

令和 2 年度 統計委員会評価分科会
審議結果報告書
(第 8 回 ~ 第 11 回審議分)

令和 3 年 6 月 23 日
総務省統計委員会

目次

【本編】

I 検討の経緯

- 1 経緯.....3
- 2 審議体制及び審議経過.....4

II 審議結果

- 1 社会教育調査(民間体育施設)の母集団の明確化及び欠測値補完の見直しについて...6
 - (1)取組の現状.....6
 - (2)評価及び課題解決に向けた今後の取組の方向性.....6
- 2 建設工事統計調査(施工調査)の欠測値補完の見直しについて.....7
 - (1)取組の現状.....7
 - (2)評価及び課題解決に向けた今後の取組の方向性.....8
- 3 経済産業省生産動態統計調査の欠測値の補完について.....8
 - (1)取組の現状.....8
 - (2)評価及び課題解決に向けた今後の取組の方向性.....10
- 4 欠測値の補完に係る主な方法等について.....10
- 5 令和3年度における評価分科会の取組について.....11
 - (1)精度検査報告書のフォローアップ.....11
 - (2)統計調査における母集団情報の利用状況について.....11

【資料編】

- 資料1 文部科学省が実施する社会教育調査のうち、民間体育施設に係る母集団について(追加検証).....15
(令和2年10月30日第8回評価分科会 資料2)
- 資料2 建設工事施工統計調査における欠測値補完の見直しについて(案).....23
(令和2年10月30日第8回評価分科会 資料3)
- 資料3 経済産業省生産動態統計調査の欠測値補完方法の分析結果を受けて.....33
(令和3年4月15日第10回評価分科会 資料1)
- 資料4 経済産業省生産動態統計調査欠測値補完法の妥当性に関する分析結果報告書(追加検証).....35
(令和3年4月15日第10回評価分科会資料 資料2)
- 資料5 欠測値の補完に係る主な方法等について.....49
(令和3年3月26日第9回評価分科会資料を分科会審議等を踏まえ修正)

【参考URL】

- (1) 2018年3月30日 平成28年度統計法施行状況に関する審議結果報告書
(統計精度検査関連分)

<https://www.soumu.go.jp/main_content/000543217.pdf>

- (2) 評価分科会 (第1回～第11回)

<http://www.soumu.go.jp/main_sosiki/singi/toukei/hyokabunkakai/kaigi.html>

【本 編】

検討の経緯

1 経緯

平成30年3月6日に閣議決定された「公的統計の整備に関する基本的な計画」（以下「第Ⅲ期基本計画」という。）によって、統計委員会に統計の品質面や統計作成の技術面等を改革する評価チームを設置して検討を行うこととなったことを受けて、平成30年8月31日に統計委員会令が改正され、統計法の施行に関し、主として統計技術の観点から評価を行い、その結果に基づき意見を述べる評価分科会が統計委員会に設置された。

評価分科会は、平成30年11月28日に第1回分科会を開催し、統計委員会の精度検査報告書^(注)により分科会の検討事項として提示されていた同報告書のフォローアップを検討事項として採用し、同報告書提言のうち、「建設工事統計調査（施工調査）の欠測値補完の見直し」、「建築着工統計調査の補正調査の見直し移行期の集計方法」及び「経済産業省生産動態統計調査の欠測値の補完」について4回の審議を行い、令和元年9月20日に「令和元年度統計委員会評価分科会審議結果報告書（第1回～第4回審議分）」をとりまとめた。

続いて、令和2年1月～3月に第5回～7回評価分科会を開催し、精度検査報告書提言のうち、「法人企業統計調査の欠測への対応に関する検討」、「民間給与実態統計調査の欠測への対応に関する検討」、「造船造機統計調査の欠測への対応に関する検討」、「自動車輸送統計調査（バス関連）の欠測への対応に関する検討」、「自動車輸送統計調査（貨物営業用自動車）の欠測への対応に関する検討」、「建築着工統計調査の補正調査の見直し」、「賃金構造基本統計調査の欠測への対応に関する検討」及び「薬事工業生産動態統計調査の欠測への対応に関する検討」について審議を行い令和元年度統計委員会評価分科会審議結果報告書（第5回～第7回審議分）」をとりまとめた。加えて、欠測値への対応に関する各府省研究成果について、内閣府、総務省・独立行政法人統計センター及び経済産業省より報告を受け、各府省研究成果の共有化を図った。

更に、令和2年10月～3月に第8回～11回評価分科会を開催し、精度検査報告書提言のうち、「社会教育調査（民間体育施設）の母集団の明確化及び欠測値補完の見直し」、「建設工事統計調査（施工調査）の欠測値補完の見直し」、「経済産業省生産動態統計調査の欠測値の補完」について審議を行うとともに、これまでの各統計調査における欠測値への対応に関する審議結果等も踏まえ、欠測値の補完に係る主な方法等について審議を行った。

(注) 統計委員会では、統計精度向上のための取組として総務省が行った検査結果（「平成 28 年度統計法施行状況報告 統計精度検査編」）の報告を平成 29 年 11 月 21 日に受けて、当時の横断的課題検討部会及び統計の精度向上及び推計方法改善ワーキンググループを開催して同報告に対する審議を実施し、検査結果を踏まえた個別統計の改善方策及び改善を行う上での留意事項を整理するとともに、精度向上のために取り組むべき事項についての提言を「平成 28 年度統計法施行状況に関する審議結果報告書（統計精度検査関連分）」（以下「精度検査報告書」という。）として平成 30 年 3 月 30 日に取りまとめた。同報告書のフォローアップについては、統計委員会から検討事項として評価分科会に提示され、評価分科会において検討事項として採用し、審議が行われることとなった。

2 審議体制及び審議経過

審議は、令和 2 年 10 月から令和 3 年 6 月までに 4 度開催して行った。

【評価分科会の構成員】

- ・分科会長 椿 広計（大学共同利用機関法人情報・システム研究機構理事、情報・システム研究機構統計数理研究所長）
- ・分科会長代理 岩下 真理（大和証券チーフマーケットエコノミスト）
- ・臨時委員 久我 尚子（株式会社ニッセイ基礎研究所生活研究部主任研究員）
- ・臨時委員 山本 渉（電気通信大学大学院情報理工学研究科准教授）
- ・臨時委員 美添 泰人（青山学院大学名誉教授）
- ・専門委員 神林 龍（一橋大学経済研究所経済制度・経済政策研究部門教授）

具体的な審議経過は以下のとおりである。

令和 2 年 10 月 30 日 第 8 回評価分科会

- (1) 精度検査報告書提言に対応した取組について（社会教育調査（民間体育施設））
- (2) 建設工事統計調査（施工調査）の欠測値対応の見直しについて

令和 3 年 3 月 26 日 第 9 回評価分科会

- (1) 欠測値の補完に係る主な方法等について

令和3年4月15日 第10回評価分科会

- (1) 経済産業省生産動態統計調査の欠測値補完方法の検証について

令和3年6月23日 第11回評価分科会

- (1) 令和2年度統計委員会評価分科会審議結果報告書（案）（第8回～第11回審議分）について

審議結果

1 社会教育調査（民間体育施設）の母集団の明確化及び欠測値補完の見直しについて

（１）取組の現状

社会教育調査のうち、民間体育施設を対象とした調査については、全数調査で実施されているが、「平成 28 年度統計法施行状況審議結果報告書（統計精度検査関連分）」（平成 30 年 3 月 統計委員会）（以下「報告書」という。）において、回収率が約 60%であるが、全部無回答に対し、単一補完やウェイト調整などの欠測値への対応が行われることなく単純合算集計がなされ、調査結果が過小になっていると見込まれると指摘されている。さらに、同報告書では、平成 33 年調査に間に合う時期までに、母集団を明確にした上で、明確になった母集団及び現状の調査実施の状況を踏まえて、同調査における対応を検討し、結論を得ることが必要である、とされている。

この報告書を踏まえ、文部科学省は、令和 2 年 8 月～10 月に有識者による「社会教育調査の改善に関する研究会」を開催し、検討を行い、令和 3 年度からの調査に関する以下の対応方針について研究会の了承を得た。

- ① これまでできるだけ広く民間体育施設の全体像を把握する観点から、施設の定義を示してきたが、様々な主体が含まれていて多様過ぎるために一義的に定義を示すことが困難であり、都道府県・市町村教育委員会において母集団名簿を整える際に差が生じていたことを踏まえ、独自に母集団全体の名簿を捉えることは極めて困難なことから、総務省が作成している事業所母集団データベースを母集団名簿として活用する。
- ② 現在の回収率・回収数から考えると、標本調査とすることは適当でなく、回収数を確保する観点から、引き続き全数調査とする。
- ③ 集計の際の補完方法については、集計表の表章形式に鑑み、都道府県別にウェイトの調整を行う。

（２）評価及び課題解決に向けた今後の取組の方向性

文部科学省が、事業所母集団データベースを母集団名簿として活用することとした対応について、母集団が明確化されたことに加え、名簿の更新に関する調査主体の負担の軽減につながるものであることから妥当であると評価する。また、引き続き全数調査としたことや、集計の際の補完方法として回収率を反映させたウェイトの調整について、全国一律での調整ではなく、都道府県別に調整を行うこととした対応についても妥当である。ただし、基幹統計として現状では低い回収率について、上げる努力をすることが必要と考えられる。

2 建設工事統計調査（施工調査）の欠測値補完の見直しについて

（1）取組の現状

- ① 建設工事統計調査（施工調査）については、「平成 28 年度統計法施行状況審議結果報告書（統計精度検査関連分）」（平成 30 年 3 月 統計委員会）（以下「報告書」という。）において、同調査の回収率が約 60%であるにもかかわらず、全部無回答を「生産なし」と見なして線形推定が行われていることから、調査結果の過小推計の可能性が指摘された。

報告書において、平成 29 年度に、経済センサス等との比較検証等を行った上で、当該比較検証に基づき、平成 30 年度に調査方法及び推計方法の見直しに関する結論を得ることが必要であるとされたことを踏まえ、国土交通省は、平成 29 年度に調査研究を実施した。

