**心理学院 R 编程语言**

**分析与计算可复现性检验报告**



**Group13**

**对*Eisenberg et al. (2019)*研究结果的计算可复现性检验**

小组成员分工

|  |  |  |  |
| --- | --- | --- | --- |
| 组长 | 孙心茹 | | |
| 组员 | 何嘉铭 陈铭玉 李雪 王孜欣 | | |
| 分工 | | | |
| 数据分析 | 孙心茹 何嘉铭 陈铭玉 李雪 王孜欣 | PPT 制作 | 孙心茹 |
| 文字报告制作 | 何嘉铭 李雪 陈铭玉 王孜欣 | PPT 展示 | 孙心茹 |

**摘要：**心理学研究在揭示认知过程与行为现象方面成果丰硕，但知识的累积整合仍面临挑战，这源于学科传统孤立及重解释轻预测的倾向，尤其在自我调控(self-regulation)等多维度构念的研究中表现突出。为促进理论整合与知识累积，本研究旨在复现并拓展一项先前研究，该研究利用数据驱动的本体论方法(包括探索性因素分析、层次聚类等)分析了自我调控指标。我们采用R语言重新进行了数据分析，不仅复现了原研究的探索性因素分析、聚类分析及预测分析，还创新性地引入了双因子(bifactor)模型进行深入考察。尽管具体分析结果与原文存在一定差异，但核心结论高度一致：认知任务与心理测量虽同声称测量自我调控，却缺乏实证关联；二者内部均显现出稳定的因子结构，提示理论整合的潜力；预测分析亦表明，量表对现实结果的预测力有限且异质性高，任务则基本无预测力，印证了自我调控作为单一构念的一致性不足。此外，本研究对双因子模型的应用条件进行了反思与探讨。综上，本研究不仅验证了原研究的结论，双因子分析的应用也为理解多维度构念的内部结构提供了新视角，进一步凸显了数据驱动方法在构建累积性心理科学中的价值。

**关键词**：自我调控;心理本体论;计算可复现性;探索性研究

**1 引言**

**1.1 所选文献信息**

**表 1 文献信息表**

|  |  |  |  |
| --- | --- | --- | --- |
| **1 文献基本信息** | | | |
| 所选文献 | Eisenberg, I. W., Bissett, P. G., Enkavi, A. Z., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-10301-1 | | |
| 数据来源 | https://osf.io/zk6w9/ | | |
| **2 文献选取** | | | |
| 文献主题是否包含不止一篇研究？ | * 是，且包含元分析研究   ☑ 是，但不包含元分析研究   * 否 | 文献此前被其他研究者重复过？ | * 是(附上原文链接)   ☑ 否 |
| **3 研究假设选取** | | | |
| 重复的研究假设 | 原文采用了数据驱动的方法，揭示自我管理的结构。有以下几点发现：  1.将有关自我调节的心理测量与认知任务的数据进行探索性因素分析，得到潜在的认知维度；使用Graphical lasso的方法可视化也表明两种测量类别之间的相对独立，即认知任务与心理测量所发现的相关变量应当分属于两个不同的心理空间。  2.研究通过层次聚类和动态树切割算法在心理空间中建立了集群。使用这种分析，部分加载在相似因子上的相关变量(dependent variables)被聚集在一起,从而揭示心理空间中的结构，并识别出不同的心理维度或簇。这些集群代表了不同的心理维度和决策策略。  3.使用预测技术评估了心理测量对一系列自我报告的现实世界结果的预测能力。 | | |
| 重复的研究假设是否在其他研究中经过重复？ | * 是(附上原文链接)   ☑ 否 | 文献共几个实验，重复的研究假设是第几个实验中的？ | 仅1个（一认知任务一心理测量） |
| 选择该假设的原因 | 该研究为探索性研究，未明确提出假设。重点在于对数据进行建模。 | | |
| **4 数据集选取** | | | |
| 是否采用原始数据？ | ☑ 是   * 否 | 是否对样本量进行修改？ | * 是(说明原因)   ☑ 否 |
| 若修改，报告原文样本量大小和修改后的样本量大小 | / | 若修改，报告使用 G-power 计算的修改后的样本量对应的效应量 | / |

