

# 博士課程進学率に関する因果モデルの構築に向けた 試行的分析のデータセット — 2022 年 3 月 17 日版 —

2022 年 3 月 17 日

文部科学省 科学技術・学術政策研究所  
第 1 調査研究グループ 研究官 高山正行

## 1. 概要

博士課程進学率，及びそれに関連すると思われるいくつかの数値指標について，年単位で 2007 年度～2019 年度の 13 年分をまとめたものです。

博士課程進学率の向上に関する政策的議論への貢献を目指した定量的な因果推論のため，「積の構造的因果モデル」に基づき，統計的因果探索アルゴリズム“LiNGAM”<sup>※1</sup>により試行的に分析を行うにあたって使用したデータを公開するものです。

本データを用いた分析結果については，オンライン環境である Google Colaboratory の形で，以下の通りコードと共に公開しております。また，これらの分析結果については別途レポートなどの形で公開を予定ですので，併せてご参照ください。

(本データを用いた分析例)

[https://colab.research.google.com/drive/1bQ5u-yy\\_1qBSb4Uf8lx2D03FS6FbxMNd](https://colab.research.google.com/drive/1bQ5u-yy_1qBSb4Uf8lx2D03FS6FbxMNd)

※1 Linear Non-Gaussian Acyclic Model

## 2. 形式

UTF 8 でエンコードされたシンプルな CSV ファイルです。1 行目はヘッダ行です。1 列目に年度を記載しており，年度ごとの時系列データになっています。

## 3 データの内容

本研究で使用した 2007 年度～2019 年度の年度ごとの各統計値を掲載しています。これらの統計値は学校基本調査や科学技術研究調査など各種の公開情報に基づくもので，それらの公開情報から分析に関連すると思われるものを，必要な範囲で改めてとりまとめたものです。

変数は x0～x8 までの 9 つあり，それぞれ以下に対応します。

x0	博士課程進学率		
x1	前年度 DC1 採択者数 (人)	x2	国全体の基盤的経費 (億円)
x3	大学研究本務者数 (人)	x4	一人当たりの基盤的経費 (億円)
x5	研究時間割合	x6	博士修了直後の大学教員就職割合
x7	博士修了直後のポストドク就職割合	x8	DC1 以外の経済的支援 (億円)

変数 x0～x8 の詳細やデータの引用元，欠損値の補填の方法については，お手数ですが，上述のコード内で冒頭に記載されているデータ説明欄に記載した論文のリンクからご確認ください。

#### **4. 利用上の注意点**

- 本研究は、変数の選定や欠損値の補填も含めて、試行的な側面を多く含みます。このデータセットはあくまで本研究テーマにおける分析にあたっての一例であり、この変数の選定や欠損値の補填方法が最善であるとは限りません。今後より適切な分析を行うにあたっての議論が多くの研究者・行政関係者から湧き起こることを期待し、データを公開することとしています。
- 本研究では、各統計値をできるだけ正確に反映するように心がけていますが、これらのデータセットを利用した上での不利益等について、作成者らは一切の責任を負いません。
- 本データは年単位のため、期間は 2007 年度～2019 年度の 13 年と長期ですが、データ点数は 13 点と多くありません。今後、最新の統計値の追加などでデータ点数は増やしていくことが可能ですが、データ数の少なさに起因し、分析結果についてはデータ点数の増減により、大きく変化する可能性もありますので、利用の際にはご注意ください。
- 本データセットについては、論文本編とは別途、別途データベースとしての著作権が付与されており、Creative Commons の“CC BY 4.0”の条件で、ご利用いただけます。

例えば、以下のような記載が考えられます。

---

高山正行：博士課程進学率に関する因果モデルの構築に向けた試行的分析のデータセット — 2022 年 3 月 17 日版 —, NISTEP (2022) DOI: 10.15108/data\_doctoral\_2022\_0317

---