该式为杜鲁门·李·凯利(Truman Lee Kelley)提出的 Kelley 回归分数方法. Kelley 回归分数方法是增广分数方法的单维版本.

在项目反应理论下, 该方法基于能力估计的信度建立能力对其估计的回归方程, 用来对能力进行估计^[16, 32]. 计算增广能力子分数的基本步骤如下:

(i) 估计被试 i 在内容领域 c 下的能力 $\hat{\theta}_i = (\hat{\theta}_{1i} \ \hat{\theta}_{2i} \ \cdots \ \hat{\theta}_{Ci})^T$, 可以采用期望后验估计方法估计得到 $\hat{\theta}_{ci} = \int_{-3}^3 \theta P(\theta | \mathbf{U}_i^{(c)}) d\theta$, 其中后验概率 $P(\theta |$

$U_i^{(c)}$ 可由似然函数 $L(U_i^{(c)} | \theta) = \prod_{j \in J(c)} P_j(\theta_i)^{u_{ij}} (1 - P_j(\theta_i))^{1-u_{ij}}$ 和能力先验分布 $g(\theta)$ 计算得到.

(ii) 将能力估计转换为回归变量 $\theta_i^* = (\theta_{1i}^*, \theta_{2i}^*, \dots, \theta_{Ci}^*)^T$, 其中 $\theta_{ci}^* = \hat{\theta}_{ci} / \rho_{\theta_c}$ 且信度 $\rho_{\theta_c} = s_{\theta_c}^2 / (s_{\theta_c}^2 + \bar{s}_{\theta_c | x_c}^2)$. 其中能力估计方差 $s_{\theta_c}^2 = \sum_{i=1}^N (\hat{\theta}_{ci} - \bar{\theta}_c)^2 / (N-1)$ 和条件方差均值

$$\bar{s}_{\theta_c | x_c}^2 = (1/N) \sum_{i=1}^N \int_{-3}^3 (\theta - \hat{\theta}_{ci})^2 P(\theta | U_i^{(c)}) d\theta,$$

用到的能力均值为 $\bar{\theta}_c = \sum_{i=1}^N \hat{\theta}_{ci} / N$.

(iii) 分别计算 θ_i^* 的协方差矩阵 $S^{(\theta^*)} = (s_{cc}^{(\theta^*)})$ 和 $\tilde{\theta}_i$ 的协方差矩阵 $S^{(\tilde{\theta})} = (s_{cc}^{(\tilde{\theta})})$:

$$S^{(\theta^*)} = \frac{1}{N-1} \sum_{i=1}^N (\hat{\theta}_i^* - \bar{\theta}^*) (\hat{\theta}_i^* - \bar{\theta}^*)^T,$$

$$S^{(\tilde{\theta})} = S^{(\theta^*)} - D^{(\theta^*)},$$

其中 $\bar{\theta}^* = \sum_{i=1}^N \hat{\theta}_i^* / N$, $D^{(\theta^*)} = \text{diag}((1 - \rho_{\theta_1}) s_{11}^{(\theta^*)}, (1 - \rho_{\theta_2}) s_{22}^{(\theta^*)}, \dots, (1 - \rho_{\theta_c}) s_{cc}^{(\theta^*)})$. $S^{(\theta^*)}$ 概率为1的 为正定阵的充要条件是 $N > C$ [31].

(iv) 建立能力对其估计的回归方程

$$\tilde{\theta}_i = (\bar{\theta}^* + S^{(\tilde{\theta})})^{-1} (\hat{\theta}_i^* - \bar{\theta}^*).$$

在该方法中,采用子测验信度修正期望后验估计的能力.这主要是为了克服贝叶斯估计存在向总体均值收缩而造成方差偏小的现象.若采用最大后验估计方法得到能力估计,则也需要进行相应修正 [30]. 在此不详细介绍.若采用极大似然估计方法估计能力,则无须修正,估计能力可直接作为回归变量 [11].