この調査研究は、平成 24 年経済センサス-活動調査と平成 23 年度実績分の建設工事統計調査（施工調査）で調査票情報のマッチングを行い、建設工事統計調査（施工調査）の全部無回答の調査票情報を

- i) 経済センサスとマッチングした調査票
- ii) 経済センサスとマッチングしなかった調査票

の 2 種類に分別し、ii) には 0 値を補完、i) をウェイト調整法により補完することで、建設工事統計調査（施工調査）の無回答部分を過剰推計することなく補完した。

この取り組みに対し、統計委員会評価分科会は、国土交通省が行った調査研究の成果を高く評価し、「令和元年度統計委員会評価分科会審議結果報告書（第 1 回～第 4 回審議分）」により、調査研究で実施した補完方法を適用した推計方法の導入に向け、令和 2 年度実績分の調査に間合おうよう平成 28 年経済センサス-活動調査の情報を用いて、調査研究で用いた方法による欠測値補完の効果の確認を行う必要があると指摘した。

- ② 以上の評価分科会による指摘を踏まえ、国土交通省は、令和元年度および 2 年度に調査研究を実施した。

この調査研究は、平成 28 年経済センサス-活動調査と平成 27 年度実績分の建設工事統計調査（施工調査）で調査票情報のマッチングを行い、加えて、経営事項審査結果^{*}との照合を行うことにより、建設工事統計調査（施工調査）の全部無回答の調査票情報を

- i) 経営事項審査結果との照合により完成工事高等のデータを補完できる調査票
- ii) 経営事項審査結果から補完できないが、経済センサスとマッチングした調査票

iii) 経営事項審査結果から補完できず、経済センサスともマッチングしない調査票

の3種類に分別した。i) は経営事項審査結果のデータを用いて補完、iii) には0値を補完、ii) を資料編資料2 P.27 (シート9) に記載されたウェイトを調整する方法により補完することで、無回答による偏りが、より小さくなるように推定している。[資料2 P.26~28 (シート7~10) 参照]

当該手法により建設工事統計調査(施工調査)の無回答部分を補完することにより、平成28年経済センサスー活動調査の結果とほぼ合致することを確認した。

※公共工事の入札参加を希望する業者が受ける審査であり、建設許可業者47万業者のうち約14万業者が受審。透明性の確保等の観点から審査結果はWEB上で公表されており、完成工事高等の情報を活用することが可能。

(2) 評価及び課題解決に向けた今後の取組の方向性

国土交通省が、評価分科会の指摘を踏まえて欠測値補完の効果に関する検証を行った上、今後の対応として、行政記録情報(経営事項審査結果)を活用して補完を行った後に、残った部分に関して経済センサスとの照合結果を踏まえたウェイトの調整により補完を行うとしていることについては妥当であると評価する。なお、当該手法については、行政記録情報を活用した補完方法として、他の統計にも参考となる事例であると考えられる。ただし、基幹統計として現状では低い回収率について、上げる努力をすることが必要と考えられる。

3 経済産業省生産動態統計調査の欠測値の補完について

(1) 取組の現状

- ① 経済産業省生産動態統計調査については、経済産業省所管の他の2つの統計調査と共に、「平成28年度統計法施行状況審議結果報告書(統計精度検査関連分)」(平成30年3月 統計委員会) (以下「報告書」という。)において、欠測値が発生した場合、過去の回答結果を使用した単一補完 (Last Observation Carried Forward。以下「LOCF」という。) が明確な使用期限を定めることなく行われていることに関し、回収率90%以上を保持していることから、致命的な対応とは判断されないが、平成30年度から総務省の支援を得て、計画的なシミュレーションによってデータの使用期限も含めた補完方法の検証を行うことが必要であると指摘された。

そのため、平成 30 年度において、総務省統計研究研修所の協力を得て、経済産業省生産動態統計調査の「自動車等」の調査票情報を基にクロスバリデーションに類似したシミュレーションが行われた。

その結果を受けて、第 3 回評価分科会（令和元年 6 月 27 日）において、LOCF 法は、平均値を補完する方法、年の伸び率で補正して補完する方法、前年同月比で補正して補完する方法など、LOCF 法より手間のかかる方法と比較しても補完効果はほぼ変わらないことが報告されるとともに、シミュレーション結果を踏まえて総務省から提案として示された同調査における LOCF 法の利用に関する以下の i) から iii) の考え方について審議が行われた。

- i) 長期的に LOCF 法を行えば、精度悪化のリスクは高まるものの、数年程度であれば極端に悪化するものではないことから、自動車等に含まれる品目のうち、金額の変動が大きくない品目であって補完期間が 5 年に至るものについて、5 年毎に実施される経済センサス-活動調査を活用し、経済センサス-活動調査の当該事業所の伸び率等で LOCF 値を補正するなどの対応を行うこと
- ii) 自動車等に含まれる品目のうち、複数年に渡り金額の大きな変動が続いている品目については、2 年以内に同じ品目内の伸び率等に乗じて LOCF 値を補正し、併せて、5 年毎の経済センサス-活動調査の当該事業所の結果を用いて補正するなどの対応を行うこと
- iii) 検証を行っていない自動車等に含まれる品目以外の品目についても、上記 i) 及び ii) のルールを暫定的に導入すること

これについて、評価分科会では、2012 年以降のシミュレーション結果に関しては、2003 年から 2012 年の間の結果で判断すると自動車等の品目に関しては、LOCF 法は一定程度、有効に機能していること、また、自動車等の品目については、上記 i) 及び ii) による対応は妥当であると評価した。ただし、上記 iii) の対応については、多数の品目で検証を行うまでの必要はないものの、自動車等の品目以外の 2～3 の品目においても同様の検証を行い、その結果を踏まえて判断することが適当であると指摘した。

- ② 上記の評価分科会による指摘を踏まえ、総務省統計研究研修所の協力を得て、令和 2 年度において、「電子部品」及び「段ボール」の品目を取り上げ、「自動車」の品目と同様にクロスバリデーションに類似したシミュレーションが行われた。

経済産業省では、今回のシミュレーション結果について、i) 時間が経過すれば、乖離は拡大するものの、一部の特殊な品目を除き、長期的

に乖離が極端に拡大する傾向は観測されなかった、ii) 数年程度であれば、一部の特殊な品目を除き、LOCF法は有効に機能すると評価するとした上で、今後、報告をいただけない事業所への依頼・要請を最大限実施することが最も重要との認識を示した上で、報告を頂けない場合は、LOCF法により欠測値補完を行うこととした。変動の大きな品目への対応としては毎年の年間補正の中で行い、安定した推移をしている品目への対応としては経済センサスー活動調査において、生産動態統計調査の無回答事業者と同一の従業者規模等の層に属すると思われる事業所集団の傾向を用いて、LOCF法による値を5年毎に見直すこととしている。

(2) 評価及び課題解決に向けた今後の取組の方向性

経済産業省が、生産動態統計調査におけるLOCF法の利用について、今回のシミュレーション結果を踏まえ、数年程度であれば一部の特殊な品目を除き、LOCF法は有効に機能していると評価した上で、今後の対応として、変動の大きな品目への逐次対応は年間補正の中で行い、安定した推移をしている品目については、経済センサスー活動調査における、無回答事業者と同一の層に属すると思われる事業所集団の傾向を用いてLOCF法による値を5年毎に見直すこととしたことは妥当であると評価する。

また、今回のシミュレーションを通じて、時系列的なデータに対する補完手法を比較検討する中で、LOCF法を利用することの効果を確認されたことを評価する。

4 欠測値の補完に係る主な方法等について

公的統計を作成するための統計調査において、調査の実施段階においては調査客体への確認や督促なども含め調査票への記入及び回収の確保に努めているところであるが、一部の調査項目が無回答である場合や、回答そのものが得られずにすべての調査項目が欠測となった場合には、結果の有用性を確保するため、統計的な補完を適切に行うなどの対応が考えられる。一方で、欠測値については統計調査ごとに欠測の発生状況や補完に利用できるデータなどに違いがあり、特定の補完方法の適用など一律の対応は困難な面がある。

このため、統計調査ごとの状況を踏まえた適切な対応が重要となるが、補完を行うに当たっての主な方法・手順や利用上の注意点など、実務上参考となる事項を整理してまとめて示すことは、公的統計の正確性や有用性の確保を促進する上で有意義と考えられる。

以上の観点から、これまでの評価分科会における審議や各府省における取組状況等も踏まえて事務局において作成した欠測値の補完に係る主な方法等についての資料を基に評価分科会において審議を行い、【資料編；資料5】の

とお取りまとめた。

統計調査においては、欠測値をできる限り発生させないようにすることが最も重要であるが、最終的に発生した欠測値への対応を行う上で、今回の取りまとめが各府省の実務担当者に広く共有され、欠測値への補完等の方法について理解を深めるための参考となるとともに、補完等の対応が適切に行われ、かつその適切性が国民に対して十分な説明がなされていることで、公的統計の正確性や有用性の一層の確保に寄与することを期待する。

5 令和3年度における評価分科会の取組について

(1) 精度検査報告書のフォローアップ

精度検査報告書Ⅱ-3-(2)-③から⑧までで対応が求められている事項のうち、まだフォローアップが行われていない経済産業省企業活動基本調査について、その進捗状況を聴取し、審議を行う。

(2) 統計調査における母集団名簿の整備状況について

正確な統計作成を行う上での基本的な事項として、基幹統計調査における母集団名簿の整備状況について確認することとし、事務局において現状の把握・整理の上、それらの結果等を踏まえて、評価分科会において必要な審議を行う。

