

H31 個人企業経済調査の補完について
(設備投資関連項目及び専従者給与の補完案)

■ 補完対象項目

ここでは、図 1 に示す全補完対象項目のうち、経理項目及び従業員数関連項目について補完を終えたデータに対して、点線で囲んだ設備投資関連項目と専従者給与について補完を行う問題について取り上げる。

図 1. 全補完対象項目一覧



■ 設備投資関連項目のための事前分析

1. 設備投資の有無について

設備投資関連項目 [19, 20, 201, 21] は、投資がないために数値が 0 となる企業が多く、同一企業であっても毎年類似額が計上されるわけではないという点で、経理項目とは異なる性質を持つ。H24 センサスデータで、H31 新調査の対象となる企業のものは 190 万レコード弱存在するが、その中で設備投資があるものは 1 割に満たず、補完にあたっては、補完値だけでなく、各欠測レコードについて設備投資の有無を推定するか、あるいは比率を決

めてランダムに割り付ける必要がある。

ここでは、H24 経済センサス-活動調査（以下 H24 センサス）データを用いて、投資の有無と関係のある項目の特定を試みた。まず設備投資の有無と他の経理項目及び従業者数とのピアソンの相関を調べたが、相関があるとみなせる項目は存在しなかった。次に、投資の有無を目的変数、表 1 の使用項目リストにある 15 項目を説明変数として、決定木 (CART) による分析を行った。説明変数は、H24 及び H28 センサスデータで利用可能な経理項目を網羅するが、分析の性質上、設備投資額（有形固定資産と無形固定資産の合計値）は除外している。

表 1. 使用項目リスト

変数項目	変数名	変数項目	変数名
都道府県番号	K_KEN	動産・不動産賃借料	KB_URIAGE_6
開設時期	KB_KAISETSU	減価償却費	KB_URIAGE_7
開設月	KB_KAITSUKI	租税公課	KB_URIAGE_8
従業者数合計	MTX_KI_TJUGYO	設備投資の有無	KB_SETSUBI
売上金額	KB_URIAGE_1	有形固定資産	KB_YUKEI
費用総額	KB_URIAGE_2	無形固定資産	KB_MUKEI
給与総額	KB_URIAGE_4	産業大分類	KB_SANGL

全産業を一括で CART により分析した結果、最初に選択される説明変数は [13 減価償却費] であった。産業大分類別の場合は、樹形図は様々なものになるが、最初に選択される説明変数は 14 産業のうち 10 産業が [13 減価償却費]、4 産業が [22 開設時期] であった。例として、図 2 に産業 I (卸売業・小売業) の樹形図を示す。この結果は、全体の 9 割を超えるデータについて、投資の有無を表 1 の変数により正しく分離できることを示すが、量の少ない設備投資ありのデータに着目すると、実際に投資のあるデータが正しく分類されるのは 1%程度である。さらに他の産業でも、投資ありに分類される企業数が実際よりも過小になる傾向があり、CART による判別は回答企業の投資の有無の比率が補完後に維持されない可能性が示唆される。

相関係数及び CART 分析の結果から、産業により設備投資の有無と [13 減価償却費] や [22 開設時期] に関係はあるが、単純な線形関係ではない可能性が高いことがわかる。また、今後設備投資の有無についてより良い推定ができる可能性は残されている。

2. 設備投資額について

産業別の設備投資額と、経理項目および従業者数との相関係数を、表 2 に示す。一部の産業に、ある程度の相関が認められる項目があるが、その項目は産業により異なる。また、ある程度の相関が認められる項目がない産業も多い。