2.4 基于 PRMSE 标准的回归法

记被试 i 在测验上总分为 $x_{zi} = \sum_{j=1}^J u_{ij}$, 测验总分

均值为 $\bar{x}_z = \sum_{i=1}^N x_{zi} / N$. 所有被试的测验总分为列向量 $x_z = (x_{z1}, x_{z2}, \dots, x_{zN})^T$, 所有被试在内容领域 c 上所有项目所组成的子测验的总分为列向量 $x_c = (x_{c1}, x_{c2}, \dots, x_{cN})^T$. 第1种基于 PRMSE 标准的回归法是考虑特定内容领域上的真分数对其观察总分的回归 [21]:

$$\tau_{ci} = \bar{x}_c + \beta(\tau_{x_c} | \tau_{x_z}) (x_{zi} - \bar{x}_z),$$

其中回归系数 $\beta(\tau_{x_c} | \tau_{x_z}) = \hat{\rho}(\tau_{x_c}, \pi_{x_z}) \hat{\rho}(\tau_{x_z}, x_z) / (\hat{\rho}(\tau_{x_c}, x_c) s_{x_c} / s_{x_z})$. $\hat{\rho}(\tau_{x_z}, x_z)$ 为整个测验的信度系数的开方 $\hat{\rho}(\tau_{x_c}, x_c)$ 为特定内容领域的子测验的信度系数的开方 $\hat{\rho}(\tau_{x_c}, \pi_{x_z})$ 表示子测验真分数与测验真分数的相关系数. $\hat{\rho}(\tau_{x_c}, \pi_{x_z})$ 计算公式如下:

$$\hat{\rho}(\tau_{x_c}, \pi_{x_z}) = \hat{\rho}(x_c, x_z) / (\hat{\rho}(\tau_{x_c}, x_c) \hat{\rho}(\tau_{x_z}, x_z)) - \hat{\sigma}^2(e_{x_c}) / (\hat{\sigma}^2(\tau_{x_c}) \hat{\sigma}^2(\tau_{x_z})),$$

其中 $\hat{\rho}(x_c, x_z)$ 表示所有被试的子测验分数向量 x_c 和所有被试的总分向量 x_z 的样本相关系数,

$$\hat{\sigma}^2(e_{x_c}) = (1 - \hat{\rho}^2(\tau_{x_c}, x_c)) s_{x_c}^2, \hat{\sigma}^2(\tau_{x_c}) = \hat{\rho}(\tau_{x_c}, x_c) s_{x_c}, \hat{\sigma}^2(\tau_{x_z}) = \sum_{i=1}^N (x_{zi} - \bar{x}_z)^2 / (N-1).$$

第2种基于 PRMSE 标准的回归法是考虑特定内容领域上的真分数对其观察子分数和观察总分的回归 [21]:

$$\tau_{ci} = \bar{x}_c + \beta(\tau_{x_c} | x_c \cdot x_z) (x_{ci} - \bar{x}_c) + \beta(\tau_{x_c} | x_z \cdot x_c) (x_{zi} - \bar{x}_z),$$

其中 $\beta(\tau_{x_c} | x_c \cdot x_z)$ 表示在固定 x_z 保持不变时 π_{x_c} 对 x_c 的偏回归系数估计; $\beta(\tau_{x_c} | x_z \cdot x_c)$ 表示在固定 x_c 保持不变时 π_{x_c} 对 x_z 的偏回归系数估计. 这2个偏回归系数计算公式分别为

$$\beta(\tau_{x_c} | x_c \cdot x_z) = (\hat{\sigma}(\tau_{x_c}) (\hat{\rho}(\tau_{x_c}, x_c) - \hat{\rho}(\tau_{x_c}, x_z) \hat{\rho}(x_c, x_z))) / (\hat{\sigma}(x_c) (1 - \hat{\rho}^2(x_c, x_z))),$$

$$\beta(\tau_{x_c} | x_z \cdot x_c) = (\hat{\sigma}(\tau_{x_c}) (\hat{\rho}(\tau_{x_c}, x_z) - \hat{\rho}(\tau_{x_c}, x_c) \hat{\rho}(x_c, x_z))) / (\hat{\sigma}(x_z) (1 - \hat{\rho}^2(x_c, x_z))),$$

$$\text{其中 } \hat{\rho}(\tau_{x_c}, x_z) = \hat{\rho}(\tau_{x_c}, \pi_{x_z}) \hat{\rho}(\tau_{x_z}, x_z).$$