【資料編】

文部科学省が実施する社会教育調査のうち、 民間体育施設に係る母集団について

令和2年10月30日

文部科学省総合教育政策局調査企画課

1

1. 社会教育調査の概要

調査所管課

文部科学省総合教育政策局調査企画課

調査の目的

社会教育行政に必要な社会教育に関する基本的事項を明らかにすることを目的とする。

調査の概要

調査の沿革	<ul style="list-style-type: none"> 昭和30年以降3～5年間隔で実施しており、令和3年度調査は21回目 	調査票及び調査事項	<ul style="list-style-type: none"> ① 社会教育行政調査票 社会教育関係職員に関する事項、社会教育委員等に関する事項、社会教育関連事業の実施状況等 ② 公民館調査票 ③ 図書館調査票 ④ 博物館調査票 ⑤ 青少年教育施設調査票 ⑥ 女性教育施設調査票 ⑦ 体育施設調査票 ⑧ 劇場、音楽堂等調査票 ⑨ 生涯学習センター調査票 ②～⑨：職員数、施設・設備の状況、事業実施状況、施設の利用状況等
調査期日	<ul style="list-style-type: none"> 調査年10月1日（活動状況については前年度間） 3年周期 		
調査範囲及び報告者数	<ul style="list-style-type: none"> 都道府県教育委員会、市町村教育委員会、公民館、図書館、博物館、青少年教育施設、女性教育施設、社会体育施設、民間体育施設、劇場・音楽堂等、生涯学習センター（約64,000：全数） 		
調査組織及び調査方法	<ul style="list-style-type: none"> 調査組織：文部科学省－都道府県教育委員会－市町村教育委員会－報告者 調査方法：郵送又はオンライン 		
結果公表	<ul style="list-style-type: none"> 中間報告：調査翌年の7月に公表 報告書：調査翌々年の3月に公表 		

2

2 . これまでの経緯・課題等

「平成28年度統計法施行状況に関する審議結果報告書（統計精度検査関連分）」
（平成30年3月30日 総務省統計委員会）

社会教育調査のうち、民間体育施設（約1万5千施設）を対象とした調査は、全数で実施されている。同調査は、回収率が約60%であるが、全部非回答に対し、単一補完やウエイト調整などの欠測値への対応が行われることなく単純合算集計がなされ、調査結果が過小となっていると見込まれる。

このため、次々回調査（平成33年調査）に間に合う時期までに、母集団を明確にした上で、明確になった母集団及び現状の調査実施の状況を踏まえて、次々回調査における対応を検討し、結論を得ることが必要である。



上記の指摘を踏まえ、民間体育施設の母集団の明確化及び調査・補完方法について見直しを行った。

3

3 . 検討方法

社会教育調査の改善に関する研究会を8月～10月に開催し、検討を行った。
委員名簿は下記のとおり。

社会教育調査の改善に関する研究会 委員名簿

（敬称略，五十音順）

青山 鉄兵（文教大学人間科学部 准教授、国立青少年教育振興機構青少年教育研究センター 副センター長）
金山 喜昭（法政大学キャリアデザイン学部 教授）
岸 正人（公益社団法人全国公立文化施設協会 事務局次長）
澁谷 茂樹（公益財団法人笹川スポーツ財団スポーツ政策研究所 シニア政策アナリスト）
関 福生（愛媛県新居浜市教育委員会生涯学習センター 所長、新居浜生涯学習大学 学長）
丹間 康仁（千葉大学教育学部 准教授）
土屋 隆裕（横浜市立大学データサイエンス学部 教授）
富松 文夫（福岡県教育委員会 社会教育課長）
納米 恵美子（特定非営利活動法人全国女性会館協議会 代表理事）
野口 武悟（専修大学文学部 教授）
福田 範史（鳥取県南部町教育委員会 教育長）

4

4 . 母集団名簿における課題

(1) 現状

- ・調査名簿については、「社会教育施設等名称ファイル」で管理しており、各都道府県教育委員会及び市町村教育委員会が更新を行っている。
- ・上記ファイルの更新にあたっては、これまでは「経済センサス（事業所母集団データベース）」等を参考として、新設・廃業の確認などに利用している。

(2) 課題

- ・新設・廃業の情報として参照している経済センサス（事業所母集団データベース）において調査対象と対象外の事業所が混在している分類があるが、名簿に記載された事業所の名称で調査対象業種か否かを判断することが困難な可能性がある。
- ・母集団名簿の更新は、経済センサス調査の対象事業所以外についても幅広く対象となり得るものであるが、その様な施設（個人がボランティア的に運営している施設等）は、都道府県教育委員会及び市町村教育委員会担当者が統一的に把握しきれていない可能性がある。
- ・また、現状の母集団名簿は、調査票発送のための住所情報しか保有していない。そのため、欠測値補完については、地域（都道府県）をベースにしたもの以外はできない（施設の種別等の欠測値補完をすることはできない。）

(参考) 民間体育施設調査の対象となる施設について

調査の範囲は、一般の利用に供する目的で独立行政法人又は民間が設置した体育館、水泳プール、運動場等のスポーツ施設を対象とします。

※ 民間の設置する体育施設のうち、企業の職員の福利・厚生用の施設は除きます。

5

(参考) 「スポーツ施設に関する調査研究」報告書における記載

平成27年度 スポーツ庁スポーツ政策調査研究 「スポーツ施設に関する調査研究」報告書では、『調査対象の一部市町村に電話でのヒアリングを実施したところ、「（スポーツ担当部局の）所管外の施設は把握できていないため、回答困難」との回答が多数あった。例えば、教育委員会が所管している場合、障害者スポーツセンター、都市公園、農業公園、民間の施設といった他の部局が所管する施設について把握できていないといったケースである。特に民間の施設については、多くの市町村で把握困難との回答があった。』といった記載がある。

(参考) 研究会での有識者からの意見

- ・自治体での母集団名簿の作成については、前回調査の施設名簿や経済センサスをもとに市町村でチェックをして更新している。また、政令市・中核市のような大きい規模になると、なかなか全数を把握するのは難しい。
- ・母集団名簿作成は、前回名簿をベースに、経済センサスで新たに掲載された企業や施設について、各市町村の担当者が電話や直接出向いて確認する。対象が多くなると各市町村の事務に携わることのできる職員も限られているので対応が難しい面もある。

6

5. 母集団名簿の改善案

(1) 母集団名簿の変更

上記の課題を踏まえ、母集団を明確にするための改善案として、経済センサスを母集団名簿とする。なお、経済センサスを母集団とした場合は、対応表のうち、赤枠の体育施設を対象とし、従前、一部の事業所のみが対象となっていた分類については、統一的な名簿を整備する観点から除外する。

事業所母集団名簿と社会教育調査の調査対象施設の対応表

事業所母集団名簿の産業分類			対応する社会教育調査の調査対象
大分類	中分類	小分類	
M 宿泊業、飲食サービス業	75 宿泊業	752 簡易宿所	体育施設（山小屋等）
		75B 他に分類されない宿泊業	体育施設（キャンプ場等）
N 生活関連サービス業、娯楽業	80 娯楽業	802 興行場（別掲を除く）、興行団	劇場、音楽堂等
		80A スポーツ施設提供業（別掲を除く）	体育施設
		80B 体育館	
		80C ゴルフ場	
		80D ゴルフ練習場	
		80E ボウリング場	
		80F テニス場	
		80G バッティング・テニス練習場	
		80H フィットネスクラブ	
		O 教育、学習支援業	82 その他の教育、学習支援業
80P 他に分類されない娯楽業	体育施設		
82A 公民館	公民館		
82B 図書館	図書館		
82C 博物館、美術館	博物館		
82D 動物園、植物園、水族館	博物館		
82E その他の社会教育	青少年教育施設、女性教育施設、生涯学習センター		
82L スポーツ・健康教授業	体育施設		
R サービス業（他に分類されないもの）	95 その他のサービス業	82M その他の教養・技能教授業	体育施設（ダンススタジオなど）
		951 集会場	女性教育施設、劇場、音楽堂等（県民会館、婦人会館等）

[出所]平成30年度社会教育調査の手引[教育委員会用1]

一部の事業所のみが対象となる小分類は除外。

7

(2) 母集団名簿の変更に伴う影響

- 母集団名簿の変更により、以下の課題の解決が可能となる。
 - ・現状では母集団名簿の更新が教育委員会担当者に任されていたため、教育委員会担当者が把握できていない部分・判断できない部分については不明確な個所もあったが、母集団を経済センサスに基づいて、統一的に整備することで、母集団を明確化することができる。
 - ・母集団について、住所情報以外の経済センサスが有する情報を十分活用することが可能となる。
 - ・母集団の明確化により、教育委員会担当者の名簿更新作業の負担が大幅に軽減され、回収率の向上に取り組むことが可能となる。
 - ・なお、母集団が過去調査と変わるため、調査結果の時系列比較ができなくなることについて、報告書等への注釈を行うこととする。

8

6. 欠測値を補う改善案

平成30年度調査のデータを用いて、都道府県別に欠測値を補完するための調整値を算出したところ、全国一律の調整値が1.85、最大値が3.82、最小値が1.04であり、3倍以上の差があるため、都道府県別の調整を行うことが考えられる。

