

## 企業調査のサンプリングバイアスの推定の試み — 回答するのは成功した企業ばかりなのか？

田中 辰雄<sup>1</sup> 山口 真一<sup>2</sup>

### 要 旨

企業調査の場合、調査テーマに関し失敗した企業は回答をしぶり、成功した企業は積極的に回答するというバイアスが考えられる。このバイアスがどれくらいかをデータ活用を事例にして推定を試みた。手法は、同じ問いを従業員に対して行い、すでにある企業調査の結果と比較する方法である。従業員は匿名でたずねられれば勤務先の企業がデータ活用に失敗していても気にせず回答するだろう。だとすれば二つの調査の結果のずれはサンプリングバイアスから生じたと解釈できる。この方法で比較したところ、データ活用が成功したという答えの率は、企業調査の方が従業員調査より 7.1%ポイント程度高かった。これがサンプリングバイアスによって生じたとすると、成功企業はそうでない企業よりも3割程度多めに回答したことになる。補正ウェイトに換算すると、回帰分析などをするとき成功企業に 1/1.3 すなわち 0.77 程度を乗じることになる。ただ、この程度のウェイトで結果が大きく変わる例は限られると思われ、そうだとすれば本稿の推定結果は、今回の事例については成功バイアスはあるにはあるが、それほど深刻なものではないことを示唆している。事例がひとつだけなので、本稿の結果が直ちに一般化できるわけではなく、一般的な結論を得るためには今後の研究の蓄積が必要である。ただ、今回の事例は官庁によるビジネス分野の調査だったという点で比較的サンプリングバイアスが小さいケースと考えられる。それでも一定のサンプリングバイアスがあったということは、企業の評判に関わるようなセンシティブな調査ではバイアスは大きくなる可能性があり、留意が必要であろう。

**キーワード：サンプリングバイアス、データ活用、情報通信産業、企業調査、バイアス補正**

### 1. 問題意識

企業に対するサーベイ調査では調査に答えないことによるバイアスの問題を考える必要がある (Peytchev, 2013)。調査テーマに関して良い成績を上げている企業は喜んで答えるが、うまくいっていない企業は答えをしぶるだろうからである。結果として、調査に答えてくれた企業は成功した企業ばかりになり、いわば成功バイアスとでも呼ぶべきバイアスが生じうる。

たとえば、顧客情報の保護措置をとっていますか、これまでデータ漏えいはありましたか、LGBT に配慮していますか、等の問いで調査した場合、それぞれのテーマに積極的に取り

<sup>1</sup> 慶應義塾大学経済学部教授

<sup>2</sup> 国際大学グローバル・コミュニケーション・センター講師

組んでいる自信満々の企業はすぐに回答してくるだろうが、取り組みの遅れている企業や失敗経験のある企業は回答をしぶり、後回しにすることは十分に考えられる。後回しにしているうちに忘れてしまって未回答になるとすると、回答企業と未回答企業にはバイアスがかかる。回答企業だけで計算すれば、これらの課題に取り組んでいる企業あるいは成功している企業の比率は高めに推計されてしまう事になる。

この成功バイアスは全体の回収率が低いほど、深刻になる。残念なことに企業調査の回収率は低く、半分以下のことが多い。なかには2~3割のこともあり、成功バイアスの大きさは相当の大きさになりうる。エビデンスベースの政策が提唱されるようになって以来、政策効果の評価は盛んに行われるようになったが、効果の評価が企業調査に基づく場合はこの成功バイアスの問題は避けられない。

この成功バイアスの大きさはどれくらいだろうか。回収率の低さから言ってバイアスは大きそうに思える。しかし、回収率が低いのは単にどんな調査であっても答えるのが面倒だという一般論のためで、意外と成功バイアスは小さいかもしれない。成功企業が多くなるバイアスがありそうだという事は誰もが感じているであろうが、その大きさは不明である。本研究はこの成功バイアスの大きさを推定することを目的とする。もし、成功バイアスの大きさがわかればウェイトをつけてこれを調整する道も開けてくる。

方法としては、従業員に直接聞くアンケートと企業に訪ねたアンケートの結果を比較する。ネット調査会社のモニターから集めた個人に、勤務先の企業について聞けば近似的に従業員に聞いたことになる。従業員であれば企業を代表するわけではなく、匿名ということもあり、失敗している場合でもかまわず回答するだろう。すなわちバイアスのないサンプリングに近いと想定できる。もしそうなら、従業員にたずねた結果と会社宛てのアンケートの結果と比較すれば、サンプリングバイアスの大きさをおおむね把握できる。

具体的な調査テーマとしては、データ活用に取り組んで効果があったかどうかをとりあげる。「データ活用」はビッグデータが話題になって以降のキーワードで、総務省は2015年の情報通信白書で、データ活用をして効果があったかどうか企業対象のアンケート調査を行い、その結果を公表している。本論文では、別途調査会社のモニターのなかからデータ活用を勤務先で体験した人をスクリーニングして集め、彼らに総務省アンケートとほぼ同じ設問を行って結果を比較した。モニターがどの企業に属しているかはわからないので、比較は個別企業単位ではなく、産業別ならびにデータ活用分野別での比較にとどまるが、それでも比較はできる。

結果として、企業調査ではデータ活用成功した企業が従業員調査より7.1%ポイント多かった。これがサンプリングバイアスのためだと考えると、成功企業はそうでない企業よりも3割多めに回答していることになる。したがってサンプリングバイアスは存在する。ただし、その大きさはそれほど大きくない。補正ウェイトに換算すると、回帰分析などをするときは回答企業に1/1.3 すなわち0.77程度を乗じることになるが、この程度のウェイトで結果が大きく変わる例は限られるからである。すなわち調査結果の信頼性はおおむね維持される。

今回の推定は、総務省のデータ活用という一事例について行ったもので、すぐに一般化はできない。取り上げる事例によって成功バイアスは変わりえて、たとえばデータ漏えいやLGBT配慮など企業の社会的評判に関わる事例では、成功バイアスは大きくなるだろう。

今回の推定は多くの事例の一つにとどまる。しかし、企業調査のサンプルバイアスの推定はこれまでほとんど行われておらず、実態はほとんど知られていなかった。今回の調査は日本ではおそらく初めてのものであり、今後積み重ねるべき事例研究の第一歩として意味があると考え、研究レポートとして報告する。

## 2. 調査方法

### 2. 1. 先行研究

調査に答えない対象者がいる事によるバイアス (nonresponse bias) は、古くから研究者を悩ませた問題であり、対処方法もいくつか提案されてきた。たとえば Rogelberg and Stanton(2007)は、調査後に答えなかった相手に電話などで追加調査を行う、あるいは、売上や業種などの公開情報について答えなかった企業と答えた企業の違いがあるか見てみる、など 9 通りの対処法を列挙している。Groves and Peytcheva (2008) は、設問に答えられないバイアス研究を 59 個集めてメタ分析をした包括的なサーベイである。近年は調査の回収率が低下傾向なので、企業調査ではバイアスがあるかどうか分析前に検討する論文は増える傾向にある。たとえば Werner et.al (2007)によれば、マネジメント分野の 5 つジャーナルの 5 年間のサーベイ論文 2,096 本のうち、未回答者バイアスをチェックしている論文は 31%あったという。

ただし、これらのチェックの多くは、未回答者バイアスがありそうかどうかを間接的に検定しているだけで、バイアスの大きさを直接に推定しているわけではない。たとえば、良く行われるチェック方法として、売上・業種など公開済情報について、その分布が回答企業と未回答企業で一致しているかどうかを見る方法がある。このチェックは無意味ではないものの、企業が回答をし渋る理由が売上・業種と関連していないと意味が無い。たとえば、データ活用について言えば、データ活用成功するか失敗するかが、売上・業種と関連していればチェックになりうるが、無関連ならチェックにはならない。無関連の場合、仮に成功企業がより多く回答するバイアスが働いていても、未回答企業と回答企業の売上・業種の分布は同じになってしまうからである。