2.5 高阶项目反应理论模型方法

高阶项目反应理论模型通过引入一个高阶能力 (overall ability) 来考虑各个内容领域下子能力 (domain abilities) 的相关,即建立各个内容领域下子能力 θ_{ci} 对高阶能力 θ_i 回归 $\theta_{ci} = \lambda_c \theta_i + \varepsilon_{ci}$. λ_c 为潜在回归系数且满足 $|\lambda_c| \leq 1$, 随机误差 $\varepsilon_{ci} \sim N(0, 1 - \lambda_c^2)$ [24]. 假设在给定高阶能力情况下,不同内容领域下子能力相互独立,加上高阶能力服从标准正态分布的约束,可以计算出任意2个内容领域下子能力 θ_{ci} 和 $\theta_{ci'}$ 之间的相关系数为 $\lambda_c \lambda_{c'}$. 同时,可知子能力 θ_{ci} 也服从标准正态分布,即高阶能力和各

子能力位于同一量尺之上。

若高阶项目反应理论模型中项目参数已经根据得分矩阵由马尔科夫链蒙特卡罗方法(Markov Chain Monte Carlo, MCMC)估计完成。在给定项目参数之后,根据高阶模型的相关分布 $\theta_i \sim N(0, 1)$, $\lambda_c \sim U(-1, 1)$ 和 $\theta_{ci} | \theta_i, \lambda_c \sim N(\lambda_c \theta_i, 1 - \lambda_c^2)$,可得全条件分布^[16]:

$$P(\theta, \lambda, \Theta | U) = L(U | \Theta) P(\Theta | \theta, \lambda) P(\lambda) \cdot P(\theta),$$

其中似然函数 $L(U | \Theta) = \prod_{i=1}^N \prod_{c=1}^C \prod_{j \in J(c)} P_j(\theta_{ci})^{u_{ij}} (1 - P_j(\theta_{ci}))^{1-u_{ij}}$, $P(\lambda) = 1/2^C$, $P(\Theta | \theta, \lambda) = \prod_{i=1}^N \prod_{c=1}^C \varphi((\theta_{ci} - \lambda_c \theta_i) / \sqrt{1 - \lambda_c^2})$, $p(\theta) = \prod_{i=1}^N \varphi(\theta_i)$, 其中 φ 为标准正态分布的概率密度, $\Theta = (\theta_{11}, \theta_{21}, \dots, \theta_{Ci})^T$ 。对此全条件分布采样,可采用MCMC中吉布斯内M-H采样法(Metropolis Hastings within Gibbs)算法依次对 Θ, λ 和 θ 多次采样,然后去掉前面一部分采样值,剩下采样值取平均,即可得到相关参数的估计,如子能力向量组成的矩阵 $\hat{\Theta}$ 、回归系数向量 $\hat{\lambda}$ 和所有被试的高阶能力向量 $\hat{\theta}$,同时可计算各被试的能力子分数向量。

2.6 多维项目反应理论模型下子分数估计方法

基于多维项目反应理论模型的子分数估计方法主要有2种方法:一种方法是视各内容领域为不同的能力维度,直接采用多维项目反应理论模型估计各被试的能力向量^[24-25];另一种方法是采用子测验特征曲线将被试的能力向量估计进行转换,得到原始观察分数量尺上的各内容领域分数^[33]。设能力向量 θ 的协方差的先验分布 $\Sigma \sim W^{-1}(A_0^{-1}, \nu_0)$ 、能力向量的多元正态分布 $\theta_i = (\theta_{1i}, \theta_{2i}, \dots, \theta_{Ci})^T \sim MVN(0, \Sigma)$,则 Σ 和 θ 的联合分布为

$$P(\Sigma, \theta | U) = L(U | \theta) P(\theta | \Sigma) P(\Sigma),$$

其中 $L(U | \theta)$ 为似然函数, $P(\theta | \Sigma)$ 为被试能力向量的多元正态分布的概率密度乘积, $P(\Sigma)$ 为逆Wishart分布的概率密度。类似于高阶项目反应理论模型下参数估计方法,对上面全条件分布采样,可采用MCMC中吉布斯内M-H采样法算法依次对 θ 和 Σ 多次采样,然后去掉前面一部分采样值,剩下采样值取平均,即可得到多维项目反应理论模型所有被试的能力向量组成的矩阵的估计 $\hat{\theta}$ 和能力向量