**1.2 文献介绍**

心理学领域积累了大量认知过程与行为现象，但缺少累积性理解与系统理论(Meehl, P. E,1978)。例如，“自我调节(self-regulation)”这一概念被广泛使用，但相关的测量（如认知任务与心理测量或自我报告）多存在低一致性问题(Eisenberg et al., 2019)。这一现象不仅存于自我调节这一概念中，包括“冲动性(Impulsivity)”这一概念也存在相同问题(Huang et al., 2024)，同样地，在青少年群体中自我报告的社交焦虑症状与客观心理生理指标的对应程度也较低(De Los Reyes et al., 2012)。这些研究说明相关的测量（如认知任务与心理测量或自我报告）存在低一致性问题在心理学研究中是非常普遍的(Copas, J. B,1983)，也就是说，这些心理学概念的本体并没有在各项研究或测量中完整地体现出来。研究的作者呼吁采用“数据驱动本体建构(ontology discover)”的方式，通过分析个体在多种认知任务(Task)、心理测量(Survey)和与自我调节相关的自我报告现实世界结果之间的差异，构建一个心理学本体论以揭示自我调节的心理过程和特征，并评估了认知任务与心理测量的预测能力：评估这些认知任务与心理测量的结果对一系列自我报告的现实世界结果(如社会经济结果、药物和酒精使用、以及身心健康)的预测能力。

研究属于结构探索性分析，未显式列出传统假设并验证，而是基于数据驱动对变量进行探索。

**2 方法**

**2.1 样本**

522名参与者(Amazon MechanicalTurk平台收集)，其中150人进行了重测。

变量包括37 项认知任务（含129个变量）和 22 类心理测量（含64个变量）。

**2.2 原研究方法简介**

首先研究对心理测量(Survey)的64个因变量和认知任务(Task)的129个因变量进行了**探索性因素分析**，旨在探索大量变量之间的关系，识别共同的潜在变量。两种测量类别(心理测量与认知任务)均无法预测另一类别的因变量指标，并且两类别之间的相关性较弱，使用Graphical lasso的方法可视化也表明两种类别之间的相对独立。使用探索性因素分析EFA来定义认知任务和心理测量的心理空间，使用贝叶斯信息准则(BIC)估计维度数量。对其中的150人进行追踪，进行数据的重测。

接着，研究通过**层次聚类**和**动态树切割算法**在心理空间中建立了集群。使用这种分析，部分加载在相似因子上的相关变量(dependent variables)被聚集在一起。因子分析将相关变量嵌入到一个低维度的心理空间中，但因子本身并不代表具体的心理结构。聚类分析可以将因子载荷相似的相关变量聚在一起，从而揭示心理空间中的结构，并识别出不同的心理维度或簇。这些集群代表了不同的心理维度和决策策略。

最后，研究使用**交叉验证**结合**L2正则化线性回归**评估了心理测量对一系列自我报告的现实世界结果的预测能力。

研究分析数据使用了Python和R语言进行分析。

**2.3 重复思路说明**

复现作业主要选文章的重要的部分进行。原研究主要想探究自我调节的结构，而在对心理结构探究的过程中，原研究有使用到且比较重要的方法为探索性因素分析EFA和聚类分析(Thurstone, 1947; Everitt et al., 1974)，所以主要重点对这两个数据分析方法进行了复现。在心理健康干预与预防中，预测分析方法对现实生活中心理健康的促进有重要意义，具有较高的实际意义(Coppersmith et al., 2018)，因此也对这部分进行了复现。

此外，对原研究的复现使用新的方法bifactor进行重复验证后发现，问卷提取的维度一部分与原文章一致。主要原因可能是心理测量的数据结构不是特别复杂，所以提取出来与原方法比较一致；而认知任务的数据结构太复杂了，提取出来的与原方法差异较大，可能是需要对共性较小的因子进行剔除。