バイアスを直接的に推定するためには、別途何らかの追加情報を必要とする。たとえば Holmes and Schmitz(1996)は、個人企業の廃業率の決定式を推定する時、パートナー (共同経営者) の情報を使う事で未回答バイアスを制御した。彼らは個人起業家が 5 年後に事業を廃業しているかどうかを推定しようとしたが、個人起業家の調査の回収率が 20%しかなく、大量の未回答者がいた。事業の継続意欲をなくした人はアンケートに回答する気なくなることは十分にありそうなことなので、回答者だけで式を推定したのではバイアスが発生する。ここで、彼らは個人起業家はパートナーを組んで企業を立ち上げる事が多いことに注目した。何人かが組んでパートナーとして企業を共同保有している時、一人が廃業した時にはほとんどの場合、同時に他の人もその事業を廃業しているはずである。彼らはこの関係を使って未回答者のなかから廃業したと思われる人を特定し、ここから未回答者を含む推定を行った。

また、Tomaskovic-Devey et.al. (1994) は 2 段階調査を行う事で未回答バイアス自体を推定した。まず、労働者に電話サーベイを行い勤務先についてたずねる (この時の回収率は 73%)。次にこうして得た勤務先 583 社の住所に郵送で調査票を送り、306 社から回答を得

た。未回答企業は 277 社で、回答企業と未回答企業を比較分析した。この調査のポイントは、一段目の調査での労働者の未回答・回答の差は勤務先とは無相関と考えられることである。すなわち労働者についてはランダムサンプルと見なせるので、企業調査での回答企業 306 社と未回答企業 207 社の違いから、未回答バイアスを知ることができる。一段目の調査の時に、労働者に分析者の関心事（たとえば勤務先企業の意味決定は中央集権的かなど）を聞いておけば、これら関心事の変数について未回答バイアスがあるかどうかを直接に測ることができる。たとえば意思決定が中央集権的な企業は回答しない傾向があるかというようなことが直接にわかる。

この二つの例が示すようにバイアスの大きさを知るためには追加的な情報が必要になる。本稿で利用する追加的な情報は、独自に行った従業員調査である。従業員に匿名でたずねれば、企業を代表するわけではないので、勤務先企業がデータ活用に失敗していても遠慮なく答えるだろう。すなわち、従業員調査に成功バイアスはなく、比較的ランダムサンプリングに近いと想定できる。これは前出の 2 段階調査を行った Tomaskovic-Devey et.al. (1994) と似た発想のアプローチである。

ただし、企業調査の調査対象である企業の名前は非公開であり、その従業員へアクセスする方法は無い。そこで、別途調査会社のモニターからデータ活用を経験した個人を抽出する。抽出した個人は企業調査のときの対象企業に勤めているとは限らないので、その企業の従業員とはいえない。しかし、集団としての統計的な特徴がほぼ等しいなら比較は可能である。本研究はこの「統計的な特徴がほぼ等しい」こと仮定して行われる。この仮定の妥当性は適宜検討していく。こうして得られた従業員調査の結果と企業調査の結果のずれからサンプリングバイアスを推定することができる。

## 2. 2. 総務省調査

情報通信白書でのデータ活用の効果分析は、2014 年に総務省が情報通信総合研究所へ委託した調査が元になっている（情報通信総合研究所(2015)）。この調査では企業に対してアンケート調査を行い、4,672 社から回答を得ている。まず、データ活用の分野として 5 分野、すなわち、1)経営全般、2)企画・開発・マーケティング、3)生産・製造、4)物流・在庫管理、そして 5)保守・メンテナンスの 5 分野をあげ、それぞれの分野でデータ活用をしているかどうかをたずねている（同調査 Q9）。表 1 の 1 列目がその結果で、データ活用をしていると答えた企業の比率を記してある。比率を見ると、経営全般と企画・開発・マーケティングではデータ活用している企業は 5 割程度に達しており、この 2 分野でのデータ活用が進んでいることが分かる。次に全企業に対しデータ活用で効果があったかどうかをたずねたのが表 1 の 2 列目である（同調査 Q17）<sup>3</sup>。データ活用して効果があった割合、いわば成功率は、データ活用をしている企業のなかでデータ活用に効果があったと答えた企業の比率なので、この二つの列の割り算で求めることができる。それが 3 列目で、55%から 65%の間の比率が得られる<sup>4</sup>。本稿の問題意識はサンプリング時の成功バイアスによってこの比率が

<sup>3</sup> Q9 と Q17 はこの調査での問い番号である。Q17 の調査票の具体的な文言は付録(2)に記した。

<sup>4</sup> 情報通信白書ならびに同報告書では、この比率は文章中にはあるが、表として示されていない。これはサンプリングの偏りを意識したからではないかと推測される。

過大になっていないかという点にある<sup>5</sup>。

表 1

分野	Q9	Q17	比率(%)
	データ活用 している企 業(%)	データ活用 で効果が あったと答え た企業(%)	
	(a)	(b)	(c)=(b)/(a)
経営全般	57.2	30.9	54.0
企画、開発、マーケティング	39.5	22.6	57.2
生産、製造	15.7	9.6	61.1
物流、在庫管理	18.6	12.4	66.7
保守、メンテナンス	17.8	9.7	54.5

n=4,672、unit=%

情報通信総合研究所(2015)「ビッグデータの流通量の推計及びビッグデータの活用実態に関する調査研究」

## 2. 3. データ活用の定義と効果の意味

この問いに答えるため、従業員調査として、調査会社の個人モニターにほぼ同じ形式のアンケート調査を行う。

調査のためにまず「データ活用」として何を意味するかを明らかにしておく必要がある。データ活用は幅のある概念である。極端に言えば、ある支店の一つの課の売上データの推移をグラフにして壁に貼り、営業目的をたてるだけでデータ活用と呼ぶこともできる。しかし、通常はその程度ではデータ活用とは呼ばず、全社的に大規模なデータを集め、統計ソフトを駆使して分析し、あるいは各種センサーのデータやスマホ履歴、GPS データなど新たなデータを使うことをデータ活用と呼ぶことが多い。データ活用の概念には幅があるので調査に先だって決めておかないと比較ができない。

我々は総務省調査と比較するので、総務省調査の定義を使うのが筋である。が、残念ながら総務省調査ではデータ活用を定義していない。企業に対してどれくらいデータを持っているかを最初に細かくたずね、次にそれらのデータを活用していますかと聞いている。たずねる相手は企業の ICT 担当者なので、データを明示するとその活用の仕方は業界の常識としてある程度了解できるからと思われる<sup>6</sup>。リストアップされているデータは、通常の販売・

<sup>5</sup> なお、サンプリング時のバイアスとしては、そもそも回答するのがデータ活用している企業に偏ると言うバイアスもありうる。表 1 でいえば、経営全般についてデータ活用をしている企業のデータ活用率 57.2%が実態より過大であるというバイアスである。本稿で扱うのは成功率、すなわち表 1 でいえば  $30.9/57.2=54\%$  という比率を持つバイアスである。後者の成功バイアスは前者の活用率バイアスから切り離して扱える。なぜならデータ活用していない企業はそもそも成功も失敗もしないため、成功バイアスを計算する時の分母・分子から除かれるためである。

なお、もしデータ活用する企業と活用していない企業で、成功バイアスを引き起こす”傾向”が異なれば、実際のデータ活用率が変化すれば全体の成功バイアスは変化する。本稿で推定した成功バイアスは、データ活用率が現時点の時の成功バイアスである。

<sup>6</sup> 調査報告書には回答者が ICT 担当者であるとは明示されていないが、1 カ月当たり受け

管理・生産データに加えて、電子取引の顧客データ、スマホの履歴、GPS、監視カメラ画像、コールセンター音声など広範である。これにあわせるなら、本稿の行う従業員調査でも、これらのデータを念頭においたデータ活用の説明を、しかも一般従業員にわかるように行う必要がある。本調査では次のような文言を用いた。

「本調査において、データ活用とは、POS データや SNS データ、位置情報、カメラの映像など企業にあるデータを用いて、業務を効率化したり、新事業・新サービスの開発、顧客分析、経営方針の作成など企業の活動に活かすことを指します。

たとえばスマホでの購買履歴から新商品を企画したり、取引データから営業・在庫管理を改善したり、顧客からの問い合わせ音声記録を分析して経営方針に反映させたりします。GPS データや監視カメラ映像を使う事もあります。」