协方差矩阵估计 $\hat{\Sigma}$,同时可计算各被试的能力子分数向量。

双因子模型是多维项目反应理论模型的一类特殊模型。在双因子模型下, Yao Lihua^[15]视双因子模型中被试 i 的一般能力 θ_{gi} 为整体领域能力估计,而其他维度 c 上能力视为各内容领域能力 θ_{ci} ,作为各内容领域分数估计。下面仅介绍一种较新且表现不错的加权合成方法^[22],用以估计整体领域能力和各内容领域能力:

$$\theta_{overall_i} = w_{1g} \theta_{gi} + \sum_{c=1}^C w_{1c} \theta_{ci},$$

$$\theta_{domain_ci} = w_{2g} \theta_{gi} + w_{2c} \theta_{ci}.$$

限于篇幅,权重 w_{1g}, w_{1c}, w_{2g} 和 w_{2c} 的各种设置方法请参见文献[22]。

2.7 群组子分数或领域分数估计方法

基于PRMSE标准的回归法已经推广用于群组水平的子分数的真分数估计^[34],由于其方法与个体水平的子分数的真分数估计方法类似,在此不进行详细介绍。为了评价特定群组在特定内容领域上的掌握程度,可以估计和报告群体在特定内容领域上的领域分数。当测验长度较短(如测验题数为5),个体领域分数信度不高、内容效度不好等情况时,不太适合直接报告个体领域分数。在内容领域 c 下,记 $N_c(g)$ 表示群组 g 的被试集合, N_{cg} 表示 $N_c(g)$ 中被试数, $\hat{\theta}_{ic}$ 表示群组 g 中被试 i 在 $J(c)$ 试题集上的能力估计值, $\hat{\theta}_{cg}$ 表示群组 g 的能力均值, $P_{cg}(\theta_{cg})$ 表示基于EM算法迭代估计的群组 g 的连续型能力后验分布 $f_g(\theta)$ 的离散化近似^[10, 35], Q 表示能力结点。 $\hat{\theta}_{ic}$ 可采用极大似然估计或期望后验估计方法进行估计, $\hat{\theta}_{cg}$ 为群组 g 中所有被试的能力估计值平均,或由群组项目反应理论模型直接估计得到。下面仅给出在0-1评分下群组领域分数估计方法。群组领域分数估计方法主要有以下4类方法。

(i) 基于领域个体观察分数平均的领域分数(OBS),它是直接计算特定群组所有被试在内容领域 c 上测验题分的均值:

$$O_{BS_{cg}} = \frac{1}{N_{cg} J_c} \sum_{i \in N_c(g)} \sum_{j \in J(c)} u_{ij}.$$

(ii) 基于领域个体能力的期望后验估计或极大似然估计的期望得分均值(记为EAP1、MLE1)的领域分数(ES),它是直接计算特定群组所有被试在内容领域 c 上测验项目的期望得分的均值:

$$E_{S_{cg}} = \frac{1}{N_{cg} M_c} \sum_{i \in N_c(g)} \sum_{j \in M(c)} P_j(\hat{\theta}_{ic}).$$

(iii) 基于个体能力的期望后验估计或极大似然估计均值(EAP2、MLE2) 或群体能力估计(MU) 的群组领域分数,它是直接计算群组能力在内容领域 c 上测验项目的期望得分均值:

$$M_{U_{cg}} = \frac{1}{M_c} \sum_{j \in M(c)} P_j(\hat{\theta}_{cg}).$$

(iv) 基于 EM 算法的群组领域分数,它是基于群组 g 的离散近似能力后验分布 $P_{cg}(\theta_q)$ 计算在内容领域 c 上测验项目的期望得分均值:

$$E_{M_{cg}} = \frac{1}{M_c} \sum_{j \in M(c)} \sum_{q=1}^Q P_j(\theta_q) P_{cg}(\theta_q).$$

已有研究显示^[10] 在作答题目数较小且群组样本量较小情况下,基于 EM 算法的群组领域分数估计方法表现好于基于个体能力点估计均值的群组领域分数估计方法. 当被试作答题目数增加时, EAP2、MLE2 方法表现接近于 EM 方法. 例如,当 $J_c = 5 \sim 10$ 、 $M_c = 30$ 且 $N_{cg} \geq 50$ 时,或者当 J_c 增加到 20 且 $N_{cg} \geq 25$ 时,基于 EM 算法的群组领域分数估计方法表现较好. 而在矩阵取样设计下, EM 方法和 OBS 方法表现优于 EAP2 方法^[36].