**3 结果**

**3.1 探索性因素分析**

对数据进行探索性因素分析，复现后发现认知任务的最佳维度数是5，而心理测量的最佳维度数是11，与原文差异不大。其中，研究复现心理测量的11个维度能够解释56%的方差，认知任务的5个维度仅能解释20%的方差。

**表 2 探索性因素分析结果的比较**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Task | | Survey | |
|  | 因子数 | 累计方差 | 因子数 | 累计方差 |
| 原研究  报告结果 | 12 | 58% | 5 | 23% |
| 本研究 | 11 | 56% | 5 | 20% |

**3.2 聚类分析**

**3.2.1 使用与原文献相同方法及新方法的聚类分析**

为了识别集群，研究对相关变量(dependent variables)的因子载荷进行了层次聚类分析。使用这种分析，部分加载在相似因子上的相关变量被聚集在一起。对数据进行与原文献相同的聚类分析进行复现后，发现认知任务中的聚类数是20，也就是说在复现中识别了20个任务相关变量集群；心理测量的聚类数是9，也就是在复现中识别了9个调查集群。这一结果与原文有所差异。

除了原文提及的方法，本研究还对该数据进行了新方法的尝试。Bifactor是一种潜变量模型，在心理学领域多有应用，用于分析多维数据的潜在结构。其核心思想是同时考虑一个一般因素(general factor)和多个特殊因素(specific/group factors)，从而更灵活地描述变量间的关联模式(Chen et al., 2012)，也曾有研究将自我控制(self-control)提取出来在Bifactor模型中作为一般因素(Ward et al., 2014)。除此之外，有一项对心理学中人格五因素模型(FFM)的一项心理测量学评估表明，Bifactor的拟合度优于其他模型，这表明一个一般因素可能具有理论意义(Biderman, 2014)。那么类似于自我控制，在自我调节这样的一个心理学概念中，考虑也可以使用Bifactor模型进行拟合。于是尝试使用Bifactor的方法对数据进行拟合，结果发现从认知任务中提取出1个一般因素，26个特殊因素，即在任务中识别了27个与自我调节相关的因素；从心理测量中提取除1个一般因素和11个特殊因素，即在测量中从识别了12个与自我调节相关的因素。

**表 3 聚类分析结果的比较(原文献方法与新方法)**

|  |  |  |
| --- | --- | --- |
|  | 聚类数 | |
|  | Task | Survey |
| 原研究  报告结果 | 15 | 12 |
| 本研究原方法 | 20 | 9 |
| 新方法 | 27 | 12 |

**3.3 对原文计算可复现性进行评估**

**3.3.1 使用与原文献相同方法**

在使用原研究方法进行复现分析时，本研究对原文中报告的6项核心指标进行了计算层面的比对，包括因子数量、解释方差与预测性能等（见表5）。结果显示：1项完全一致（任务因子数），3项偏差较小（如问卷解释方差、任务预测R²等），2项偏差较大，分别是任务解释方差(δ ≈13%)与问卷预测R²(δ ≈20%)。

整体结果表明，本次复现在数值层面表现出良好的可复现性，主要差异出现在模型性能的微幅提升上。

**表 5 结果可复现性的评估表**

|  |  |  |
| --- | --- | --- |
| 可复现性情况 | 数量及占比 | |
| *N* | *%* |
| 完全一致(*δ* = 0%) | 1 | 16.7% |
| 偏差较小(0% < *δ* < 10%) | 3 | 50.0% |
| 偏差较小(*δ* ＞ 10%) | 2 | 33.3% |
| 因舍入导致的偏差 | 0 | 0% |

\* 结果数量*N*指在重复分析中，对重复分析结果与原结果进行配对比较的次数。对于每个目标效应，结果包括一组数值，如汇总效应估计(summary estimate，如*t*值/ F值)、置信区间界限(confidence interval bound)、效应量(effect size)样本大小(size effect)等，应将原文中报告的每个数值与重复结果进行比较。例如，在一个*t*检验中，原文献报告了*t*值、95%置信区间、cohen’s d和样本大小，则这个效应中*N*＝4。将各效应的*N*求和即为全体数量。