大量のデータを使う事を明示し、さらに事例をあげてイメージをつかんでもらっている。自分の職場だけの売上データをエクセルでグラフにする程度ではデータ活用とは言えない事はわかるだろう。また、画像や音声データ、GPS 等に触れることで、一般の人はデータとっていないものもデータ活用に含められることを伝えている。

この説明の後で、調査会社のモニターに対して、勤務先でデータ活用を経験したことがあるかどうかをたずね、経験がある人のみを対象として選びだした<sup>7</sup>。使った調査会社はマイボイス社である。たずねたモニター総数は 17,122 人で、そのうち勤務先でデータ活用が行われたことがあると答えた人は 1,804 人であった。出現率は 10%程度となる。この 1,804 人が今回の分析対象者である。

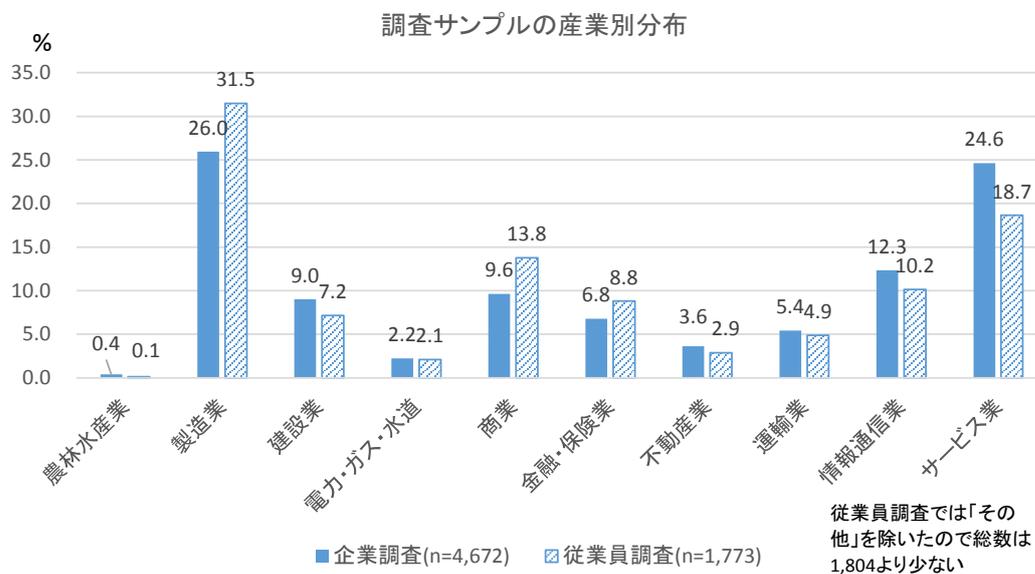
この分析対象者と総務省企業調査の対象者が比較可能であるためには、すでに述べたように統計的な特徴が一致している必要がある。そこでまず産業の分布を見てみよう。従業員調査のモニターは PC 利用者から選んでいるので、情報通信産業に勤務する人が増えるようなバイアスは無いだろうか。これを確かめるために、総務省調査での企業の産業分布と、今回の従業員調査での従業員の勤務先の産業分布を描いてみる。図 1 がその結果である。一見して明らかのように、両者は良く似ている。

---

取るデータの件数や通信端末台数、情報の受信頻度など ICT 担当者でなければ知りえないことをたずねており、現実に答えているのは ICT 担当者と考えられる。

<sup>7</sup> 調査票の具体的な文言は付録(1)参照

図 1

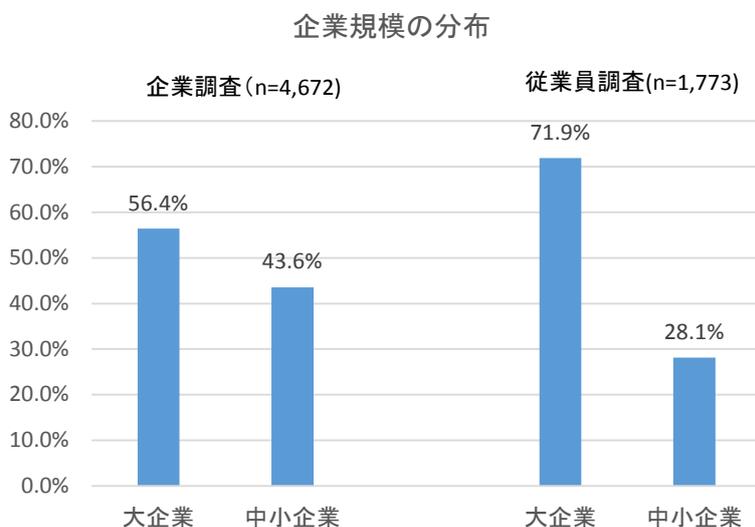


もう一つの比較として企業規模の分布を見る。残念ながら総務省調査の企業規模は大企業と中小企業の区分しか得られないので、比較はこの2区分での比率の比較となる<sup>8</sup>。図2がその結果である。これを見るとかなりの違いがあり、総務省調査に比べてモニター調査では大企業に勤務する人が多い。これは考えてみると当然のことで、総務省の企業調査は企業単位なので、従業員数1万人以上の巨大企業でも調査票は1通である。しかし、モニター調査の従業員調査は個人単位なので、大企業からは人数比に応じて多数のモニターが選ばれる。ゆえにモニター調査がランダムサンプリングであっても、モニターを使った従業員調査では大企業が多くなる。

補正するには、モニター調査で大企業に勤めている人の回答を  $0.564/0.719=0.785$  倍し、中小企業に勤めている人の回答を  $0.436/0.218=1.550$  倍する必要がある。両者をあわせると補正は2倍近くなり無視できない大きさなので、補正が必要である。以下の分析では産業分布の補正と企業規模の補正の両方を行った結果を報告する。

<sup>8</sup> 大企業と中小企業の区分は、総務省調査にならぬ、商業（卸売・小売業）とサービス業は従業員100人以上かどうかで切り、それ以外の産業では300人のところで切った

図2



### 3. 調査結果

#### 3. 1. 効果の有無

1,804人の従業員に対し、データ活用で効果があったかどうかをデータ活用の5分野に分けてたずねた。設問の形式は総務省調査を踏襲する。実際の設問文は次のとおりである。

- 問 あなたが勤務先で体験したデータ活用は効果があったと思いますか。  
データ活用を以下の5分野に分けますので、あなたの体験した事例についてあなた自身のお考えをお答えください。  
その分野で体験した事が無い時は「体験していない」をお選びください。  
#なお、「効果」とは売上や利益に直接結びつかなくても、「効率化した」「物事の把握が容易になった」など広い意味でとらえてください。

分野

- 1 経営全般
- 2 企画・開発・マーケティング
- 3 生産・製造
- 4 物流・在庫管理
- 5 保守・メンテナンス

選択肢

- 1 非常に効果があった
- 2 効果があった
- 3 効果がなかった
- 4 全く効果がなかった

5 わからない

6 体験していない

この設問文は総務省調査とほぼ同じ形になっている<sup>9</sup>。なにをもって効果があったと見なすかが問題であるが、これも総務省調査の文言をそのまま採用した。「・・・売上や利益に直接結びつかなくても、「効率化した」「物事の把握が容易になった」など広い意味で」捉えると言う部分がそれである。

総務省の企業調査と同じように、選択肢のうち「非常に効果があった」「効果があった」を選んだ人を、効果があったと答えたと見なし、全体に占める比率を求める<sup>10</sup>。この比率を企業調査の結果と比較する<sup>11</sup>。

ただし、ここで回答者の知識レベルの相違に注意する必要がある。回答者のスクリーニング条件は勤務先でデータ活用を経験したという事だけなので、そのなかには ICT 担当者だけでなく、それ以外の一般従業員も含まれている。一般従業員と ICT 担当者では持っている情報に大きな差がある。たとえば、営業マンに顧客を回った時の記録を毎日とってもらい、これを全社的に集めてテキスト解析をして経営判断に役立てたとしよう。このデータ活用に効果があったとしてもそれがわかるのは、経営者とシステムを組んだ ICT 担当者であり、現場の営業職員ではない。末端の営業マンからすれば、データ活用とは営業記録を取る手間が増えただけで何の効果があるのかわからないように思えるかもしれない。ゆえに、一般従業員にデータ活用の効果を聞けば、わからないと答える人が多くなるだろう。一方、企業調査で答えたのはその企業の ICT 担当者でありデータ活用についての情報を知りうる立場にある。したがって、比較条件を等しくするためには、回答者の属性を ICT 担当者にそろえる必要がある。