図 2. I 卸売業、小売業の樹形図

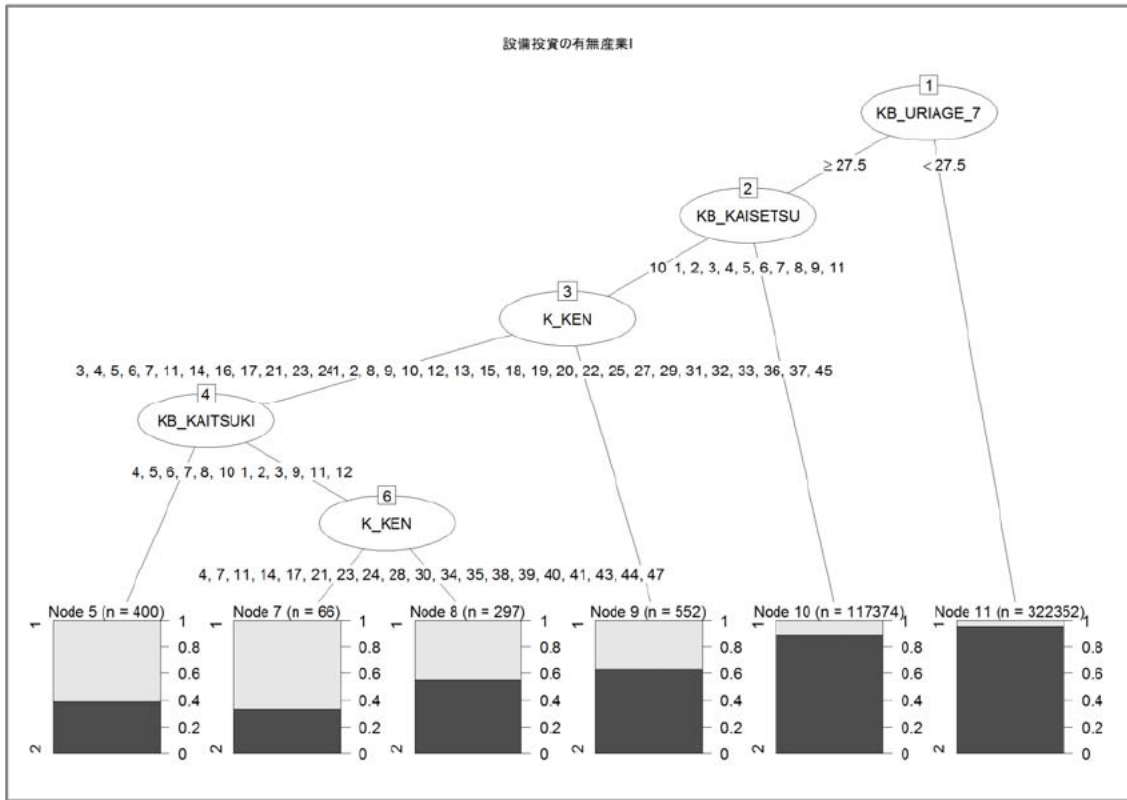


表 2. 設備投資額との相関係数

産業	従業者数						経理項目					
	個人業主	家族で無給の人	正社員	パート	臨時雇用者	従業者合計	売上金額	費用総額・経常費用	給与総額	動産・不動産賃借料	減価償却費	租税公課(法人_住民_事業税除く)
全産業	-0.01	0.01	0.14	0.11	0.08	0.16	0.13	0.12	0.12	0.06	0.03	0.04
D	0.01	-0.07	0.14	0.02	0.05	0.05	0.13	0.16	0.06	0.11	0.15	0.06
E	0.02	-0.07	0.36	0.12	0.04	0.34	0.24	0.24	0.25	0.12	0.53	0.23
G	0.03	0.13	-0.04	0.05	0.05	0.08	0.53	0.54	0.64	0.10	0.89	0.11
H	-0.07	-0.12	-0.12	0.39	0.88	0.40	0.19	0.34	0.24	0.11	0.05	0.17
I	0.01	-0.03	0.09	0.09	0.09	0.11	0.09	0.08	0.13	0.03	0.12	0.04
J	0.11	0.18	-0.01	-0.11	0.00	-0.01	0.07	0.14	0.02	0.02	0.12	-0.06
K	-0.01	0.03	-0.02	-0.01	-0.01	-0.01	0.06	0.05	0.01	0.00	0.23	0.11
L	-0.06	0.00	0.28	0.20	0.02	0.29	0.26	0.26	0.24	0.27	0.27	0.17
M	0.01	0.10	0.12	0.16	0.04	0.19	0.12	0.02	0.14	0.01	0.00	0.11
N	0.03	-0.02	0.16	0.43	0.00	0.33	0.68	0.67	0.35	0.39	0.67	0.27
O	-0.14	-0.03	0.24	0.11	0.01	0.19	0.18	0.17	0.17	0.10	0.19	0.13
P	0.01	0.07	0.14	0.09	0.14	0.15	0.11	0.11	0.12	0.02	0.16	0.05
Q	NA	0.21	-0.02	-0.25	-0.15	-0.25	-0.63	0.26	-0.17	0.23	0.67	0.58
R	-0.01	-0.03	0.09	0.09	-0.04	0.04	0.13	0.14	0.03	0.02	0.14	0.12