3 各方法的比较

表 1 从 3 个维度对上述 7 类方法进行了比较.

表 1 7 类子分数估计方法比较

估计方法	主要采用的测量模型	统计方法	所利用的辅助信息
目标表现指数(OPI) 法	单维项目反应理论模型	经验贝叶斯估计法	整个测验上的能力估计
单维项目反应理论(UM) 法	单维项目反应理论模型	矩估计方法	子测验或整个测验上能力估计
增广分数(AS) 方法	经典测验理论、项目反应理论	多元回归方法	不同内容领域下观察子分数或能力估计的相关
基于 PRMSE 标准的回归法	经典测验理论	多元回归方法	总分、子分数与子分数的真分数之间的相关
高阶项目反应模型记分法	多维项目反应理论模型	完全贝叶斯估计法	领域能力与高阶能力之间的相关
多维项目反应理论模型法	多维项目反应理论模型	完全贝叶斯估计法	领域能力之间的相关
群组子分数或领域分数估计方法	单维项目反应理论模型	经验贝叶斯估计法	群组中所有个体的子分数的信息

4 研究展望

1) 研究如何利用其他辅助信息估计群体和个体子分数. 国内外关于传统测验设计下子分数估计方法的研究已有不少,但仍有待为大尺度测量开发群体和个体子分数估计方法. 例如,如何利用已有方法的优势,开发一种适合矩阵取样技术下群体和个

目标表现指数方法和增广分数法的优点是利用了测验总体表现或其他领域表现参与领域分数估计. 基于 IRT 的期望分数百分比方法可以十分方便进行等值,因 IRT 项目参数具有不变性,增加、减少或替换项目对能力子分数影响较小. 多维方法不是单独估计各领域分数,而是兼顾领域能力之间的关系对各领域能力同时估计,还可估计整体能力,这类估计方法精度较高,但这类方法涉及比较复杂的模型和算法. OBS 计算简便,但受题册长度影响较大,如有研究者指出当题册中题数少于 20 时估计精度下降较大^[36].

前人开展相关子分数估计方法的比较研究,研究结论显示^[9,15-17,25,33]: MS 方法优于 UM 方法和双因子模型方法; MS 方法、HO 方法、AS 方法表现相当,而 OPI 方法表现最差; MS 方法和 AS 方法表现较好; MS 方法优于基于经典测验理论的回归法;项目反应理论模型下正确作答比率(真分数)估计方法优于基于经典测验理论的观察百分比记分方法. 整体而言,MS 方法和 AS 方法表现最好,能力子分数优于观察子分数. 但是 MS 方法依赖多维项目反应理论模型,其对样本量要求较高,估计算法相对较为复杂. 基于已有研究方法的比较结果,推荐使用 AS 方法.

体子分数报告方法. 对于矩阵取样技术下子分数估计方法研究并不多,矩阵取样设计下的子分数估计方法、估计的信度问题及解决办法仍缺乏深入探讨. 矩阵取样技术可缩短被试的测试时间,提高测试效率,是大规模教育质量监测实施的有力工具. 在矩阵取样设计下,被试往往只作答全部测试题中的一小部分,且每个被试所做题目不同或不完全相同,因此,其参数估计方法与传统测量中的估计方法有所

不同.在国内,辛涛等^[35-36]对矩阵抽样下群体领域分数的估计问题进行了介绍和研究,并没有涉及个体子分数估计.由于矩阵取样设计下考生在特定领域上的作答项目数有限,因此估计的考生个体能力会有比较大的偏差.针对此问题,是否可利用其他辅助信息(内容相关、维度相关、总分相关、群体信息和基础数据等)参与估计群体和个体子分数,以提高报告分数质量,值得探讨.