そこで、次の問いを使って回答者を分類した。

問 あなたが勤めている企業では、だれがデータ活用の中心になっていますか？最も近いものを1つ選んでください。また、あなた自身の立場をお答えください。

- 1 専門のデータ解析担当者（データサイエンティスト）
- 2 営業などの現場担当者（営業 SE など含む）
- 3 マーケティング担当者
- 4 情報システム部などの社内システム担当者

<sup>9</sup> 総務省調査の調査票は付録(2)を参照。

<sup>10</sup> 全体に占める比率なので、分母には、5 わからない、6 体験していない、が含まれる。6 の体験していないが含まれることには違和感があるかもしれないが、総務省調査がそのように処理しているため、比較を可能にするために同じ処理を行った（同報告書、p24,注 11）

<sup>11</sup> 企業調査では他にも多くの問いをたてており、それらの問いもサンプリングバイアスが生じうるので、それを併用できれば推定精度をあげられる可能性がある。ただ、他の問いは、昨年扱った画像データのバイト数を聞く等非常に細かく、ウェブモニター調査で聞くには回答者の負担が重すぎる。今回は簡単で、直接的に成功バイアスを測れるこの問いのみを用いた。将来的には複数の問いのバイアスを組み合わせることが有効な推定手段になると予想される。

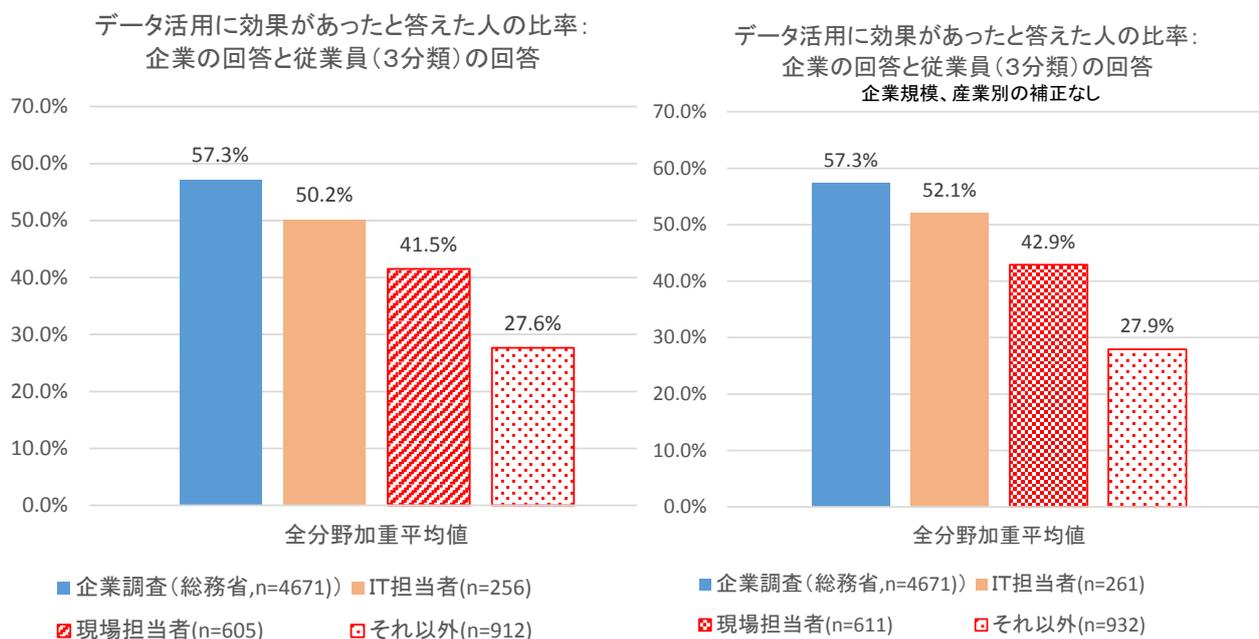
- 5 複数の部署で組んでいるチーム
- 6 外部の企業
- 7 この中にはない

まず、勤務先でデータ活用の中心になっているのはどういう人たちかを上記から選んでもらい、そのうえで回答者自身がどの類型に入るかを選んでもらった。後者の自分自身の立場を使って回答者を類型化する。

選択肢が多いので3つの大分類にまとめる。1と4を「ICT担当者」とし、2, 3, 5を「現場担当者」と呼ぶことにする。7を「それ以外」とする<sup>12</sup>。サンプルの人数は、ICT担当者が256(261)人、現場担当者が605(611)人、それ以外が912(932)人であった(括弧内は未補正の時の人数)<sup>13</sup>。この3類型に分けて、データ活用に効果があると答えた人の比率を求める。結果は図3にまとめた。

図3の左のグラフは企業規模比率と産業別分布を補正するウェイトをつけた場合、右のグラフはウェイト無しの場合である。ウェイトをつけてもつけなくても傾向はほとんど変わらない事がわかる。以下ではより事実に近いと考えられるウェイトを付けた補正後の結果について議論する。

図3



<sup>12</sup> それ以外とは、勤務先でデータ活用はあったが、自分は直接その現場に関わっていない人である。隣接部署や職場の別の人がデータ活用に関わったというケース等が考えられる。6の外部の企業は回答者の立場としてはありえない。この選択肢は設問の前半部分、すなわちデータ活用の担当者が誰かに対する選択肢である。

<sup>13</sup> 産業別のウェイトをつくる時、産業分類が「その他」の人はウェイトがつかれないのでサンプルから除かれ、わずかに人数が減少する

バーの数値はいずれも 5 つのデータ活用分野を合わせた結果である。4 本並んだバーのうち、一番左の 1 本は比較のための総務省企業調査の結果をあらわす。右 3 本が従業員調査の結果で、左から順に ICT 担当者、現場担当者、その他の 3 類型である。

まず従業員調査の 3 類型の結果を見ると、バーは右下がりである。すなわち、ICT 担当でない人の方がデータ活用に効果を見出していない。これは ICT 担当でない人は、データ活用の効果を知るだけの情報を持っていなかったからと解釈できる<sup>14</sup>。実際、効果の有無がわからないと答えた人の比率は ICT 担当者では 30% 程度であるのに、右 2 本の現場担当者とそれ以外では 40%～50% にも達しており、知識あるいは情報のレベルに大きな差があることをうかがわせる。

企業調査と比較するなら、企業調査と同質的なサンプルにすべきである、企業調査で回答しているのは ICT 担当者と考えられるので、比較すべきは従業員調査のほうでも ICT 担当者が望ましいだろう。すなわち、図 3 の左端の 2 本のバーを比較したときの差がサンプリングバイアスによるずれと見なせる。これを計算すると

$$57.3 - 50.2 = 7.1$$

で 7.1% と見積もれる。すなわちサンプリングバイアスにより、効果があったという答えは 7.1% ポイント増えていると解釈できる。

### 3. 2. 比較可能性の検討

このように解釈をするためには、繰り返し述べるように二つの調査で回答している人が同質的、すなわち統計集団としてほぼ等しいという前提が必要である。すでに産業分布と企業規模については検討したが、他の属性もできるだけ一致している事が望ましい。すなわち、回答者の年齢、学歴、勤務先企業の規模、役職、所得、さらに ICT 担当者としての知識レベルなどの属性分布が等しくなっていることが望まれる。結果の信頼性を高めるため、この点を検討しておく。

まず、ICT 担当者としての知識あるいは情報レベルが近いかどうか見てみよう。企業調査の ICT 担当者は企業が選んでいるので、従業員調査でスクリーニングした ICT 担当者よりも優秀で、知識レベルあるいは情報量で優れているということはないだろうか。これを調べるため、前出の効果の有無を問う設問に対してわからないと答えた人の比率を比べてみる。もし、知識・情報レベルに差があるなら、わからないという答えの比率に差が出てくるはずである。これを求めると、企業調査でわからないと答えた人の比率は 28.9%、従業員調査の ICT 担当者では 28.4% であった。極めて近い値であり、二つの調査での ICT 担当者の知識