都道府県番号	都道府県	H30母集団 (%)	H30回収数 (%)	調整値
1	北海道	632 (3.9%)	420 (4.7%)	1.50
2	青森県	226 (1.4%)	166 (1.9%)	1.36
3	岩手県	120 (0.7%)	74 (0.8%)	1.62
4	宮城県	291 (1.8%)	145 (1.6%)	2.01
5	秋田県	137 (0.8%)	92 (1.0%)	1.49
6	山形県	164 (1.0%)	87 (1.0%)	1.89
7	福島県	361 (2.2%)	184 (2.1%)	1.96
8	茨城県	675 (4.1%)	406 (4.6%)	1.66
9	栃木県	596 (3.6%)	330 (3.7%)	1.81
10	群馬県	346 (2.1%)	159 (1.8%)	2.18
11	埼玉県	1026 (6.3%)	476 (5.4%)	2.16
12	千葉県	962 (5.9%)	464 (5.2%)	2.07
13	東京都	1139 (7.0%)	517 (5.8%)	2.20
14	神奈川県	1333 (8.1%)	770 (8.7%)	1.73
15	新潟県	319 (1.9%)	148 (1.7%)	2.16
16	富山県	130 (0.8%)	83 (0.9%)	1.57
17	石川県	113 (0.7%)	64 (0.7%)	1.77
18	福井県	66 (0.4%)	38 (0.4%)	1.74
19	山梨県	246 (1.5%)	198 (2.2%)	1.24
20	長野県	531 (3.2%)	349 (3.9%)	1.52
21	岐阜県	288 (1.8%)	114 (1.3%)	2.53
22	静岡県	715 (4.4%)	402 (4.5%)	1.78
23	愛知県	699 (4.3%)	380 (4.3%)	1.84
24	三重県	308 (1.9%)	145 (1.6%)	2.12
25	滋賀県	183 (1.1%)	89 (1.0%)	2.06
26	京都府	298 (1.8%)	146 (1.6%)	2.04
27	大阪府	424 (2.6%)	192 (2.2%)	2.21
28	兵庫県	697 (4.3%)	293 (3.3%)	2.38
29	奈良県	141 (0.9%)	86 (1.0%)	1.64
30	和歌山県	107 (0.7%)	28 (0.3%)	3.82
31	鳥取県	67 (0.4%)	56 (0.6%)	1.20
32	島根県	49 (0.3%)	36 (0.4%)	1.36
33	岡山県	327 (2.0%)	115 (1.3%)	2.84
34	広島県	276 (1.7%)	157 (1.8%)	1.76
35	山口県	243 (1.5%)	161 (1.8%)	1.51
36	徳島県	74 (0.5%)	71 (0.8%)	1.04
37	香川県	164 (1.0%)	114 (1.3%)	1.44
38	愛媛県	196 (1.2%)	90 (1.0%)	2.18
39	高知県	72 (0.4%)	58 (0.7%)	1.24
40	福岡県	599 (3.7%)	322 (3.6%)	1.86
41	佐賀県	94 (0.6%)	51 (0.6%)	1.84
42	長崎県	94 (0.6%)	59 (0.7%)	1.59
43	熊本県	161 (1.0%)	123 (1.4%)	1.31
44	大分県	195 (1.2%)	103 (1.2%)	1.89
45	宮崎県	195 (1.2%)	112 (1.3%)	1.74
46	鹿児島県	105 (0.6%)	59 (0.7%)	1.78
47	沖縄県	189 (1.2%)	139 (1.6%)	1.36
	合計	16,373 (100.0%)	8,871 (100.0%)	1.85

・母集団及び回収数の単位は事業所。

・調整値は、「H30回収数」を「H30母集団」で除した数の逆数。

9

欠測値補完の試算結果は下記のとおり。施設の種類別でみた場合は、都道府県別及び全国一律の調整値をそれぞれかけたところ、数値に大きな差は見られない。

施設の種類別箇所数

施設番号	施設名	施設の箇所数	都道府県別	全国一律
	全体	16,397	30,734	30,264
1	陸上競技場	12	24	22
2	野球場・ソフトボール場	148	298	273
3	球技場	455	877	840
4	多目的運動広場	176	324	325
5	水泳プール（屋内）	1,360	2,588	2,510
6	水泳プール（屋外）	68	124	126
7	レジャープール	79	139	146
8	ダイビングプール	22	42	41
9	体育館	289	549	533
10	柔道場	237	448	437
11	剣道場	160	296	295
12	柔剣道場	26	49	48
13	空手・合気道場	463	853	855
14	バレーボール場（屋外）	5	10	9
15	庭球場（屋外）	607	1,151	1,120
16	庭球場（屋内）	272	516	502
17	バスケットボール場（屋外）	5	10	9
18	すもも場（屋外）	10	18	18
19	すもも場（屋内）	4	9	7
20	卓球場	105	202	194
21	弓道場	48	85	89
22	アーチェリー場	16	30	30
23	馬場	177	316	327
24	アイススケート場（屋内）	24	44	44
25	アイススケート場（屋外）	3	4	6
26	ローラースケート・インラインスケート場（屋外）	4	8	7
27	ローラースケート・インラインスケート場（屋内）	7	13	13
28	山の家	73	113	135
30	トレーニング場	1,564	2,971	2,887
31	レスリング場	3	6	6
32	ボクシング場	107	202	197
33	ダンス場	1,030	1,913	1,901
34	射撃場	154	375	284
35	ゴルフ場	4,413	8,347	8,145
36	ゴルフ練習場	1,221	2,259	2,254
37	ボウリング場	296	543	546
38	高艇場	0	0	0
39	ゲートボール・ゴッケー場	58	96	107
40	スカッシュ・ラケットボール場	43	84	79
41	ボート場	85	189	157
42	スキー・スノーボード場	332	572	613
43	キャンプ場	412	741	760
44	ハイキングコース	5	8	9
45	サイクリングコース	7	13	13
46	自転車レンタルコース	9	16	17
47	ランニングコース	4	7	7
48	冒險道具コース	22	40	41
49	海の家・海水浴場等の施設	52	91	96
50	河川・湖沼等の遊泳場	7	13	13
51	スカイボーツ施設	39	69	72
52	体験競技場	56	107	103
53	その他	1,623	2,930	2,996

10

都道府県別の施設の箇所数については、都道府県別及び全国一律の調整値をそれぞれかけたところ、数値に差が見られる。これまでも都道府県別集計を公表していることから、都道府県別の調整は必要といえる。

都道府県番号	都道府県	施設の箇所数	都道府県別	全国一律
	全体	16,397	30,734	30,264
1	北海道	726	1,092	1,340
2	青森県	272	370	502
3	岩手県	132	214	244
4	宮城県	271	544	500
5	秋田県	137	204	253
6	山形県	145	273	268
7	福島県	304	596	561
8	茨城県	912	1,516	1,683
9	栃木県	638	1,152	1,178
10	群馬県	289	629	533
11	埼玉県	1,093	2,356	2,017
12	千葉県	989	2,050	1,825
13	東京都	963	2,122	1,777
14	神奈川県	1,098	1,901	2,027
15	新潟県	229	494	423
16	富山県	125	196	231
17	石川県	162	286	299
18	福井県	88	153	162
19	山梨県	305	379	563
20	長野県	550	837	1,015
21	岐阜県	180	455	332
22	静岡県	715	1,272	1,320
23	愛知県	756	1,391	1,395
24	三重県	459	975	847
25	滋賀県	128	263	236
26	京都府	257	525	474
27	大阪府	360	795	664
28	兵庫県	660	1,570	1,218
29	奈良県	206	338	380
30	和歌山県	95	363	175
31	鳥取県	114	136	210
32	島根県	50	68	92
33	岡山県	212	603	391
34	広島県	401	705	740
35	山口県	300	453	554
36	徳島県	93	97	172
37	香川県	150	216	277
38	愛媛県	147	320	271
39	高知県	65	81	120
40	福岡県	483	899	891
41	佐賀県	106	195	196
42	長崎県	91	145	168
43	熊本県	170	223	314
44	大分県	168	318	310
45	宮崎県	189	329	349
46	鹿児島県	175	311	323
47	沖縄県	239	325	441

7. 対応方針

平成28年度統計法施行状況に関する審議結果報告書において指摘のあった社会教育調査の民間体育施設の母集団及び現状の調査実施状況を踏まえた次回調査の対応については、以下のとおり考える。

本件については、文部科学省としても、できるだけ広く民間体育施設の全体像を把握する観点から、施設の定義を示してきたが、様々な主体が含まれていて多様過ぎるために一義的に定義を示すことが困難であり、都道府県・市町村教育委員会において母集団名簿を整える際に差が生じていた。

これらを踏まえると、独自に母集団全体の名簿を捉えることは極めて困難なことから、経済センサスを活用することが適当であると考える。

また、経済センサスを活用することにより、回収率の一層の向上を図るとともに、これまでの回収数の状況を踏まえると、標本調査とすることは適当でなく、引き続き全数調査とすることが適当である。

なお、集計の際の欠測値補完については、集計表の表章形式に鑑み、都道府県別の調整を行うことが適当と考える。さらに、経済センサスを活用することにより、多くの情報が得られることから、実際の回収状況に鑑み、今後、適切と考えられる方法を丁寧に検討していくことが適当と考える。

(参考) 体育施設調査票

(様式第1号)

◎ 統計法に基づく基幹統計調査

平成30年度 社会教育調査
体育施設調査票
 平成30年10月1日現在

(注) 1. [] のような枠内には、数字を右詰めで記入します。
 2. 該当しない場合は「0」と記入する。
 3. 該当しない場合は「-」を記入し、「0」は記入しません。

※ 該当する番号または記号を全て囲み、必要の場合は数値等も記入してください。

1 施設の名前	2 施設の所在地	3 施設の長の氏名	4 取扱者氏名
甲 (市)	(市)		

郵便番号	教育委員会番号	施設部番号
7		
施設所在地	郵便番号	市区町村番号

※教育委員会へ記入します。

5 設置者 (公立) 1 市町村 2 市 3 町 4 村 5 組合 (公立以外)※ 6 独立行政法人 7 一般社団法人・一般財団法人・公益社団法人・公益財団法人 8 会社 9 その他の法人 10 任意団体 11 個人	6 所管別 (公立の施設のみ回答) 1 教育委員会 2 地方公共団体の長	7 指定管理の相手先 (公立の施設のみ回答) 1 管理者の指定無 2 地方公共団体等 3 地域による団体(自治会・町内会等) 4 任意の会等(任意)を指定 5 任意の会等(任意)を指定 6 NPO法人を指定 7 その他を指定	8 職員数(人) 区分 施設の長 指導員 その他の職員 専任 兼任 非常勤 指定管理者
---	---	--	---

※ 設置者が(公立以外の)施設は、表面の9(2)～10、11(2)～12は回答不要です。

9 施設・設備の状況

(1) 施設の種類

種類	種類番号	施設の管理状況			指導員数の状況		施設の開放状況(平成29年度間)		施設・設備の有無					
		規模1	規模2	規模3	自然施設	複数施設	夜間開放	年間利用者数	外国人受け入れ	スロースポーツ	エレベーター	障害者利用設備	点字による案内	障害者専用車庫
陸上競技場	01													
野球場・ソフトボール場	02													
球技場	03													
多目的運動広場	04													
水泳プール(屋内)	05													
水泳プール(屋外)	06													
レジャープール	07													
体育館	09													
柔道場	10													
剣道場	11													
柔道場	12													
剣道場(屋外)	13													
庭球場(屋内)	14													
テニスコート	21													
H-テニスコート	30													
ゴルフ場	35													
キャンプ場	40													
以下、上記以外の施設を記入														