本研究复现了原文中针对问卷与任务预测效能所提出的推论，并将复现结果与原文结论进行比对。如表6所示，原文指出任务预测力较弱(平均R² = 0.01)，问卷预测性能适度(平均R² = 0.10)，而本研究复现的平均R²分别为0.32与0.37，方向与原文一致，效果略有提升。因此，两项推论均被判断为一致，推论一致性为100%。

**表 6 推论的一致性的评估表(原分析方法)**

|  |  |  |
| --- | --- | --- |
| 推论的一致性 | 数量及占比 | |
| *N \** | *%* |
| 一致 | 21 | 100% |
| 不一致 | 0 | 0% |

\* 推论数量*N*指在重复分析中，对效应做出统计推断的次数。例如，仅进行了一个*t*检验，则*N*＝1；如果进行了一个2\*2的方差分析，并进行简单效应分析，则有可能有7个统计推断：两个主效应的推论，一个交互作用的推论，四个可能的简单效应分析的推论，因此*N*＝7。如果报告的*p*值相对于重复的*p*值落在显著性水平边界的另一侧，则被归类为推论不一致；反之为推论一致。

**3.3.2 使用与原文献不同方法**

在因子结构分析中，本研究采用Bifactor模型作为创新方法，对问卷与任务的维度结构进行了探索性建模。结果表明，任务数据中共提取出26个特异因子，远高于问卷的11个特异因子，验证了原研究关于“任务维度更复杂”的推论。该推论在结构层面与原研究一致，故本研究中基于创新方法的推论一致性为100%(见表7)。

**表 7 推论的一致性的评估表(创新方法)**

|  |  |  |
| --- | --- | --- |
| 推论的一致性 | 数量及占比 | |
| *N \** | *%* |
| 一致 | 1 | 100% |
| 不一致 | 0 | 0% |

**4 讨论**

**4.1 计算可复现性检验结果分析**

在对原研究进行计算可复现性检验的过程中，本研究对导致偏差或不可复现结果的潜在原因进行了系统分析。综合表8所列各类因素，以下几个方面是影响复现成功率的关键来源：

**1. 原文分析中存在未设定随机种子的问题**

原研究在聚类分析中使用了Dynamic Tree Cut等方法，这些算法在执行时包含随机初始化或计算过程中的随机性(Langfelder et al., 2008)。然而原始代码并未显式设定随机种子(Boulesteix, 2012)，导致每次运行结果可能略有不同——这一技术细节可能解释了我们在复现聚类数量与结构时出现的细微偏差(Eren et al., 2021)。

**2. 原文代码缺乏注释与结构说明，理解成本较高**

尽管原研究在OSF上提供了完整数据和分析代码，但整体结构复杂，缺乏清晰的文档说明(如 readme 文件)，函数命名和中间结果保存不清晰，增加了我们对每一步骤理解的难度，特别是在因子处理、相关图生成等模块，依赖较强的对代码语义的还原能力。

**3. 建模方法与原文不同步，影响数值可比性**

在复现过程中，我们使用Bifactor模型替代原文的层次聚类方法(Reise, 2012)，这使得部分分析结果无法进行逐值比对(δ比对)，也在一定程度上影响了计算可复现性的全面性(Goodman et al., 2016)。这种方法的改变虽然牺牲了直接可比性，但通过模型重构提供了更丰富的结构信息(Chen et al., 2012)，促使我们思考在复现研究中如何平衡方法创新与结果可比性(Patil et al., 2016)。

**4. 软件版本差异和操作环境问题**

原文未注明R版本与关键包的版本号，我们使用的 R 包(如 psych、glmnet)为当前最新版本，部分函数的默认参数或输出形式可能已发生变更，造成分析路径或结果的微小偏差。且原文有部分数据处理使用python完成。