---

<sup>14</sup> ただし、他の解釈もできることには留意する必要がある。それは担当者が自己評価を甘くしがちになるというバイアスである。ICT 担当者は自分がシステム設計をした以上、自己正当化のために評価は甘くなる可能性がある。ICT 担当者の評価は自己正当化のために高めに、その他の人の評価は情報不足のために低めになっているとすれば、真の値は両者の中間のどこかにある。すなわちデータ活用で効果があった企業の真の比率は、図 3 の 28.2% から 50.7% の間のどこかになる。本稿はサンプリングバイアスに関心があり、真の値には関心が無いのでこの問題にはこれ以上立ち入らない。

情報レベルに、特に差は無いようである。

これ以外の年齢や学歴、所得などの属性については、残念ながら総務省の企業調査の回答者の属性がわからないため、二つの調査の回答者の属性が似ているかどうかを直接確かめることはできない。ただ、従業員調査の方の回答者の属性はわかるので、一般的な企業のICT担当者の属性と極端に離れてはいないかどうかを検討することはできる。そこで、従業員調査での261人のICT担当者の年齢、学歴、役職、企業規模（人数）、個人年収の分布を見てみよう。表2がその結果である。

表 2

従業員調査でのIT担当者261人の属性分布

年齢	学歴	役職	企業規模	個人年収
20代	高卒	一般社員	0～4人	～100万円未満
30代	専門学校卒	主任・係長	5～9人	100～200万円未満
40代	短大・高専卒	課長クラス	10～19人	200～300万円未満
50代	大学卒	部長クラス	20～29人	300～400万円未満
60代	大学院卒	経営者・役員	30～49人	400～500万円未満
		個人事業主	50～99人	500～600万円未満
		その他	100～299人	600～700万円未満
			300～999人	700～800万円未満
			1,000～1,999人	800～900万円未満
			2,000～4,999人	900～1000万円未満
			5,000人以上	1000～1500万円未満
				1500万円以上

n=261

年齢は40代～50代があわせて6割と多く、ICT担当者としては末端ではなく、責任のある地位にあることがわかる。従業員調査はネットでのモニター調査なので入社したばかりの新人ばかりではないかという危惧を持つ人がいるかもしれないが、そういうことはない。学歴も大卒が6割、大学院卒が1割いて高学歴であり、ICT担当者としては妥当である。役職は一般社員から部長まで広く分布している。勤務先企業の規模は数百人以上の中堅以上の企業が多く、5,000人以上の大企業に勤めている人も2割程度存在する。個人年収は三百万円以上に幅広く分布しており、1,000万円以上の高給取りも15%存在する。ウェブモニターは一回数百円でアンケート調査に応じるので、年収200万～300万で20～30代の若い人が多いと推測する人もいるかもしれないが、ICT担当者とは異なることではない。

このようなプロフィールは一般的な企業のICT担当者のプロフィールとそれほど違っているとは思えない。企業調査でのアンケートは経営者宛てであってもICT部門に回され、ICT部門の長はそれを誰かに命じて回答させる。命じられた回答者は普通のICT担当者だとすれば、そのプロフィールは表2のプロフィールと大きく異なることは考えにくい。

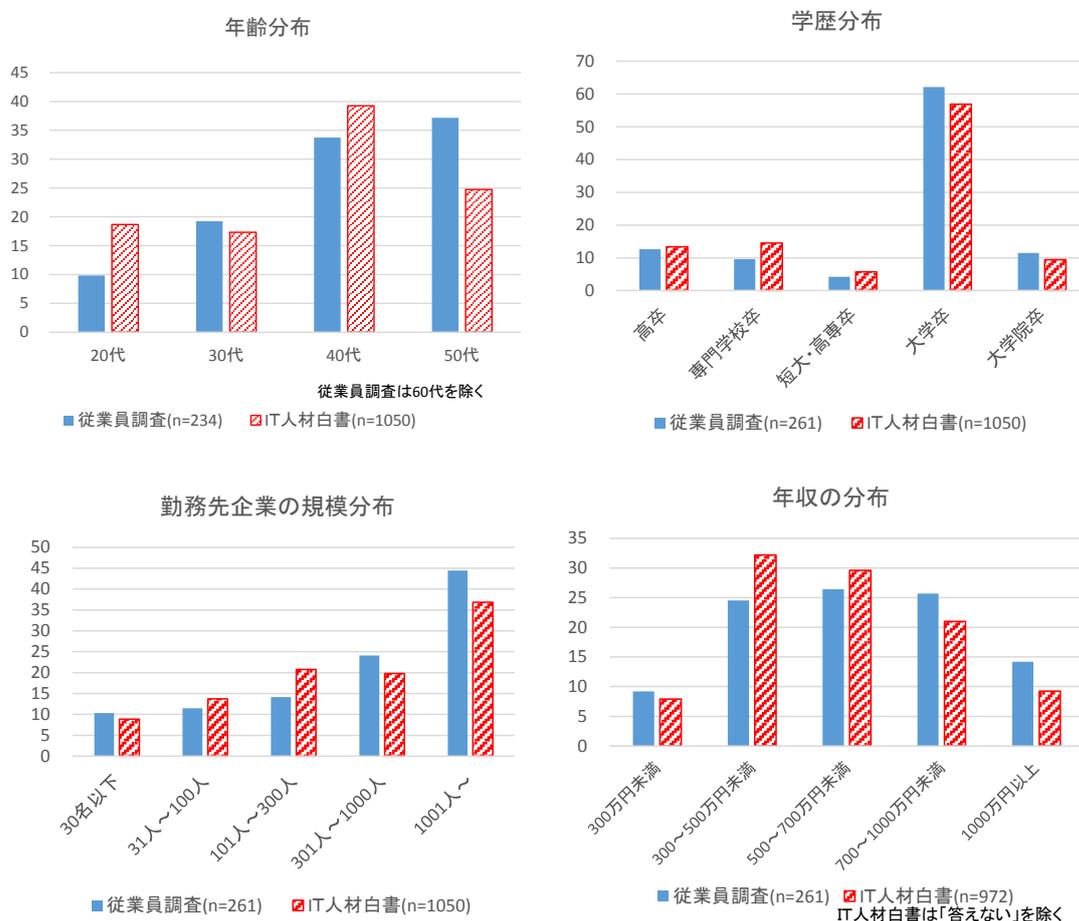
これを裏付ける間接的な証拠として、別の調査でのICT担当者の属性分布と比較してみよう。情報処理推進機構（IPA）は毎年1,000人程度のICT技術者にアンケート調査を行い、その結果を『IT人材白書』として公表している<sup>15</sup>。彼らの年齢、学歴、企業規模、そ

<sup>15</sup> 『IT人材白書2017』のなかの「IT技術者動向調査（IT人材個人向け）」に集計され

して年収の分布がわかっているので、それと比較してみよう。図4がその結果である。塗り潰したバーが我々の従業員調査での分布で、斜線のバーがIT人材白書の分布である<sup>16</sup>。どの属性についても分布がよく似ている事がわかるだろう。分布だけ見せられてどちらの調査の結果か答えることは困難である。

属性の分布すなわちプロフィールが似ており同じ統計集団と見なす事ができれば、ここで得た値のずれ7.1%はサンプリングバイアスのためという解釈が可能である。本稿の分析はこの二つのサンプルが統計集団として同質であるという前提に決定的に依存している。同質性はここで示した事実である程度担保されたと考えるが、それでも前提であることに変わりなく、この点は本論文の限界になることを確認しておきたい。

図4  
 ICT担当者の属性分布：本調査とIT人材白書の比較



ている。<https://www.ipa.go.jp/jinzai/jigyow/data.html> 2018/10/3 確認

<sup>16</sup> なお、横軸の切れ目の区間が一致するように、細分化されたデータを取りまとめた。その際、分布を比較するのが目的なので、そもそも選択肢にないものは除いた。具体的には、IT人材白書では50代までしか調査対象としていないので、従業員調査も50代までのデータだけを使って分布を描いた。年収についてIT人材白書では「答えたくない」の選択肢があったが従業員調査には無いため、これもとり除いて分布を描いた。