2) 有待分析认知诊断模型下子分数估计方法,并开展相关比较研究.许多大型标准参照测验的测验结构具有多维性,对于多维数据,可以采用多维项目反应理论对连续型能力建模,也可采用认知诊断评估对离散型属性建模.若能力、技能、知识的粒度比较粗,相互之间可能存在一定的相关关系,则可以选用多维项目反应理论模型,充分利用维度之间信息互借,从而提高领域分数估计的信度和效度.若粒度比较细,则可以选用认知诊断模型进行数据分析,由此可以获得丰富而准确的诊断信息.本文并没有叙述基于认知诊断模型的子分数估计方法,后续有待梳理和分析这方面文献.认知诊断模型下子分数估计方法能否与本文介绍的方法进行比较,也有待考虑.

3) 复杂结构下子分数估计方法的表现也有待深入研究.有研究显示标准参照测验的测量目标之间具有层级结构.如不同年级的学生会处于不同学习进阶水平、测量目标之间呈现层级结构.在学习进阶研究中,运用单维项目反应理论和潜在类别分析作为测量模型进行分析,其中前者占绝大多数^[37].认知诊断模型可以借助潜在类别分析实现^[38-39],并且有研究采用属性层级结构模型或其他测量模型分析学习进阶^[40-41].虽有研究提出考虑属性层级结构的认知诊断模型^[42-44],可以充分利用相互之间的关系并对其结构进行修正.鉴于测量目标、知识、技能和能力之间关系十分复杂,且题目与测量属性一般并非简单结构,有待开展新方法及应用研究.

4) 研究针对子分数信度的测验设计和测验实施方式,以及研究针对子分数表现水平的分类准确性和分类一致性的组卷方法.有许多大型测验正是因为测验设计不合理,无法报告可靠的子分数^[45].因此,提高子分数的可靠性,不能仅靠事后分析方法,也应该事先设计良好的测验^[46].子分数或其估计是否比总分更具价值,除了采用 PRMSE 指标进行选择,还可以采用效标关联效度方法、分类一致性指标、概化理论中概括化系数(G 系数)或其他子分

数增值评价指标.子分数等值、群体不变性和其他估计方法等最新研究方向也值得研究者关注.

5 参考文献

- [1] 戴海琦.心理测量学[M].北京:高等教育出版社,2010.
- [2] 甘良梅,余嘉元.标准参照测验分数体系的探讨研究[J].心理学探新,2006,26(3):79-83.
- [3] 辛涛,李勉,任晓琼.基础教育质量监测报告撰写与结果应用[M].北京:北京师范大学出版集团,2015.
- [4] Jiang Yu, Zhang Jiahui, Xin Tao. Toward education quality improvement in China: a brief overview of the national assessment of education quality [J]. Journal of Educational and Behavioral Statistics, 2019, 44(6): 733-751.
- [5] Carroll P E, Bailey A L. Do decision rules matter? A descriptive study of english language proficiency assessment classifications for english-language learners and native english speakers in fifth grade [J]. Language Testing, 2016, 33(1): 23-52.
- [6] Douglas K M, Mislevy R J. Estimating classification accuracy for complex decision rules based on multiple scores [J]. Journal of Educational and Behavioral Statistics, 2010, 35(3): 280-306.
- [7] Fein M. Test development: fundamentals for certification and evaluation [M]. Danvers: ASTD Press, 2012.
- [8] 汪存友,余嘉元.标准参照测验及格线设定研究中的模拟实验法[J].心理学探新,2009,29(2):81-85.
- [9] Bock R D, Thissen D, Zimowski M F. IRT estimation of domain scores [J]. Journal of Educational Measurement, 1997, 34(3): 197-211.
- [10] Pommerich M, Nicewander W A. Estimating average domain scores [J]. Journal of Educational Measurement, 1999, 36(3): 199-216.
- [11] Fu Jianbin, Qu Yanxuan. A review of subscore estimation methods [EB/OL]. [2019-10-13]. <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ets2.12203>.
- [12] Wainer H, Sheehan K M, Wang Xiaohui. Some paths toward making praxis scores more useful [J]. Journal of Educational Measurement, 2000, 37(2): 113-140.
- [13] Welborn C A, Lester D, Parnell J. Using act subscores to identify at risk students in business statistics and principles of management courses [J]. Journal of Education for Business, 2015, 90(6): 328-334.
- [14] Reckase M D, Xu Jingru. The evidence for a subscore structure in a test of english language competency for english language learners [J]. Educational and Psychological Measurement, 2015, 75(5): 805-825.

