

TE科2年生における欠点科目数の因果分析 —構造方程式モデリングによるアプローチ—

大石 信弘^{1,*}

Causal Analysis of the Number of Failing Subjects in Sophomore of TE Department - Structural Equation Modeling Approach -

Nobuhiro Oishi^{1,*}

The causal relationship between the number of failing subjects on an exam and each subject were analyzed by means of structural equation modeling (SEM). The data used were the results of the first semester final examinations of sophomore belonging to the Department of Information, Communication and Electronic Engineering of Kumamoto National College of Technology. Since these data provide unsuitable solution, a stepwise method was applied to extract four subjects that affect the number of failing subjects. These subjects are mathematics, physics, computer engineering and engineering practice, which are the base of science and engineering. I also determined the number of latent variables by means of scree plots and dendrograms. Then I propose a causal model in which one latent variables behind the four subjects affect the number of failing subjects. The lavaan package on R was used to apply the SEM. From the goodness-of-fit indices, it was found that the model explained well the causal relationship related to the number of failing subjects.

キーワード：学業成績、教学 IR、因果分析、構造方程式モデリング、R

Keywords : scholastic performance, institutional research, causal analysis, structural equation modeling, R

1. はじめに

近年、教学マネジメントを支援するために、学業成績や学校満足度アンケートなどを定量的に分析する教学 IR の重要性が増している⁽¹⁾。しかし、大学、高専などの高等教育機関や小・中・高校における教学 IR が十分に行われていないのが現状である。その背景として、教育改善につながる質的なデータが少ないことが指摘されている⁽²⁾。この指摘は例えば、アンケートの分析や科目ごとの成績の分析が中心になっていることを指している。

一方、本校の情報通信エレクトロニクス工学科 (TE 科) においては、専門科目を苦手と感じる学生が増えており、特に 2 年生の混合クラスでは、TE 科学生が保有する赤点科目 (欠点科目) が増加している傾向にある。低学年における専門科目への苦手意識は技術者として成長する際に大きな障壁になると危惧される。

筆者はこれまで、統計解析環境 R⁽³⁾を用いて因果分析⁽⁴⁾

や教育ツールの開発⁽⁵⁾を行ってきた。そこで本研究では、科目を横断して包括的に成績不振の原因を探り、どういふ視座に立てばカリキュラムの狙い通りに学生が成長するかを分析する。このことで、教員側が学生に与えるべき観点が得られ、また、学生側としては技術者として迷いなく成長する方向性が得られるものと考えられる。

なお、分析には R ver. 4.1.0 上の lavaan ver. 0.6.8 パッケージ⁽⁶⁾を用いた。

2. 分析に用いるデータについて

分析には、令和 3 年度 TE 科 2 年生の前期期末試験の成績を用いた。学生の人数は、48 名である。ただし、本校熊本キャンパスでは 1、2 年次においては、3 学科の学生からなる 3 クラスを構成した「混合クラス」という制度を導入しており、リベラルアーツ系の科目を受講する際はこの混合クラスでのクラスメートとともに授業を受けることになる。それに対し、専門基礎科目を受講する際は、学科ごとに科目で取り扱う内容が若干異なることから、学科ごとのクラスメートとともに授業を受けている。

¹ 電子情報システム工学系
〒861-1102 熊本県合志市須屋 2659-2
Faculty of Electronics and Information Systems Engineering,
2659-2 Suya, Koshi-shi, Kumamoto, Japan 861-1102
E-mail address: oishi@kumamoto-nct.ac.jp (N. Oishi).

表1 変数選択後の観測変数の要約 (n = 48)

名称	変数名	Min	1st Qu.	Med	Mean	3rd Qu.	Max	SD
欠点科目数	<i>fault</i>	0	1	2	2.67	4	8	2.22
数学	<i>Math</i>	23	43.00	57.5	56.54	67.00	89	15.72
物理	<i>Phys</i>	29	60.50	68	68.85	79.25	95	14.25
計算機工学	<i>CE</i>	13	60.75	71	67.83	81.00	92	17.44
工学演習	<i>EP</i>	53	75.00	79.5	78.62	83.00	94	8.64

分析に用いた成績データは、氏名が分からないように加工してある。これは、分析手法の特徴からくるもので、データの平均構造を用いて因果関係を述べるためには個人の特定を必要としないためである。

成績データの形式はデータフレーム型であり、学生ごとの欠点科目数、受講科目12科目(国語、世界史、倫理社会、数学、物理、スポーツ、英語、基礎電気、プログラミング、計算機工学、電気回路、工学演習)の点数が格納されている。学生数は48名なので、48×13の2次元データに各観測変数名(科目名など)のラベルが付随している。