比較可能性の検討の最後に、もうひとつ別の比較可能性の検討をしておく。それは公式見解と実態（本音）の乖離の問題である。企業がデータ活用に失敗した時、公式見解としては成功したことになるという事は十分に考えられる。データ活用の成功・失敗は外部からは見えにくいので、経営者には責任問題を恐れて表向きには成功した事にしようとする誘因がある。企業調査でアンケートに回答を記入する人は会社を代表する立場で回答するので、会社の公式見解を答えるだろう。一方、従業員アンケートでたずねているのは、回答者個人としての評価いわば本音であり、会社としての公式の評価ではない。ここから生じるずれは、いわば公式見解と本音のずれであり、サンプリングバイアスとは無関係のずれである。したがって、サンプリングバイアスを得るためには、今回得た 7.1%と言う数値のずれから、公式見解と実態のずれを除いておく必要がある。

公式見解と実態のずれがどれくらいあるかを見積もるため、前掲の問いのすぐあとに、会社としての見解をたずねる問いを設けた。設問の文章は次のとおりである。

「では、そのデータ活用の効果について会社の見解はどうだったでしょうか？あなたの意見と一致していてもいなくてもかまいません。」

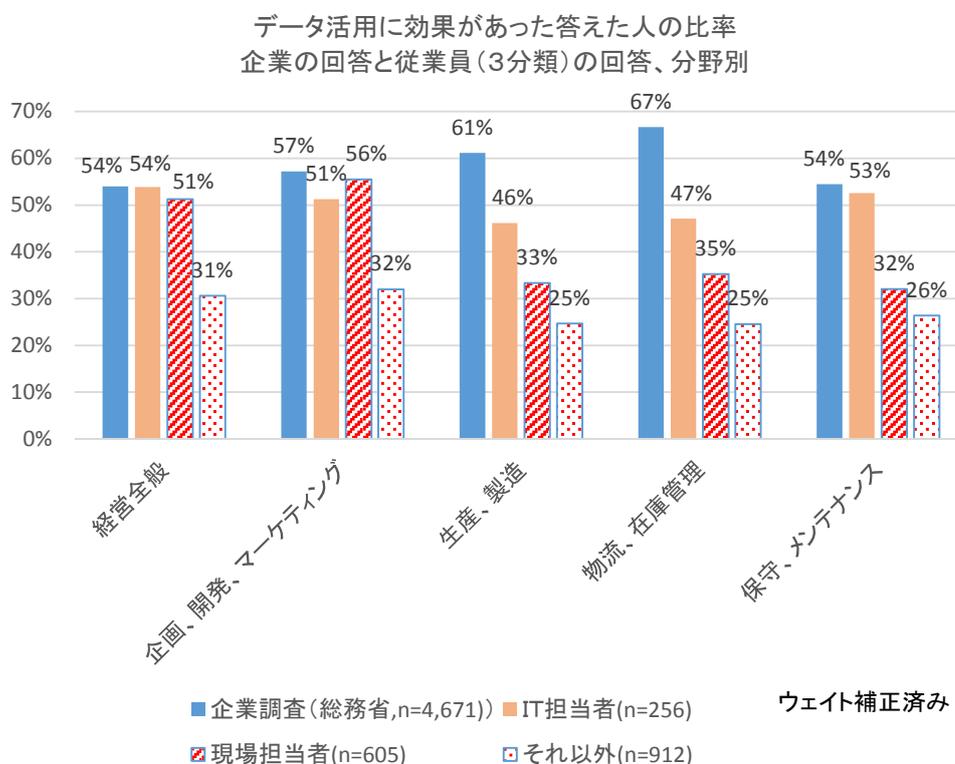
選択肢は同じである。この問いは回答者個人の見解ではなく、会社の見解をたずねており、これで異なる結果が出れば、会社の見解と個人の見解に差があったことになる。これを ICT 担当者にたずね、会社見解で効果があったと答えた人の比率を求めると 49.7%であった。図 3 での個人見解では 50.2%なのでほとんど差が無く、違いは誤差の範囲である。したがって、公式見解と実態（本音）のずれの影響はほぼないと言ってよく、7.1%はそのままサンプリングバイアスと解釈できる。

### 3. 3. データ活用分野別の違いとバイアスの実践的評価

ここで得たサンプリングバイアスの 7.1%は、5つのデータ活用分野の平均値である。これをデータ活用分野別に見てみよう。図 5 がそれで、先の図 3 のグラフをデータ分野別に分けて集計したものである。5つのデータ活用分野別に 4本のバーが立っている。このうち左端の2本のバーが企業調査と従業員の ICT 担当者の回答であり、この2本の差がサンプリングバイアスと見なせる。

図 5 を見ると、1) 経営全般、2) 企画・開発・マーケティング、そして5) 保守・メンテナンスの3分野では、企業調査と従業員調査にはほとんど差が見られない。しかし、3) 生産・製造と4) 物流・在庫管理では、10%以上の差が発生している。したがって、サンプリングバイアスが発生しているとすればこの2分野ということになる。

図 5



なぜ、この 2 分野のみサンプリングバイアスが生じたのかは、実際の企業にインタビューサーベイしなければわからない。本稿ではそこまでの用意は無いが、これを成功バイアスと見た場合、一定の解釈が可能な事を示す事が出来る。それはこの 2 分野は失敗を自覚しやすい分野だからという解釈である。

この 2 分野と他の 3 分野には違いがある。それは生産・製造、ならびに物流・在庫管理は成果を数値で測りやすい分野だということである。生産・製造なら生産個数で、物流・在庫管理も処理件数や遅延率、在庫量等で定量的に成果を把握できる。ゆえにデータ活用を行った場合、効果があったかなかったかがあからさまにわかる。大きな費用を掛けてデータ活用を試みたのに生産量は増えず、在庫も減らせずと効果がなかったことがわかれば、企業としてはショックであろう。その挫折感ゆえに調査に答えたくないという誘因が発生しうる。一方、経営全般、企画・開発・マーケティング、そして保守・メンテナンスは、成果が定量的に測りにくい。ゆえに、効果が無かったとしてもそれがすぐにはわからず、失敗したという自覚が生じにくい。それゆえ、調査を躊躇させるだけの挫折感を持つまでにはいたらないと考えられる。このように、2 分野でのみずれが生じたことは、このずれが成功バイアスから生じたと考えると一定の解釈が可能である。むろん、これは解釈であり、この解釈が正しいかどうかはインタビューサーベイ等で今後明らかにする必要がある。

ここで得られたサンプリングバイアス 7.1% 値の評価をしておこう。この値はバイアスとして大きいのだろうか小さいのだろうか。

評価のため、この 7.1% がサンプリングバイアスのために生じたとしたとき、成功企業と

失敗企業の回答傾向にはどれくらい差があるかを考える。成功企業は他の企業よりも  $1+r$  倍だけ回答する率が高くなるとしよう。成功企業の比率の真の値を  $R$  とすると、調査回答者の中の成功企業の比率は、 $R(1+r)$  となるので、観察される成功企業の比率は  $R(1+r)/(R(1+r)+(1-R))$  となる。この式に図3の従業員調査の値  $R=0.502$  を入れ、計算結果が観察された総務省の値  $0.573$  に等しくなるように、 $r$  を定めると、およそ  $0.3$  となる。すなわち、成功した企業はそれ以外の企業よりも  $1.3$  倍多めに回答するということであり、この  $1.3$  がバイアスの大きさを表す。

このバイアスの推定値の実践的な評価を試みる。  $1.3$  という値は大きいとも言えるし、小さいとも言える。  $3$  割多めに答えるとすれば、バイアスとしては確かに存在していると言わなければならない。また成功の比率が  $7.1\%$  ポイント多めにでるのも無視できない。その意味では大きな値である。

しかし、効果があったかどうかという元の値が  $5$  割程度なので、 $7.1\%$  は比率にすると  $1.5$  割である。  $1$  割強程度のずれは推定誤差の範囲とも考えられ、その意味では小さいとも言える。さらに実践的に重要なのは回帰分析などの相関関係の推定であるが、このバイアスの回帰分析への影響は限定的である。なぜなら、回帰分析でバイアスを補正するならば、成功企業に  $1/1.3=0.77$  のウェイトを乗じることになるが、このウェイトはそれほど大きくないからである。この程度のウェイトなら回帰係数の推定値自体は変わっても、元の推定式ではっきりと有意と出ていた結果が、有意でなくなるようなケースは限られる。無論、ぎりぎり  $5\%$  有意のような結果はひっくり返りうるので留意が必要ではある。しかし、 $p$  値が  $1\%$  以下のようなはっきりした結果は簡単には覆らないだろう。ゆえに、回帰分析の定性的な結果の大勢は維持されると予想される。分析結果に深刻な疑念をいだかせるほどのバイアスではなかったという点で、今回の推定結果は調査への信頼性を支持する結果となっている。