- [15] Yao Lihua. Reporting valid and reliable overall scores and domain scores [J]. Journal of Educational Measurement, 2010, 47(3): 39-360.
- [16] de la Torre J, Song Hao, Hong Yuan. A comparison of four methods of IRT subscore [J]. Applied Measurement in Education, 2011, 35(4): 296-316.
- [17] Sinharay S, Puhan G, Haberman S J. An NCME instructional module on subscores [J]. Educational Measurement: Issues and Practice, 2011, 30(3): 29-40.
- [18] Sinharay S. Added value of subscores and hypothesis testing [J]. Journal of Educational and Behavioral Statistics, 2019, 44(1): 25-44.
- [19] Yen W M. A bayesian/IRT index of objective performance [EB/OL]. [2019-40-43]. http://www.ets.org/Media/Research/pdf/Yen_OPI_1987.pdf.
- [20] Wainer H, Vevea J L, Camacho F, et al. Augmented scores—"borrowing strength" to compute scores based on small numbers of items [M]//Thissen D, Wainer H. Test scoring. Mahwah, NJ: Lawrence Erlbaum Associates, Inc, 2001: 343-387.
- [21] Haberman S J. When can subscores have value? [J]. Journal of Educational and Behavioral Statistics, 2008, 33(2): 204-229.
- [22] Liu Yue, Li Zhen, Liu Hongyun. Reporting valid and reliable overall scores and domain scores using bi-factor model [J]. Applied Psychological Measurement, 2018, 43(7): 1-15.
- [23] Reckase M D. Multidimensional item response theory [M]. New York: Springer, 2009.
- [24] de la Torre J, Song Hao. Simultaneous estimation of overall and domain abilities: a higher-order irt model approach [J]. Applied Measurement in Education, 2009, 33(8): 620-639.
- [25] Yao Lihua, Boughton K A. A multidimensional item response modeling approach for improving subscale proficiency estimation and classification [J]. Applied Psychological Measurement, 2007, 31(2): 1-23.
- [26] 马世晔, 章建石. 基于考试结果挖掘的教育评价: 理论与实践 [J]. 心理学探新, 2012, 32(5): 461-465.
- [27] Liu Ren, Qian Hong, Luo Xiao, et al. Relative diagnostic profile: a subscore reporting framework [J]. Educational and Psychological Measurement, 2018, 78(6): 1072-1088.
- [28] 康春花, 杨亚坤, 曾平飞. 海明距离判别法分类准确率的影响因素 [J]. 江西师范大学学报: 自然科学版, 2017, 41(4): 394-400.
- [29] 罗慧, 熊建华, 王晓庆, 等. 基于加权距离的一种认知诊断方法 [J]. 江西师范大学学报: 自然科学版, 2018, 42(1): 74-81, 88.
- [30] Thissen D, Wainer H. Test scoring [M]. Mahwah, NJ: Lawrence Erlbaum Associates, Inc, 2001.
- [31] 张尧庭, 方开泰. 多元统计分析引论 [M]. 武汉: 武汉大学出版社, 2013.
- [32] de la Torre J, Patz R J. Making the most of what we have: a practical application of multidimensional item response theory in test scoring [J]. Journal of Educational and Behavioral Statistics, 2005, 30(3): 295-311.
- [33] Haberman S J, Sinharay S. Reporting of subscores using multidimensional item response theory [J]. Psychometrika, 2010, 75(2): 209-227.
- [34] Haberman S, Sinharay S, Puhan G. Reporting subscores for institutions [J]. The British Journal of Mathematical and Statistical Psychology, 2009, 62(1): 79-95.
- [35] 辛涛, 谢敏. 群体水平领域分数及其估计方法 [J]. 心理发展与教育, 2010, 26(4): 416-422.
- [36] 辛涛, 谢敏. 矩阵取样设计中群体水平领域分数估计方法的精确性比较研究初探 [J]. 中国考试: 评价与测量, 2011(5): 3-12.
- [37] 姚建欣, 郭玉英. 为学生认知发展建模: 学习进阶十年研究回顾及展望 [J]. 教育学报, 2014, 10(5): 35-42.
- [38] DeCarlo L T. On the analysis of fraction subtraction data: the DINA model, classification, latent class sizes, and the Q-matrix [J]. Applied Psychological Measurement, 2011, 35(1): 8-26.
- [39] 王孟成, 毕向阳. 回归混合模型: 方法进展与软件实现 [J]. 心理科学进展, 2018, 26(12): 2272-2280.
- [40] Briggs D C, Alonzo A C. The psychometric modeling of ordered multiple-choice item responses for diagnostic assessment with a learning progression [C]//Alonzo A C, Gotwals A W. Learning progressions in science: Current challenges and future directions. Rotterdam, The Netherlands: Sense Publishers, 2012: 293-316.
- [41] 高一珠, 陈孚, 辛涛, 等. 心理测量学模型在学习进阶中的应用: 理论、途径和突破 [J]. 心理科学进展, 2017, 25(9): 1623-1630.
- [42] 宋丽红, 汪文义, 戴海琦, 等. 基于贝叶斯网的认知诊断模型构建 [J]. 心理科学, 2016, 39(4): 783-789.
- [43] 喻晓锋, 丁树良, 秦春影, 等. 贝叶斯网在认知诊断属性层级结构确定中的应用 [J]. 心理学报, 2011, 43(3): 338-346.
- [44] Zhan Peida, Ma Wenchao, Jiao Hong, et al. A sequential higher order latent structural model for hierarchical attributes in cognitive diagnostic assessments [J]. Applied Psychological Measurement, 2019, 44(1): 1-19.
- [45] Wainer H, Feinberg R. For want of a nail: why unnecessarily long tests may be impeding the progress of western civilisation? [J]. Significance, 2015, 12(1): 16-21.
- [46] Feinberg R A, Wainer H. When can we improve subscores