3. 分析結果と考察

欠点科目数と全科目の点数からなる成績データの散布図行列を図1に示す。また、これらの観測変数間の偏相

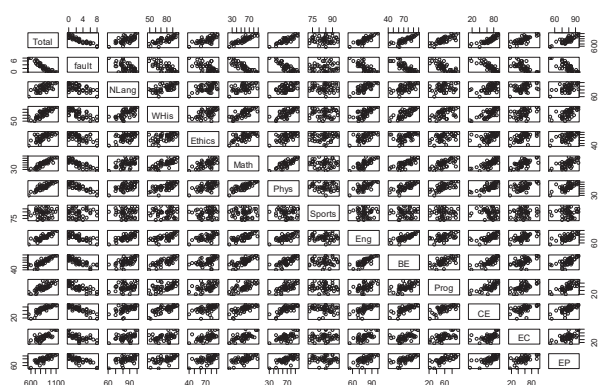


図1 欠点科目数と全科目の散布図行列

偏相関係数 > 0.2

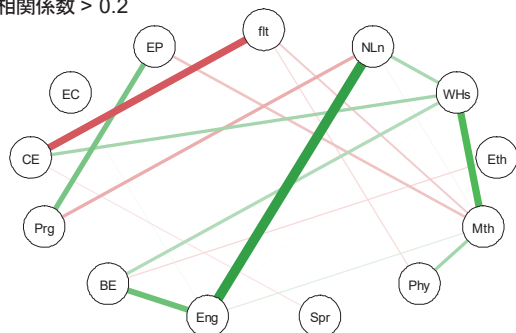


図2 欠点科目数と全科目の偏相関係数の可視化

関係数を求め、その符号と大きさを実線の色と太さで表したものを図2に示す。例えば、欠点科目数 (*fault*) と計算機工学 (*CE*) が赤色の太い線で結ばれているが、これは負の大きな偏相関係数を持っていることを示しており、計算機工学以外の科目の得点を変化させない条件下では、計算機工学の点数が高いほど欠点科目数が少くなることを示している。

全科目の点数を欠点科目数 (*fault*) と関係づける分析をしたところ、解が求まらない不適解となった。そこで、欠点科目数に結びつく科目をステップワイズ法により変数選択⁽⁷⁾⁽⁸⁾したところ、数学 (*Math*)、物理 (*Phys*)、計算機工学 (*CE*) および工学演習 (*EP*) の4科目に減らせることが分かった。この結果は変数減少法でも変数増減法でも同じ結果となった。そこで、これ以降は欠点科目数とこの4科目の点数で構成される成績データ (48×5のラベル付き2次元データ) を分析の対象とすることにした。分析対象の成績データの要約を表1に示す。

因果モデルを作る際に、観測変数の背後に潜在変数を想定することで、抽象化した概念を用いて因果関係を表現できるようになる。そのために、まず潜在変数の個数を決定する。潜在変数の個数を決める目安として、スクリープロットとカイザーガットマン基準が用いられることが多い。4つの科目の相関行列の固有値から得られるスクリープロットを図3に示す。スクリープロットにおいては、固有値が急に減少し他の固有値とほぼ同じ値になる部分を肘と呼んでいる。潜在変数の個数は、この肘の部分の1個手前の個数±1個とするのが妥当であると言われている⁽⁹⁾。そのため、今回の分析対象には多くても2個の潜在変数が存在すると予想される。もう一つの目安であるカイザーガットマン基準は、固有値の値に着目したものであり、固有値の値が1以上の個数を潜在変数の個数とするものである。カイザーガットマン基準から潜在変数の個数は、1個であると予想される。

スクリープロットおよびカイザーガットマン基準から予想された潜在変数の個数が妥当かどうかを、階層的クラスタ分析で確認した。4科目のデンドログラムを図4に示す。クラスタの統合には、Ward法を使った。分岐の深さはどの分岐においても同じ程度であるため、クラスタとして一塊にすることは難しいと考える。つまり、