#### 4. バイアスに影響と与える要因

本稿で考えるサンプリングバイアスに影響を与える要因は何だろうか。たとえば企業規模は大きい方がバイアスが起りやすいのだろうか、それとも中小企業の方が起りやすいのだろうか。このようなバイアスの決定要因について考えてみる。

今回の調査では個別企業単位のバイアスはわからないので、個々のデータを見ても何もわからない。そこでデータを小さいグループに分割し集団として分析する。総務省の企業調査では、データ活用の  $5$  つの分野と  $9$  つの産業別を掛け合わせた  $45$  個のグループについて、効果があった企業の比率を計算している<sup>17</sup>。そこで、従業員調査でも同じように分野×産業の  $45$  個のグループに分けて、効果があったかどうかの比率を計算する。両者の差をとることで、バイアスの大きい場合と小さい場合のデータがつかれる。

式で表すと、企業調査で産業  $j$  に属する企業の中でデータ活用分野  $i$  で効果のあったと答えた企業の割合を  $y_{ij}$ 、従業員調査で産業  $j$  に勤務している ICT 担当者人の中で分野  $i$  について効果のあったと答えた人の割合を  $x_{ij}$  とする。 $y_{ij}-x_{ij}$  がバイアスの大きさを表す。これを

<sup>17</sup> 同報告書、p24、図表 4-18。なお、総務省調査の産業分類は  $10$  分類であるが、農林水産業のデータ数が少なくセルに分けることができない。そこで農林水産業を除く  $9$  産業について計算する。

被説明変数とし、企業規模など説明変数に回帰する。推定式は以下のとおりである。

$$y_{ij} - x_{ij} = a + bScale_{ij} + cAge_{ij} + dTurnOver_{ij} + e_i F_i + f_j I_j + \varepsilon_{ij}$$

説明変数の候補としては、企業規模 Scale (人)、企業年齢 Age (年)、転職率 TurnOver (%) をとった。以下、期待される符号について簡単に考察する。

企業規模については、符号は両方の可能性がある。本稿で考えているバイアスは、調査内容であるデータ活用に失敗した企業が回答をしぶることで生じる。失敗して回答をしぶるというのはすぐれて心理的な現象であり、そのような心理的要因に左右されやすいのは組織として対応する大企業ではなく、属人的に対応する中小企業と考えられる。だとすれば中小企業ほどバイアスが大きくなる。一方、大企業ほど入念に準備してデータ活用を試みるので、それが失敗した時の心理的な挫折感は大きいかもしれない。だとすれば大企業ほど回答をしぶり、バイアスが大きくなる。両方の可能性があり、符号は事前には予想できない。

企業年齢も同様に符号は予想しがたい。企業年齢が若いということは、まだ経営が安定せず、浮き沈みが激しい状態なので、データ活用プロジェクトが失敗しても気にせず、故にバイアスは小さいかもしれない。一方、企業年齢が若いと組織がしっかりしておらず、調査に属人的に対応するとすればバイアスは大きいかもしれない。

転職率は、従業員調査で新しく入社した人が3年以内に離職する割合を答えてもらった。転職率が高いと、従業員の忠誠心が低くなり、プロジェクトが失敗しても気にせずに回答してくれる可能性がある。もしそうだとすれば符号はマイナスになる。Fはデータ活用分野のダミー、Iは産業ダミーであり、分野と産業特有の要因を取り除く。

3変数の記述統計は表3に掲げた。グループごとの平均値なので分布の範囲が小さくなっている。たとえば企業規模の最大値は4,000人強、最小値は1,000人弱である。個別企業としてはこの範囲外の大企業あるいは中小企業もいるが、ここでの回帰結果をそこまで外挿することはできないことに留意されたい。

表3

			N	平均値	標準偏差	最小値	最大値
バイアス	$y_{ij} - x_{ij}$	比率	45	0.064	0.229	-0.373	0.860
企業規模	scale	人	45	2087	998	880	4083
企業年齢	age	年	45	39.2	10.6	20	59
転職率	Turnover	%	45	11.6	3.9	7	21

推定結果は表4のとおりである。3つの説明変数を見ると、有意なのは企業規模のみで、符号はマイナスである。企業規模が大きいほどバイアスは小さくなるので、バイアスが大きいのは規模の小さい企業という事になる<sup>18</sup>。

<sup>18</sup> 企業規模の値-0.000513は、企業規模が1,000人増えるとバイアスの値が0.5だけ減ること意味する。バイアスは比率なので変化の取りうる値は-1~1であり、そこから考えると係数の値が大き過ぎるという疑問が出るかもしれない。しかし、この回帰は産業別の平均値についての回帰なので、企業規模の範囲は表3に見るように880~4,083にとどま

企業規模が小さい場合、意思決定が属人的であり、心理的な要因の入り込む余地が大きいからと解釈できる。大規模な企業の場合、常時多くの調査依頼が舞い込んでおり、組織的に対応するので、外部からの調査依頼に対してルーティーンが出来ていることが多いだろう。その場合、失敗した事は語りたくないという個人の心理的な要因の入り込む余地は少なくなると考えられる。これに対し規模の小さい企業ではそのようなルーティーンが無く、経営者あるいは ICT 担当者の個人的な意思決定の余地が大きい。その場合は個人の心理的な要因が大きく影響を与える。失敗したプロジェクトの調査に答えるのがおっくうになり、後回しにするうちに忘れてしまうという事が起こりうると解釈できる。

表 4

VARIABLES	データ活用に効果あり： 企業調査と従業員調査 の差
企業規模（従業員数） （人）	-0.000513** (-2.378)
企業年齢（設立からの年数） （年）	0.0189 (1.167)
離職率（新卒の3年以内離職率） （%）	-0.0854 (-1.370)
<データ活用分野ダミー（4個）>	included
<産業ダミー（8個）>	included
Constant	1.474** (2.511)
Observations	45
R-squared	0.735
Adjusted R-squared	0.597

z-statistics in parentheses

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

## 5. 結論と含意

企業サーベイの場合、調査テーマに関して失敗した企業は回答をしぶり、成功した企業は積極的に回答するというバイアスが考えられる。このバイアスがどれくらいかをデータ活用を事例にして推定を試みた。手法は、同じ問いを従業員に対して行い、すでにある企業調

り、それ以外に外挿はできない。また企業規模の効果には非線形性が見られ、次第に効果が逡減する。ゆえに推定された係数はあり得ないほど大きな値というわけではない。

査の結果と比較する方法である。従業員は匿名なら勤務先の企業がデータ活用に失敗していても回答すると予想されるので、回答のずれはサンプリングバイアスから生じたと解釈できる。この方法で比較したところ、データ活用が成功したと答えた人の比率は、企業調査の方が従業員調査より 7.1%ポイント程度高かった。これがサンプリングバイアスによって生じたとすると、成功企業の方が失敗企業よりも 3 割程度多めに回答したことになる。補正ウェイトに換算すると、回帰分析をするときは成功企業に 1/1.3 すなわち 0.77 程度を乗じることになる。この程度のウェイトで結果が大きく変わる例は限られるので、本稿で扱った企業調査に関しては成功バイアスがあるとしてもそれほど深刻なものではないだろう。

以上の結果を踏まえ、本稿の含意について考察する。まず、第一に確認すべきは本稿が扱ったのは、総務省調査のデータ活用という一事例であり、すぐに一般化できるわけではないことである。一般的な結論を得るためには、同様の調査を事例を変えて何度か繰り返す必要があり、本稿はそのための第一歩にとどまる。そこで含意として、仮に一般化するとどのような変化がありうるのかを考察する。

本稿でのバイアス推定は、総務省のデータ活用の調査について行ったものである。データ活用はビジネス分野であり、かつ調査主体は官庁である総務省である。この条件が異なればバイアスの大きさは変わってくると予想される。