この調査は、統計法に基づく基幹統計を構成するものに付属します。この調査の対象は、本団体・施設の方々に統計法に基づく報告の義務があり、報告の届出や調査報告については関係ありません。この調査の実施にあたっては、特に必要から報告には、資料の提出の必要は関係者の方々の協力を要行することがあります。

裏面に続く

文 部 科 学 省

(参考) 体育施設調査票

※ 設置者が公立以外の施設は、裏面の9(2)～10、11(2)～12は回答不要です。裏面は11(1)のみ回答してください。

9 施設・設備の状況(続き)

(2) 変動喫煙防止のための対策の方法

- 敷地内を禁煙している
- 施設内を禁煙している
- 施設内に喫煙場所を設置するとともに、喫煙場所から非喫煙場所へはこの煙が流れ出ないように措置している
- 施設内に喫煙場所を設置しているが、喫煙場所から非喫煙場所へはこの煙が流れ出ないように措置していない
- 何ら措置を講じていない

(3) コンピュータの導入状況

インターネットに接続したコンピュータの有無 1 有 2 無

10 ボランティア活動状況

(1) ボランティア登録数

区分	登録団体数(団体)	登録者数(人)	
		男	女
個人			

(2) ボランティアに対する研修の有無

1 有 → [] 回 2 無

「1」を選択した場合、平成29年度間の実施回数をお答えください。

(3) ボランティア活動の種類(複数回答可)

- 主催事業(講座等)の運営支援
- 施設利用者の活動補助
- 自主企画事業(講座等)の実施
- 環境保全(館内美化等)
- 託児
- その他

11 事業実施状況(平成29年度間)

(1) 各種事業

主 催	区分	実施件数(件)	参加者数(人)
主 催	スポーツ教室		
	指導者研修会、講習会等		
	スポーツ大会		
	スポーツテスト会		
共 催	スポーツ相談		
	スポーツ教室		
	指導者研修会、講習会等		
	スポーツ大会		
共 催	スポーツテスト会		
	スポーツ相談		

(2) 民間社会教育事業者との連携・協力の状況(1)の再掲

富 利	[] 件
非 富 利	[] 件

(3) 共催相手(複数回答可)(1)の再掲

- 他の体育施設
- 以外の社会教育施設
- 学校(大学)
- 学校(大学以外)
- 教育委員会
- 知事部局・市町村部局
- その他

(4) 情報提供方法(複数回答可)

- 情報ネットワーク
- 公共広報紙
- 機関紙、ポスター、パンフレット等
- マスメディア(放送・新聞等)
- 説明会・訪問
- その他

「1」を選択した場合の情報提供方法(複数回答可)
 a ホームページ b メールマガジン c ソーシャルメディア

12 運営状況に関する評価の実施状況(平成29年度)

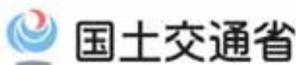
(1) 平成29年度の運営状況について評価を実施している。

- 自己評価 1 有 2 無
- 外部評価 1 有 2 無

(2) 評価結果を公表している。 1 有 2 無

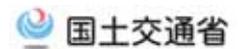
建設工事施工統計調査における欠測値補完の見直しについて(案)

令和2年10月 国土交通省総合政策局情報政策課 建設経済統計調査室



Ministry of Land, Infrastructure, Transport and Tourism

これまでの検討の経緯



○「公的統計の整備に関する基本的な計画」(平成30年3月6日、令和2年6月2日(変更)閣議決定)

建設工事施工統計について、精度向上に向けた見直しを検討する。(実施時期:平成30年度(2018年度)から実施する。)

○「平成28年度統計法施行状況審議結果報告書(統計精度検査関連分)」概要(平成30年3月30日総務省統計委員会)

- ・建設工事統計調査は、回収率が約60%だが、全部無回答に対して欠測値対応を行っておらず、調査結果が過少と見込まれる。
- ・一方、無回答業者には実績のない業者が多分に含まれている可能性があり、回収率の逆数を乗じると課題推計の恐れがある。
- ・このため、まずは**無回答事業者の実態について検証を行う必要がある**。また、**回収率向上に向けた方策の検討も必要**である。
- ・そのため、**経済センサス等との比較検証**を行い、**平成30年度に推計方法・調査方法の見直しに関する結論**を得る必要がある。

○「建設工事施工統計調査における欠測値補完の見直しについて(案)」概要(平成31年3月12日総務省統計委員会評価分科会)

<推計方法・調査方法の見直しに関する結論>

- ・建設工事統計調査の全部無回答の調査票情報を

i) **経済センサスとマッチングした調査票**

ii) 経済センサスとマッチングしなかった調査票

の2種類に分別し、ii)には0値を補完、i)を「**ウェイト調整法※**」により**欠測値補完**する。

<今後の取組内容と期限>

- ・平成28年経済センサス-活動調査の情報を用いて、ウェイトを5年間固定した場合の「ウェイト調整法」による欠測値補完の妥当性を検証し、運用実務上支障がないか確認する。
- ・回収率向上に向けた調査事項等の見直し、他統計(建設工事受注動態統計調査、建築物リフォーム・リニューアル調査、建設総合統計、産業連関表等)への影響確認等を行う。
- ・次回産業連関表に使用される**令和2年度実績を把握する調査において、推計方法・調査方法の見直しを反映**させる。

○「令和元年度 統計委員会評価分科会審議結果報告書(第1回～第4回審議分)」概要(令和元年9月20日総務省統計委員会)

- ・国土交通省が行った調査研究の成果を高く評価する。国土交通省は、調査研究で実施した補完方法を適用した推計方法の導入に向けて、**平成28年度経済センサス-活動調査の情報を用いて欠測値補完の効果の確認を行う必要**がある。
- ・その際、建設工事統計調査側における回答企業及び(経済センサス側では回答が得られている)無回答企業の標本を全て用い、経済センサス側の回答情報から回帰分析を行うことにより、**回答企業と無回答企業とでどの程度の差異が生じているのかについて、業種や資本金階層などの属性の差異も含めて検証**を行うことができれば、更に緻密な欠測値補完の実施につながり、他の統計にも応用できるものとする。

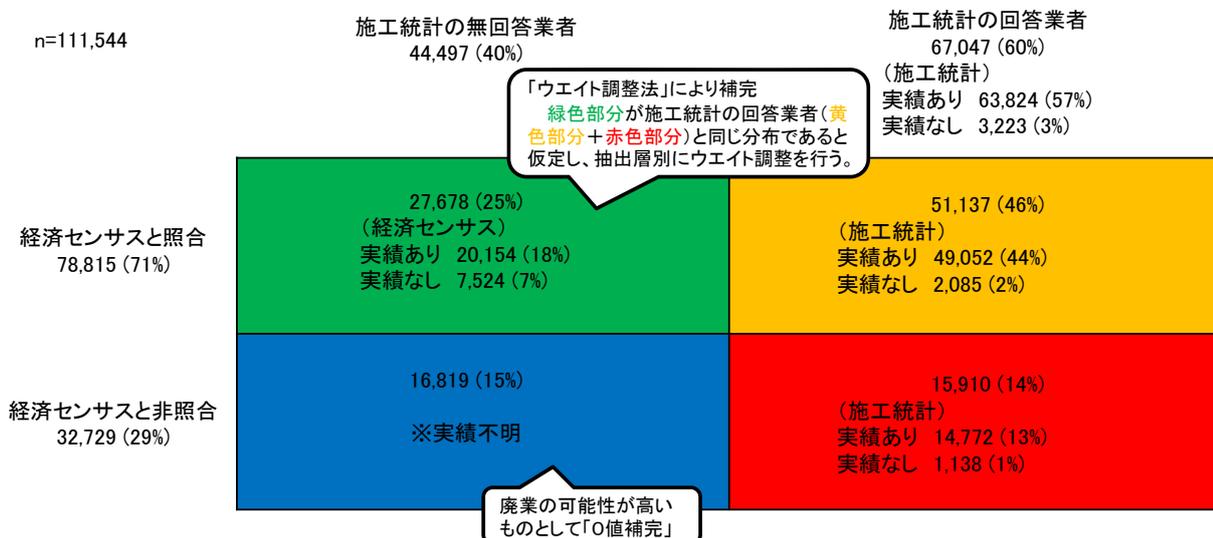
「ウェイト調整法」経済センサスと照合できた業者数を施工統計調査の層別の乗率に加えることで、欠測値を補完する方法

○建設工事施工統計調査は、回収率が約60%だが、無回答業者に対して欠測値補完を行っておらず、調査結果が過少と見込まれる。一方、無回答業者には実績なしの業者が含まれている可能性があり、回収率の逆数を乗じると過大推計の恐れがある。
⇒このため、経済センサスー経済調査の情報等を活用し、建設工事施工統計調査の無回答業者の活動実績を踏まえた欠測値補完の方法を検討。