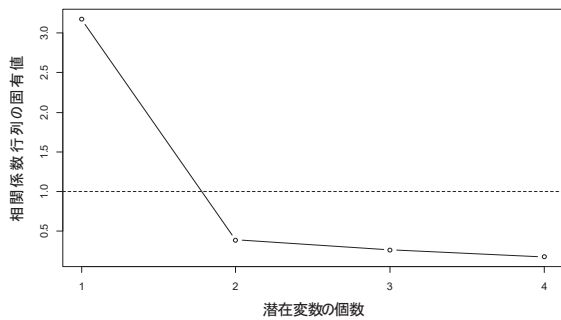


図3 理工系4科目のスクリープロット

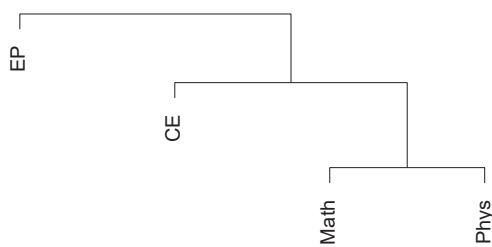


図4 理工系4科目のデンドログラム

グループは1個とみなされるので、4科目の成績の背後にある潜在変数の個数は1個であると考えられる。これは、スクリープロットおよびカイザーガットマン基準から求められた潜在変数の個数と一致する。他のクラスター統合手法でも、同じ結果が得られた。

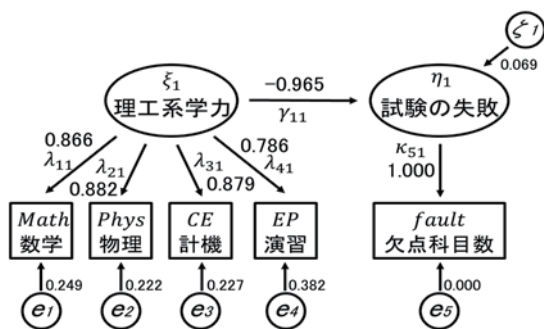


図5 提案する多重指標モデル

表2 提案モデルの潜在変数、観測変数の内容

変数名	変数の内容
ξ_1	理工系学力
Math	数学の点数
Phys	物理の点数
CE	計算機工学の点数
EP	工学演習の点数
η_1	試験の失敗
fault	欠点科目数

因果モデルを作るにあたっては、4科目の背後にある潜在変数を想定しているため、多重指標モデル^{(9),(10),(11)}を採用する。そのため、観測変数である欠点科目数を表す潜在変数をさらに導入して、潜在変数間の因果関係を表す必要がある。今回提案する多重指標モデルを図5に示す。各潜在変数、観測変数の内容を表2に示す。外生潜在変数 ξ_1 は4科目の成績の背後にあることから、「理工系学力」を表していると考えた。また、潜在変数 η_1 は欠点科目数と等価であることから「試験の失敗」とした。提案モデルでは「理工系学力」が「試験の失敗」の原因になっていることから、 η_1 は内生潜在変数となる。このモデルの測定方程式および構造方程式を次式に示す。

$$\begin{cases} \text{Math} = \lambda_{11} \times \xi_1 + e_1 \\ \text{Phys} = \lambda_{21} \times \xi_1 + e_2 \\ \text{CE} = \lambda_{31} \times \xi_1 + e_3 \\ \text{EP} = \lambda_{41} \times \xi_1 + e_4 \\ \text{fault} = \kappa_{51} \times \eta_1 + e_5 \\ \eta_1 = \gamma_{11} \times \xi_1 + \zeta_1 \end{cases} \quad (1)$$

(1)式をR上のlavaanパッケージを用いて解き、パス係数、残差誤差および適合度指標を求めた。lavaan スクリプトを図6に示す。求められたパス係数と残差誤差も図5に記入している。

このモデルの適合度指標を表3に示す。GFI (Goodness of Fit Index)、AGFI (Adjusted GFI) および CFI (Comparative Fit Index) は、観測変数の分散に対するモデルの説明率という観点から分析の精度を評価する指数で、0から1までの値を取り、0.95より大であると良適合と判断される。RMSEA (Root Mean Square Error of Approximation) と SRMR (Standardized Root Mean Square Residual) は実際のデータとモデルによる予測値とのずれの小ささに着目した評価指数で、0.05より小であると良適合と判断される。AGFIが0.9より小さいが、他の適合度指標から提案モデルは適合が良いと判断できる。

図5の左の部分では「理工系学力」という概念で表さ

```
library(lavaan) #lavaan パッケージ

#fault,Math,Phys,CE,EP のデータは変数 score に入っている
model <- '
f1 =~ Math + Phys + CE + EP #測定方程式
f2 =~ fault #測定方程式
f2 ~ f1 #構造方程式
'

fit <- lavaan::sem(model, data=score, std.lv=T, std.ov=F)
summary(fit, standardized=T)
fitMeasures(fit)
```