まず、データ活用のようなビジネス分野ではなく、より企業の評判・名声に関わるような社会的・政治的に敏感な問題ではサンプリングバイアスは大きくなるだろう。たとえば、個人情報保護への取り組み、データ漏えいやハッキングの実態、社内での女性の登用・処遇、セクハラ・パワハラ問題への対策、LGBT への配慮、不正会計への監査体制、反社会勢力との関わりなどでは、取り組みの遅れ自体が社会的批判を伴う。データ活用のようなビジネス事例なら経営だけの問題であるが、これらの敏感な話題は社会的批判を伴う分だけ、取り組みが遅れた企業、あるいは対策がうまくいっていない企業は回答をよりためらうようになるだろう。ゆえに、回答するのは対策がうまくいった企業だけになる度合いが、単なるビジネス活用事例より強まると予想される。サンプリングバイアスの推定値は本稿の得た 1.3 より大きくなるだろう。

また、調査主体が官庁ではなく民間のシンクタンクや研究者である場合、そもそもの回収率が低くなりやすい。これからもつきあいのある官公庁の調査ならしぶしぶでも答えるが、民間の調査会社あるいは研究者からの初めての調査依頼の場合、無視するのは簡単だからである。その場合、対策がうまくいって進んで答える企業とそうでない企業の回答率の差は広がるだろう。そうだとすれば、この場合も成功バイアスは 1.3 よりも大きくなる。

したがって、本稿でとりあげた総務省によるデータ活用調査は、「官庁」が行う「ビジネス分野」でのアンケートだったという点で、企業アンケートとしては比較的バイアスが少ないケースだったと考えられる。ゆえに本稿で得た結果の適用は、そこに限定した方が安全である。すなわち、バイアスが 1.3 程度と比較的少ないという本稿の結論は、官庁が行うビジネス分野での効果の調査に限定した方が良い。

視点を変えると、これと真逆の場合ではサンプリングバイアスがより大きくなるという点も含意として重要である。今回の調査は比較的サンプリングバイアスが小さいケースと述べたが、そのような場合でさえ 1.3 倍のバイアスがあったと見することもできる。そうだとすると、サンプリングバイアスがこれより大きくなるようなケース、たとえば官庁ではない

民間の研究機関や研究者がセクハラ、反社会勢力との関わり等の敏感な問題について企業アンケートを行った場合、サンプリングバイアスは1.3をはるかに超え、深刻化する恐れがある。サンプリングバイアスが2倍を超えるようなケースも考えられないわけではない。本稿の推定結果は、そのような恐れがある事に警鐘をならしているとも解釈できる。

本稿の第二の含意として、サンプリングバイアスが大企業で小さく、中小企業で大きいという結果の実践的な意義が指摘できる。サンプリングバイアスが大企業と中小企業でどちらが大きいかは事前の段階では予想がつかないが、本稿の結果は中小企業の方が大きいことを示している。無論、この結果が一般的に成立するかどうかはまだわからない。が、仮にこれが一般的な傾向として成立するとすれば、サンプリングバイアスについて、限定的であるが対策がたてられる。

たとえば、セクハラや反社会勢力との関わりなどサンプリングバイアスが大きいテーマで企業調査したとしよう。サンプリングバイアスが気になっても、なにも情報がなければ補正のしようがない。しかし、大企業のほうがサンプリングバイアスが少ないということがわかっているなら、サンプルを企業規模別に分けて推定を行う事で、バイアス補正の手掛かりを得ることができる。たとえばあるテーマで回帰分析を行って全データでは1.5という推定値を得たとする。この時大企業だけで計算した推定値が1.1、中小企業だけの推定値が1.8なら、大企業の方がサンプリングバイアスが少ないのであるから、バイアス補正すれば真の値は1.5よりも小さくなるだろうということが予想できる。このような方法でサンプリングバイアスの補正方向についての手掛かりが得られる。ここで得た企業規模の効果に一般性があるかどうかは他の事例を調査しないといけないので結論付けるわけにはいかない。しかし、このように相関関係を出す事でバイアス補正への手掛かりをえるというのは実践上の有意義なアプローチであろう。

## 参考文献

Groves, Robert M. and Emilia Peytcheva. 2008. "The impact of nonresponse rates on nonresponse bias: A meta-analysis," *Public Opinion Quarterly* 72 (2): 167-89.

Holmes, Thomas J. and James A. Schmitz, JR. 1996, "Nonresponse Bias and Business Turnover Rates: The Case of the Characteristics of Business Owners Survey," *Journal of Business & Economic Statistics*, Vol. 14, No. 2 (Apr., 1996), pp. 231-241

Peytchev, Andy, 2013, "Consequences of Survey Nonresponse," *The Annals of the American Academy of Political and Social Science*, Vol. 645, *The Nonresponse Challenge to Surveys and Statistics (JANUARY 2013)*, pp. 88-111

Rogelberg Steven G. and Jeffrey M. Stanton 2007, "Introduction Understanding and Dealing With Organizational Survey Nonresponse," *Organizational Research Methods* Volume 10 Number 2 April 2007 195-209

Tomaskovic-Devey, Donald, Jeffrey Leiter, and Shealy Thompson, 1994,  
Organizational Survey Nonresponse, Administrative Science Quarterly, 39  
(1994): 439-457

Werner, Steve, Moira Praxedes and Hyun-Gyu Kim, 2007, The Reporting of  
Nonresponse Analyses in Survey Research, Organizational Research Methods  
Volume 10 Number 2 April 2007 287-295

情報通信総合研究所、2015、「ビッグデータの流通量の推計及びビッグデータの活用実  
態に関する調査研究」  
[http://www.soumu.go.jp/johotsusintokei/linkdata/h27\\_03\\_houkoku.pdf](http://www.soumu.go.jp/johotsusintokei/linkdata/h27_03_houkoku.pdf), 2018/10/04  
確認

総務省, 2015, 情報通信白書 平成 27 年版

情報処理推進機構(IPA), 2017 『IT 人材白書  
2017』、<https://www.ipa.go.jp/jinzai/jigyouback.html> 2018/10/04 確認

## Appendix

### (1) スクリーニングの設問

データ活用を経験した 1,804 人の抽出の仕方は以下のとおりである。下記の設問を 17,122 人に対して行った。

あなたが勤めている企業でのデータ活用状況についてお聞きします。自社のデータ、データ会社のデータ、他社(データ会社以外)のデータについて、それぞれ当てはまるものを1つ選んでください。

<マトリクス個別設問>
【自社のデータ】
【データ会社のデータ】
【他社(データ会社以外)のデータ】
<マトリクス選択肢>
1=データ活用についての検討はされていない
2=検討中ではあるが、データ活用はまだされていない
3=最近1年以内に、データ活用が行われ始めた
4=1年以上前から、データ活用が行われている
5=企業に勤めていない
6=わからない
999=無回答

この問いに対して、データの種類を問わずに、

3 最近1年以内に、データ活用が行われ始めた

4 1年以上前から、データ活用が行われている

のどちらかを選んだ人をデータ活用の経験者として選んだ。呼びかけたモニターの総数は 17,122 人で、うち 3, 4 を選んだデータ活用の経験者は 1,804 人である。

### (2) 総務省調査での効果の設問

総務省調査でのデータ活用の効果を問う設問は下記のとおりである。

Q17

貴社では、各領域でデータを活用することによって、どのくらい効果がありましたか。 必須 ひとつだけ

\*効果については、売上高や営業利益の増加には結びついていなくても構いませんので、「効率化」、「物事の把握が容易になった」など広い意味での効果としてお答えください。

回答方向 ➔	非常に 効果があった	効果が あった	効果が なかった	全く効果が なかった	わからない
経営全般	<input type="radio"/>				
企画、開発、マーケティング	<input type="radio"/>				
生産、製造	<input type="radio"/>				
物流、在庫管理	<input type="radio"/>				
保守、メンテナンス	<input type="radio"/>				

情報通信総合研究所(2015)「ビッグデータの流通量の推計及びビッグデータの活用実態に関する調査研究

(掲載決定日：平成 31 年 4 月 11 日/オンライン掲載日：令和元年 6 月 14 日)