by making them shorter the case against subscores with overlapping items [J]. Educational Measurement: Issues

and Practice 2014 33(3):47-54.

The Subscore Estimation Methods for Score Reports in Criterion-Referenced Tests

SONG Lihong¹, WANG Wenyi², DING Shuliang²

(1. Elementary Education College, Jiangxi Normal University, Nanchang Jiangxi 330022, China;

2. College of Computer Information Engineering, Jiangxi Normal University, Nanchang Jiangxi 330022, China)

Abstract: Criterion-referenced tests focus on mastery statuses of students in different content, knowledge or skill areas by subscore, which can be beneficial for playing the learning function of the examination. Subscore only estimated from a small number of test items makes it difficult to establish high reliability. Subscore estimation methods can make full use of auxiliary information on associated test data to obtain subscore with higher reliability, which is crucial to remedial teaching. After a review of measurement models is briefly introduced, the seven subscore estimation methods and their detailed computational procedures are explained. And the application and performance of each method are analyzed. Finally, subscore estimation of group-levels and individual-levels or under complex structures, optimal test design and test administration needs to be focused on.

Key words: subscore; objective performance index; augmented scores; regression method; item response model

(责任编辑: 冉小晓)

(上接第 246 页)

The Amide-Directed C—H Functionalization Catalyzed by Rh-Catalysts

LIN Junyue¹, YANG Jiefang², DING Qiuping^{1*}, HUANG Yaru³, GONG Chun², LIN Hong², GAO Wei^{1,2*}

(1. College of Chemistry and Chemical Engineering, Jiangxi Normal University, Nanchang Jiangxi 330022, China;

2. Jiangxi Academy of Forestry, Nanchang Jiangxi 330032, China;

3. The Attached Middle School, Nanchang University, Nanchang Jiangxi 330013, China)

Abstract: The direct functionalizations of C—H bond catalyzed by transition metal is a new strategy to construct C—C bond and C—X (O, S, N, etc), providing an alternative, environmentally friendly and efficient way to construct C—C and C-hybrid bonds. Among them, Rh catalyzed C—H functionalizations of groups directed by amide group has attracted extensive attention. In recent years, the research in this field has become hot. The research progress of this strategy in the field of organic synthesis is provided for readers in this review, mainly including the alkynylation, arylate activation/ cyclization and amination of C—H bond.

Key words: Rh catalysis; amide; C—H bond activation

(责任编辑: 刘显亮)