**5. 重复者对R语言仍处于掌握过程之中**

本研究由学生团队完成，虽然具备一定R语言操作基础，但在面对复杂函数调用、建模逻辑还原时仍面临较大挑战，特别是在EFA模型设置与结果提取、网络图构建等环节，均需花费大量时间查阅帮助文档与官方文献。

总体而言，原研究在数据与代码共享方面具备较高的开放性，但由于部分随机过程未固定、代码缺乏注释、部分方法选择不同，加之我们对分析细节掌握尚不完全，导致部分计算结果与原文出现差异。

**表 8 计算上（不）可重复的原因分析表**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **可能原因** | | | **研究一** | **简要说明** |
| **原文献开放性问题** | **一般性开放获取问题** | 个别结果的微小差异，可能是由于分析中使用了没有设置固定种子的随机数； | √ | Dynamic Tree Cut 算法包含随机性，原文未固定随机种子 |
| 个别结果的微小差异，可能是由于印刷或复制粘贴错误； | × |  |
| 文章文本中程序报告不明确，包括纳入亚组的标准、缺乏或不正确报告用于回归模型的变量、以及未报告的单侧分析； | × |  |
| 在文章的开放实践声明中对研究的模糊标记。 | × |  |
| **OSF 开放获取特定问题** | OSF 中缺乏对数据和/或代码内容进行说明的文档(readme文档)； | × |  |
| OSF 上的数据与代码文件不一致，如代码中对部分数据进行了操作，但这部分数据在数据文件中无对应； | × |  |
| OSF上的数据存储问题，包括文件损坏或无法下载。 | × |  |
| **数据开放获取特定问题** | 没有提供原始数据； | × | 数据完整，使用原始OSF文件成功读取 |
| 没有提供处理后的数据； | × | 已严格按原文说明预处理，无此问题 |
| 没有提供数据处理过程的描述或代码。 | √ | 原文代码复杂但缺乏详细注释，部分过程需自行理解重构 |
| **代码开放获取特定问题** | 缺乏共享的分析代码或建模代码； | × |  |
| 软件包或软件版本的问题。 | √ | 原文分析可能使用旧版包(如psych、DynamicTreeCut)，部分函数输出不同 |
| **重复过程的原因** | **重复研究与原研究的区别** | 是否使用同样的数据集； | √ |  |
| 是否使用同样的数据分析软件及软件包； | × | 全部使用R代替原文的python＋R分析  使用 Bifactor 替代原文的层次聚类 |
| 是否使用同样的数据分析方法。 | × |
| **重复者相关因素** | 复制者此前是否有过 R 使用经验； | √ |  |
| 复制者对关于 R 的知识或操作上存在漏洞，较难理解原文章中的部分操作(可做简单说明)。 | √ |  |
| **其他影响因素** | **文献年份** | 文献发表年份是否较为久远，是否在开放科学运动之前； | × |  |
| **文献质量** | 文献引用量大小； | × |  |
| 是否有其他研究支持本文献结果； | × |  |
| 是否有其他研究对本文献结果进行了重复，重复结果如何(可做简单说明)。 | × |  |

**4.2 其他思考**

在本次复现过程中，我们不仅重新实现了原研究中的核心分析流程，也在尝试引入替代方法(如 Bifactor 分析)时，对数据分析本身的逻辑、策略与解释方式有了更深入的理解。这种从“模仿执行”到“自主判断”的过渡，是整个复现训练中最具启发性的部分。

首先，我们意识到：研究中所采用的分析方法会深刻影响结果的可解释性与稳定性(McNeish & Wolf, 2020)。以因子结构分析为例，原文使用的是层次聚类，而我们在复现中改用 Bifactor 分析——这种方法不仅提取出更多潜在维度，也揭示了不同因子间的共享与特异成分(Reise, 2012; Chen et al., 2012)。这种方法的替代虽然使得计算指标难以逐值对比，但在推论层面却反而增强了结构理解的深度(MacCallum et al., 1993)。这也促使我们重新思考：在面对开放性数据时，是否存在比原文方法更合适的策略(Hayton et al., 2004)？