図6 lavaan スクリプト

表3 提案モデルの提案モデルの適合度指標

GFI	AGFI	CFI	RMSEA	SRMR
0.959	0.877	0.999	0.027	0.020

れるものが「数学」、「物理」、「計算機工学」および「工学演習」の点数（観測変数の値）に、それぞれ 0.866、0.882、0.879 および 0.786 の強さで現れていることを表している。「理工系学力」があれば、「物理」、「計算機工学」、「数学」の順に高い得点を取れることが分かる。「工学演習」は、これらの科目よりは若干低い点数になると考えられる。同様に図 5 の右の部分では「試験の失敗」という概念で表されるものが「欠点科目数」という観測変数に反映されていることを表しており、パス係数が 1 で残差誤差が 0 である。これは 1 個の潜在変数に結びついているのが 1 個の観測変数しかないためであり、「試験の失敗」と「欠点科目数」とが等価であることを意味している。そして、「理工系学力」が-0.965 という負のパス係数で「試験の失敗」に反映されている。これは、「理工系学力」が高いほど「試験の失敗」が小さくなる（失敗しない）ことを意味している。パス係数の絶対値が大きいことから、「理工系学力」だけで前期期末試験の欠点科目数が決められることが分かる。

これらのことから、TE 科 2 年生の前期期末試験においては、「欠点科目数」は「理工系学力」によって決まり、「理工系学力」があるかどうかを知るためには全 12 科目の中でも「数学」、「物理」、「計算機工学」および「工学演習」の理工系 4 科目だけを見れば見当がつくことを表している。

今回取り上げた理工系 4 科目以外の残りの 8 科目については、それらの得点が「欠点科目数」の傾向を表していなかったと解釈できる。例えば、理工系 4 科目の得点は高い学生でも「電気回路」では得点が低く欠点を取った、理工系 4 科目の得点は低かったが「プログラミング」では高い得点を取って欠点を取らなかったなどがあったためと考えられる。その他、平均点が高く、欠点を取った学生のいない科目で、分散も小さい場合にも、パス係数が小さくなるため、「欠点科目数」の傾向を表わさない例と考えられる。

4. まとめ

今回、TE 科 2 年生の前期期末試験における欠点科目数にかかわる因果モデルを提案し、構造方程式モデリングにより分析を行った。その結果、「欠点科目数」を決定するのは「理工系学力」であることが分かった。この「理工系学力」をモニターするには、全 12 科目のうち「数学」、「物理」、「計算機工学」および「工学演習」を見ればよいことが分かった。

提案した因果モデルより、専門基礎科目に苦戦してい

る TE 科 2 年生には、「理工系学力」を付けることを目標としてもらい、科目としては「数学」、「物理」、「計算機工学」および「工学演習」に励んでもらうことを推奨する。また、TE 科のカリキュラムを改訂する際には是非とも理工系学力を定着させる方策をご検討いただきたい。

今回分析に用いたデータは期末試験の成績データであり、分析のために新たにデータを取得したり複雑な加工をしたものではなかった。成績データは伝統的に成績評価に用いられてきた経緯があるが、総得点や平均点、偏差値などばかりでない新たな情報を引き出せる可能性があることが示唆された。

(令和 3 年 10 月 11 日受付)

(令和 3 年 12 月 24 日受理)

参考文献

- (1) 文部科学省：「大学における IR の現状とあり方に関する調査研究報告書」, pp.110-112 (2014).
- (2) 徳山工業高等専門学校：「グローバル高専事業（展開型）“青い鳥”グローバル教育プログラム 成果と課題 ～グローバル社会における高専教育～」, p.14 (2019).
- (3) “The R Project for Statistical Computing”, <https://www.r-project.org/>, (Retrieved Oct. 11, 2021).
- (4) N. Oishi, N. Yamamoto, A. Ishida and J. Murakami, “A Causal Analysis by Structural Equation Modeling of Sleep Monitoring Sensor Data”, IJEEE, vol. 8, no. 3, pp. 58-62 (2020).
- (5) 石田明男, 扇崎和希, 山本直樹, 大石信弘, 村上純：「第 87 号 60 頁の 4×4×4 立方体パズルについて（後編）」, 初等数学, 第 90 号, pp.18-22 (2021).
- (6) Yves Rosseel : “The lavaan tutorial”, pp.8-29, Ghent University (2019) .
- (7) 水野欽司：「多変量データ解析講義」, pp.61-69, 朝倉書店 (1996).
- (8) 「AIC を使った変数選択」, <http://www.hnami.or.tv/d/index.php?radvance>, (2021.10.11 閲覧).
- (9) 豊田秀樹：「共分散構造分析 [R 編] —構造方程式モデリング—」, pp.18-195, 東京図書 (2014).
- (10) 山田剛史, 杉澤武俊, 村井純一郎：「R によるやさしい統計学」, pp.309-319, オーム社 (2008).
- (11) 豊田秀樹, 前田忠彦, 柳井晴夫：「原因をさぐる統計学 —共分散構造分析入門」, pp.99-132, 講談社 (1992).