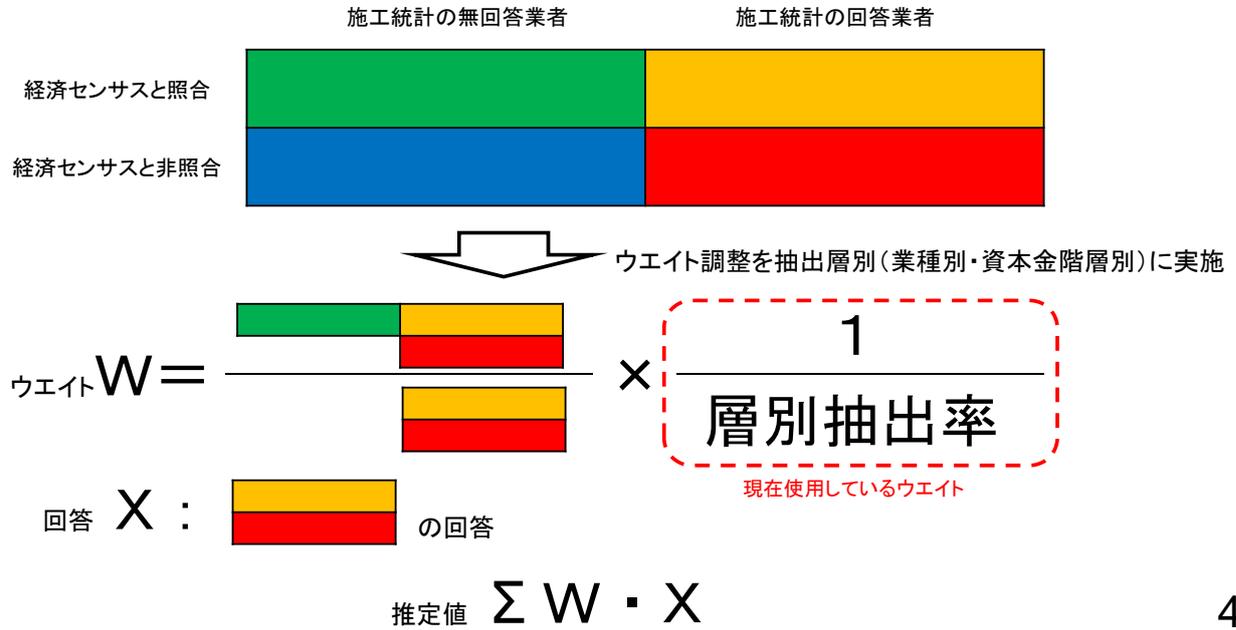
活動実績を踏まえた欠測値補完方法を検討

○平成28年経済センサスー活動調査と照合したところ、建設工事施工統計調査の無回答業者約4.4万業者のうち、経済センサスと約2.8万業者で照合でき、うち約2.0万業者が実績あり。
○前回は、建設工事施工統計調査の無回答業者のうち、経済センサスと非照合の業者(青色部分)を「0値」補完、経済センサスと照合できた者(緑色部分)を「ウエイト調整法」により欠測値補完する案を提示。



<「ウエイト調整法」による補完>

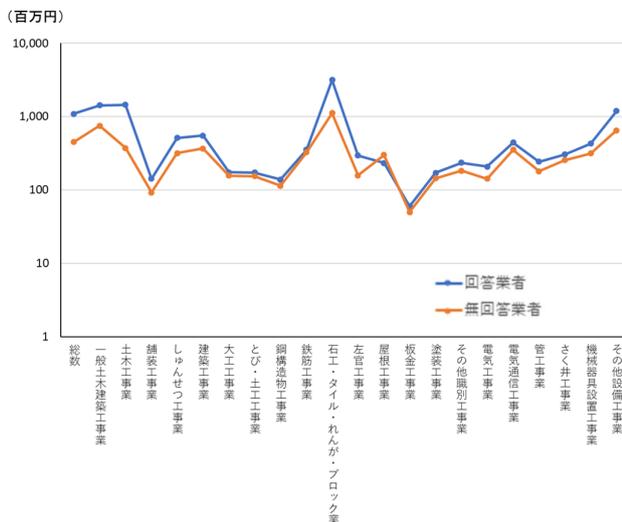
施工統計の無回答業者のうち経済センサスと照合できた業者(緑色部分)が施工統計の回答業者(橙色部分+赤色部分)と同じ分布であると仮定し、抽出層別にウエイト調整を行う。



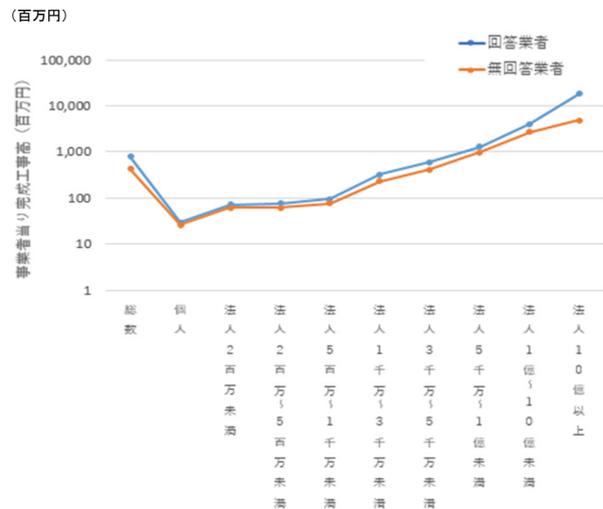
4

○ここで、建設工事統計調査における回答業者と(経済センサス側で回答が得られている)無回答業者とで業種や資本金階層の属性別に完成工事高を確認したところ以下のとおりであった。

業種別の1業者当たり完成工事高



資本金階層別の1業者当たり完成工事高



5

○建設工事施工統計調査における回答業者と(経済センサス側で回答が得られている)無回答業者の標本に対し、業種や資本金階層などを変数として回帰分析を試みたが、決定係数が低く、層別ではマイナスになるなど、活用には課題が残る結果となった。

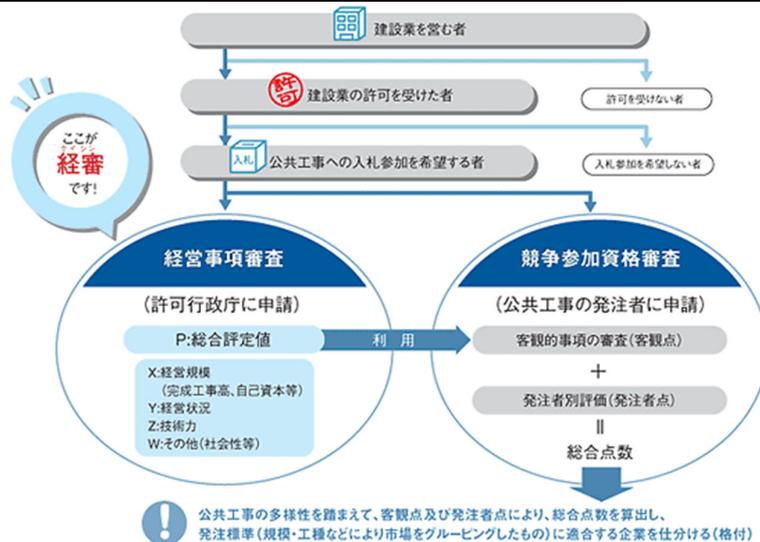
$$Z = c + b_1X_1(\text{業種ダミー}) + b_2X_2(\text{資本金階層別ダミー}) + b_3 \cdot \text{Dummy1}(\text{主業ダミー}) + b_4 \cdot \text{Dummy2}(\text{施工統計無回答ダミー})$$

業種ダミー(設備工事業が基準)		
総合工事業	429.4	(133.0) ***
職別工事業	956.0	(153.0) ***
資本金階層別ダミー		
(5000万円以上1億円未満が基準)		
個人企業	-1665.1	(346.2) ***
法人企業(1000万円未満)	-1451.6	(200.1) ***
法人企業(1000万円以上5000万円未満)	-1112.6	(163.1) ***
法人企業(1億円以上10億円未満)	2870.7	(304.2) ***
法人企業(10億円以上50億円未満)	10626.8	(642.6) ***
法人企業(50億円以上)	54578.0	(717.4) ***
主業ダミー	1770.0	(110.6) ***
施工統計無回答ダミー	-311.9	(104.7) ***
定数項	-129.7	(189.7)
adj-R-sq	0.083	
N	76274	

		法人企業					合計		
		1000万円未満	5000万円未満	1億円未満	10億円未満	50億円未満		50億円以上	
総合工事業	推計値	-234,267	-932,038	-281,323	1,280,601	1,570,449	703,020	2,757,139	4,863,581
	実績値	13,687	190,333	3,137,488	1,485,249	1,356,719	450,932	1,755,161	6,809,924
職別工事業	推計値	-30,556	-15,509	1,205,724	784,522	917,753	392,106	1,461,581	4,715,621
	実績値	10,633	112,652	1,370,169	720,071	907,685	151,087	160,187	3,432,484
設備工事業	推計値	-179,277	-1,160,579	-2,673,950	-97,502	778,875	620,519	2,374,626	-337,488
	実績値	2,974	60,056	599,994	317,949	210,830	23,495	108,484	1,323,782
全業種	推計値	-444,100	-2,108,126	-1,749,549	1,967,621	3,266,877	1,715,645	6,593,346	9,241,714
	実績値	27,294	363,041	5,107,651	2,523,269	2,475,234	625,514	444,187	11,566,190

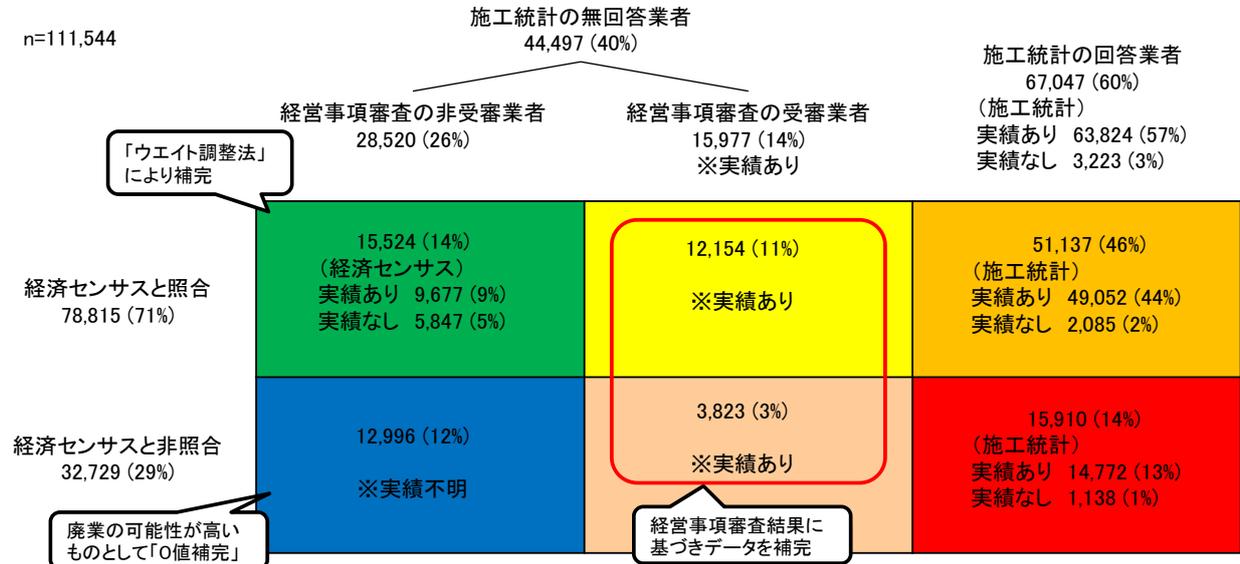
注) ()内は標準誤差。
***, **, *は、それぞれ1%、5%、10%水準で有意であることを示す。