其次，复现过程中我们也遇到了多个挑战，包括原始代码缺乏注释、数据结构预处理方式未详述、部分算法过程存在随机性等(Stodden et al., 2018)。这些都提醒我们：在科研实践中，可复现性不仅依赖于方法的选择，更依赖于研究透明度与共享细节的充分性(Nosek et al., 2015)。未来我们也将更加重视这一点，在自己的研究中尽可能提供完整、注释清晰的分析流程(Sandve et al., 2013)。

最后，针对本课程的教学安排，相比于传统课堂学习单一分析方法，复现训练不仅锻炼了我们的代码能力和文献解读能力(Wilson et al., 2017)，也让我们深刻体会到“分析是有选择的、科学是需要判断的(Leek & Peng, 2015)”。建议未来课程在讲解复现流程时，进一步加强对“如何处理方法不一致(Patil et al., 2016)”和“如何写可复现报告(Markowetz, 2015)”这些实际问题的指导，将技术训练与科研写作结合，帮助学生在完成任务的同时，建立起更完整的研究思维链条。

**参考文献**

Biderman, M. (2014). Against all odds: Bifactors in EFAs of Big Five data [Conference presentation]. 29th Annual Conference of the Society for Industrial and Organizational Psychology, Honolulu, HI, United States.

Boulesteix, A.-L. (2012). Randomness in statistical computing: The case of the stochastic rounding error. Computational Statistics, 27(1), 3-17. <https://doi.org/10.1007/s00180-011-0236-6>

Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J.-P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. Journal of Personality, 80(1), 219-251. <https://doi.org/10.1111/j.1467-6494.2011.00739.x>

Copas, J. B. (1983). Regression, prediction and shrinkage. Journal of the Royal Statistical Society Series B: Statistical Methodology, 45(3), 311-335.

Eisenberg, I. W., Bissett, P. G., Enkavi, A. Z., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. Nature Communications, 10(1), 3329. <https://doi.org/10.1038/s41467-019-10301-1>

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.

Fabrigar, L. R., & Wegener, D. T. (2011). Exploratory factor analysis. Oxford University Press.

Huang, Y., Luan, S., Wu, B., Li, Y., Wu, J., Chen, W., & Hertwig, R. (2024). Impulsivity is a stable, measurable, and predictive psychological trait. Proceedings of the National Academy of Sciences, 121(24), e2306735121. <https://doi.org/10.1073/pnas.2306735121>

Langfelder, P., Zhang, B., & Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. Bioinformatics, 24(5), 719-720. <https://doi.org/10.1093/bioinformatics/btm563>

Markowetz, F. (2015). Five selfish reasons to work reproducibly. Genome Biology, 16, 274. <https://doi.org/10.1186/s13059-015-0850-7>

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., ... Yarkoni, T. (2015). Promoting an open research culture. Science, 348(6242), 1422-1425. <https://doi.org/10.1126/science.aab2374>

Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten simple rules for reproducible computational research. PLoS Computational Biology, 9(10), e1003285. <https://doi.org/10.1371/journal.pcbi.1003285>

Stodden, V., Seiler, J., & Ma, Z. (2018). Enhancing reproducibility for computational methods. Science, 354(6317), 1240-1241. <https://doi.org/10.1126/science.aah6168>

Thurstone, L. L. . (1947). Multiple-factor analysis. a development and expansion of the vectors of mind. Quarterly review of biology(4), 88-89.

Ward, J. T., Nobles, M. R., & Fox, K. A. (2014). Disentangling self-control from its elements: A bifactor analysis. Journal of Quantitative Criminology, 31(4), 595–627. <https://doi.org/10.1007/s10940-014-9236-3>

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. PLoS Computational Biology, 13(6), e1005510. [https://doi.org/10.1371/journal.pcbi.1005510](https://doi.org/10.1371/journal.pcbi.1005510" \t "https://chat.deepseek.com/a/chat/s/_blank)