■ 事前分析結果を踏まえた状況の整理

1. 設備投資の有無の推定について

将来的に、CART による設備投資の有無の推定可能性はあるが、樹形モデルの時系列的安定性等の検討には、ある程度のデータの蓄積が必要と思われる。

また、投資の有無を推定する項目として、[13 減価償却費] と [22 開設時期] が有望であり、特に [22 開設時期] については、開設から 1 年経過した企業に設備投資があるという結果が得られている。ただし、H31 の新調査時に、事前に利用可能なデータは H28 センサスであり、H28 年調査当時の新設企業は設立後 3 年経過している。また、開設時期はセンサスの調査項目であるが、新調査の調査対象項目ではないため、H32 年以降も開設から 1 年経過した企業を特定することできない。

2. 補完クラスについて

経理項目の補完において、補完クラス内でドナーとなる補完対象項目について欠測のないデータの大きさ n を確保するために、標本層の設定に採用された都道府県の別を、補完クラスの設定には使用しない方針である。設備投資関連項目については、投資額 0 の企業も多いため、さらに n が少ないことが予想され、やはり都道府県別のクラス設定はかなり困難が予想される。一方で、設備投資は、災害等により特定地域で集中して発生するといったことも想定される項目である。

3. 設備投資額等の推定について

既存データのある [19 設備投資 (新規設備取得)] について、推定に用いることのできる有望な項目が産業別に異なり、またそのような項目が存在しない産業も複数ある。また、[20 うち車両・・・]、[201 (車両 機械 工具 器具 備品について) 取得額が最も多かった時期] 及び [21 設備投資 (中古設備取得)] については、事前に分析できるデータ自体が存在せず、ホットデックのドナーを決める項目の妥当性を検証することができない。

■ [19 設備投資 (新規設備取得)]、[20 うち車両・・・] 及び [21 設備投資 (中古設備取得)] についての補完

補完クラス別に、異なる推定モデルによる補完を行うのは、調査集計上あまり現実的ではないため、ドナーを決める項目は [13 減価償却費] とする。

[A. 減価償却費によりドナーを決める比率ホットデック]

補完クラス内で、新調査の [13 減価償却費] が近いデータをドナーとして、ドナーとの減価償却費の比率をドナーの各項目の数値に乘じ、[19 設備投資 (新規設備取得)]、[20 うち車両・・・] 及び [21 設備投資 (中古設備取得)] について比率ホットデックを行う。ドナーが 0 の値である場合は補完値も 0 とする。

[B. 新調査データの設備投資の有無の割合に基づいた比率ホットデック]

- ① 新調査データで [19 設備投資 (新規設備取得)] の有無の割合を算出
- ② 上で計算した、投資がある比率分の欠測データをランダムに選び、[13 減価償却費] の値に基づき、[19 設備投資 (新規設備取得)] 及び [20 うち車両・・・] について比率ホットデックを行う。このとき、[19 設備投資 (新規設備取得)] に数値があるデータのみをドナーとする。
- ③ 新調査データで [21 設備投資 (中古設備取得)] の有無の割合を算出
- ④ 上で計算した、投資がある比率分の欠測データをランダムに選び、[13 減価償却費] の値に基づき、[21 設備投資 (中古設備取得)] について比率ホットデックを行う。このとき、[21 設備投資 (中古設備取得)] に数値があるデータのみをドナーとする。

※ H28 センサスの調査項目には [21 設備投資 (中古設備取得)] が存在しないため、設備投資の有無の割合は新調査データで決める必要がある。

■ 設備投資の時期の判定 (うち車両機械工具器具備品) の補完

事前に分析できるデータが存在せず、推定に役立つ可能性のある項目も存在しないため、この項目を補完する場合は、新調査データの比率に基づき、ランダムに付与する以上のことができない。本来 0 値が多い設備投資項目について、実測データに何らかの傾向がある場合、このような補完を行えば、ランダムなノイズを加えることになる恐れがある。このため、この項目は、カテゴリ項目の欠測の代表的な処理方法である、「不詳」の符号を設定することを推奨する。

■ 専従者給与の補完

[32 専従者給与] は、H28 センサスや H30 までの個人企業経済調査において調査項目ではない。また、それを推定するための候補データとして、専従者の人数が最も有望と思われるが、この項目は H28 センサスや H30 までの個人企業経済調査にも、H31 新調査にも存在しない。次案として、専従者を含む常用雇用者数が考えられるが、専従者給与データがなく、推定可能性も検討できない。このため、少なくとも H31 新調査においてこの項目の補完は推奨できない。