○欠測値補完方法のさらなるブラッシュアップのため、経済センサスに加え、経営事項審査結果を活用する方法を検討。
○経営事項審査とは、公共工事の入札参加を希望する業者が受ける審査であり、建設業許可業者約47万業者のうち約14万業者が受審。透明性の確保等の観点から経営事項審査結果はWEB上でも公表されており、完成工事高等の情報を活用することが可能。



公共工事の多様性を踏まえて、客観点及び発注者点により、総合点数を算出し、発注標準(規模・工種などにより市場をグルーピングしたものに)に適合する企業を仕分ける(格付)

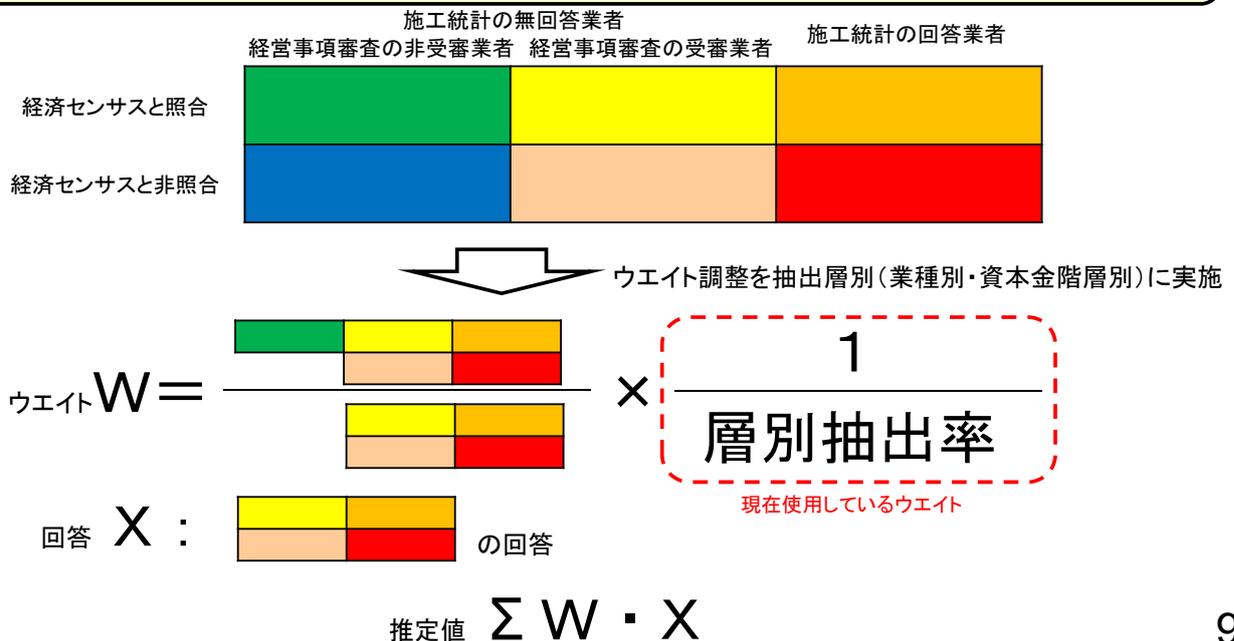
- 経営事項審査結果も活用し、欠測値補完方法を以下のとおり見直すこととしたい。
- ①施工統計の無回答業者約4.4万業者について、経営事項審査結果と照合したところ、約1.6万業者の実績を捕捉。経営事項審査結果により把握できる完成工事高等のデータを補完
 - ②その上で、残る無回答業者に対し、平成28年経済センサスー活動調査と照合し、さらに約1.6万業者の実績を捕捉。「ウエイト調整法」により補完



8

<「ウエイト調整法」による補完(改良後)>

施工統計の無回答業者かつ経営事項審査の非受審業者のうち、経済センサスと照合できた者(緑色部分)が施工統計の回答業者および経営事項審査の受審業者(橙色部分+赤色部分+黄色部分+桃色部分)と同じ分布であると仮定し、抽出層別にウエイト調整を行う。



9

○平成28年経済センサス-活動調査及び経営事項審査結果を活用した欠測値補完の試行結果は、平成28年経済センサス-活動調査の結果とほぼ合致。

	建設業が主産業	建設業が従産業
建設業許可あり		建設工事施工統計調査
建設業許可なし	経済センサス-活動調査の建設業部門集計	

建設工事施工統計調査の
部分の結果
(公表結果)

	施工統計	
	現行 A	
集計対象サンプルサイズ	67,047	
事業者数	217,566	
従業者数(千人)	2,788	
完工高(10億円)	88,247	
元請完工高	56,413	
土木工事	14,598	
建築工事・建築設備工事(住宅)	15,540	
新設	12,290	
維持・修繕	3,250	
建築工事・建築設備工事(非住宅)	19,805	
新設	13,512	
維持・修繕	6,293	
機械装置等工事	6,467	
下請完工高	31,833	
土木工事	9,212	
建築工事・建築設備工事(住宅)	5,958	
建築工事・建築設備工事(非住宅)	13,294	
機械装置等工事	3,368	

建設工事施工統計調査の
部分の結果
(補完試行結果)

	施工統計		C/B
	現行 B	補完 C	
集計対象サンプルサイズ	45,280	45,280	100%
事業者数	153,244	191,488	125%
従業者数(千人)	1,929	2,212	115%
完工高(10億円)	70,427	77,355	110%
元請完工高	48,310	51,990	108%
土木工事	14,915	15,713	105%
建築工事・建築設備工事(住宅)	11,952	13,469	113%
新設	9,890	10,990	111%
維持・修繕	2,056	2,474	120%
建築工事・建築設備工事(非住宅)	18,755	19,904	106%
新設	13,403	14,092	105%
維持・修繕	5,352	5,812	109%
機械装置等工事	2,687	2,904	108%
下請完工高	22,118	25,364	115%
土木工事	7,738	8,591	111%
建築工事・建築設備工事(住宅)	3,446	4,208	122%
建築工事・建築設備工事(非住宅)	9,206	10,607	115%
機械装置等工事	1,728	1,959	113%

経済センサス-活動調査の
部分の結果
(A表集計結果)

センサス D	B/D		C/D	
集計対象サンプルサイズ	-	-	-	-
事業者数	175,350	87%	109%	
従業者数(千人)	2,348	82%	94%	
完工高(10億円)	78,225	90%	99%	
元請完工高	53,547	90%	97%	
土木工事	16,077	93%	98%	
建築工事・建築設備工事(住宅)	10,371	115%	130%	
新設	-	-	-	
維持・修繕	-	-	-	
建築工事・建築設備工事(非住宅)	19,681	95%	101%	
新設	-	-	-	
維持・修繕	-	-	-	
機械装置等工事	4,016	67%	72%	
下請完工高	24,489	90%	104%	
土木工事	8,754	88%	98%	
建築工事・建築設備工事(住宅)	3,063	113%	137%	
建築工事・建築設備工事(非住宅)	9,070	101%	117%	
機械装置等工事	2,909	59%	67%	

※ 建設工事施工統計調査の従業者数は、建設業の従業者数+建設業以外の部門の常雇数。

10

○経済センサス-活動調査に基づきウエイトを調整するため、5年に1度、ウエイトを見直すことにより、統計に段差が生じることとなる。
⇒今後、新たな経済センサスの公表に伴いウエイトを見直す場合には、過去の公表分も含めて遡及改定を行うこととする。

<遡及改定のイメージ>

施工統計	適用するウエイト
令和2年度実績分	平成28年経済センサス (平成27年度実績)
令和3年度実績分	平成28年経済センサス (平成27年度実績)

施工統計	適用するウエイト
令和2年度実績分	令和3年経済センサス (令和2年度実績)
令和3年度実績分	令和3年経済センサス (令和2年度実績)
令和4年度実績分	令和3年経済センサス (令和2年度実績)

令和4年度実施分の公表にあわせて、2か年分遡及改定

令和3年経済センサス-経済調査公表

11

建設工事施工統計調査における欠測値補完の見直しについて(案) 参考資料

令和2年10月
国土交通省総合政策局情報政策課
建設経済統計調査室

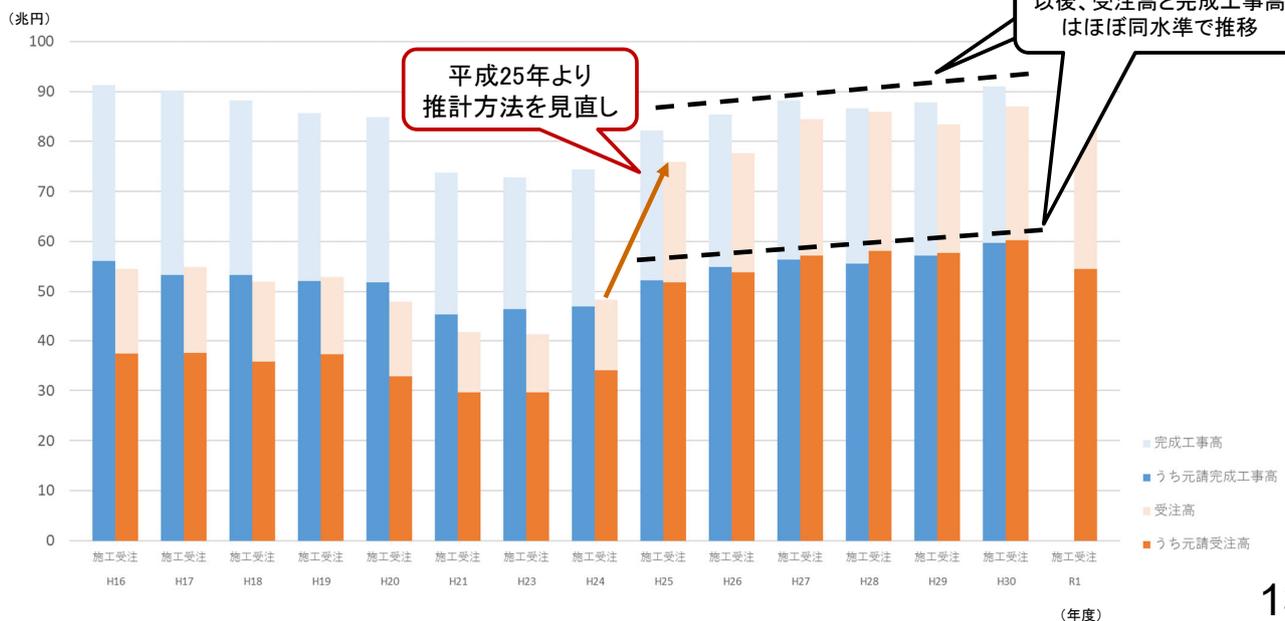


Ministry of Land, Infrastructure, Transport and Tourism

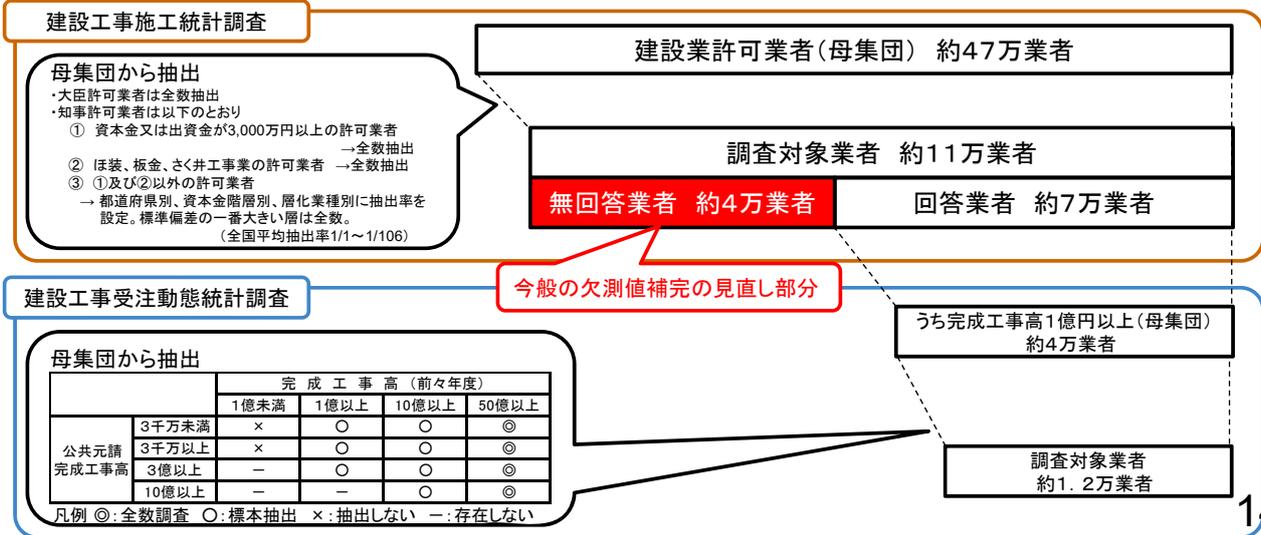
建設工事受注動態統計調査への影響①



○建設工事受注動態統計調査については、平成25年より推計方法を変更(抽出率の逆数に加えて、都道府県別・抽出層別の回収率を加味)。この結果、同調査の受注高と、建設工事施工統計調査の完成工事高はほぼ同水準で推移。



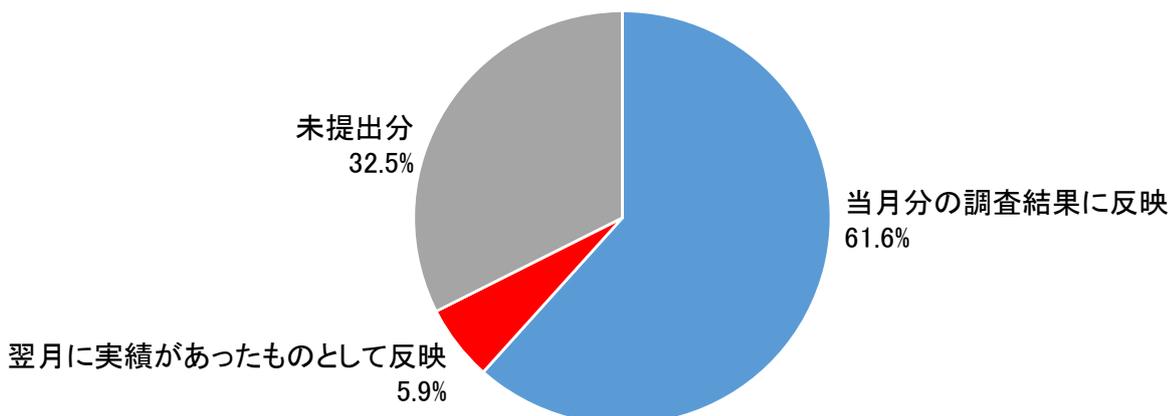
○建設工事受注動態統計調査は、前年度の建設工事施工統計調査の対象業者(約11万業者)のうち、回答業者(約7万業者)の中から抽出しているが、調査結果の復元に当たっては、無回答業者(約4万業者)は実績なしとみなして推計しており、建設工事施工統計調査と同様に、調査結果が過少になっていると見込まれる。
 ⇒このため、今般の建設工事施工統計調査における欠測値補完方法の見直しにあわせて、建設工事受注動態統計調査の推計方法についても見直しを行う。



○建設工事受注動態統計調査について、報告者のやむを得ない事情等により提出期限(翌月10日)から遅れて提出があった調査票については、可能な限り当月分の調査結果に反映させるよう柔軟な運用を行っているところであるが、それでも間に合わない調査票は、翌月に実績があったものとして計上しているところ。
 ⇒今般の建設工事受注動態統計調査の推計方法の見直しにあわせて、遅れて提出があった調査票についても当月分の調査結果に適正に反映すべく、毎年度の年度報の公表にあわせて遡及改定を行うこととする。

<調査票の反映状況>

※数字は令和2年1~6月までに調査対象業者(約1.2万業者)から回答があった調査票の反映状況を月平均したもの



○産業連関表(建設部門)においては、「建設補修」及び「その他の土木工事」の生産額の算定に、建設工事施工統計調査の元請完成工事高を利用
⇒今般の建設工事施工統計調査における見直し内容は、次回の産業連関表に反映。

IO部門分類	生産額推計元データ
建設補修	建設工事施工統計における維持・修繕工事の元請完成工事高を利用
その他の土木建設	民間構築物の生産額の推計に、建設工事施工統計における民間の新設土木工事等の元請完成工事高を利用

経済産業省生産動態統計調査の欠測値補完方法の分析結果を受けて

経済産業省調査統計グループ

鉦工業動態統計室

令和3年4月15日

シミュレーションの追加分析について

今回は、LOCFの以下の条件からLOCFの影響を受けやすい「機械器具月報(その40)自動車」を選定しシミュレーションをお示したところ。

- ア) 販売金額が高額
- イ) 当該品目を生産又は販売している事業所数が少ない
- ウ) 販売金額別の事業所数の分布に偏りがみられる

前回分科会において「サンプルサイズが小さい自動車は特殊、電気機械等、機械類でもう一つ検証してもらいたい。」との意見をいただき、別の調査票(月報)を用いて再検証を行う必要があるとされた。

LOCFの継続による結果精度への影響を明らかにするため、引き続き追加分析を総務省統計研究研修所に実施して頂いた。

追加分析は、サンプルサイズの大きい「段ボール」と、委員より具体的に指摘のあった「電気機械」の中からサンプルサイズの大きい「電子部品」を対象として検証。

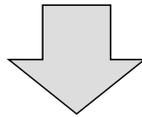
作業量の制約から、「段ボール」(3品目)、「電子部品」のうちからサンプルサイズの大きい品目を8品目(2003年1月～2017年12月通して存在する品目)選定して検証作業を行った。

シミュレーションの結果評価について

- 時間が経過すれば、乖離は拡大するものの、一部の特殊な品目を除き長期的に乖離が極端に拡大する傾向は観測されなかった。
- 数年程度であれば、一部の特殊な品目を除き、LOCFは有効に機能

考察されること

- (1) 前回分科会において示された総務省の提案
金額の変動が大きくない品目であって補完期間が5年に至るものについて、5年毎に実施される経済センサス-活動調査を活用し、経済センサス-活動調査の当該事業所の伸び率等でLOCF値を補正するなどの対応を行う。
複数年に渡り金額の大きな変動が続いている品目については、2年以内に同じ品目内の伸び率等を乗じてLOCF値を補正し、併せて、5年毎の経済センサス-活動調査の当該事業所の結果を用いて補正するなどの対応を行う。
- (2) これらを踏まえるとともに、今回の統計研究研修所による検証や実務への適用可能性を踏まえ、以下の通り対応することとしたい。



変動の大きな品目への逐次対応は年間補正の中で行い、安定した推移をしている品目は、経済センサス-活動調査の無回答事業者と同一の従業者規模等により客体を分割した情報を用いてLOCFの値を5年毎に見直すこととする

2

統計調査実施上の運用について

- 月次統計は、月遅れ提出調査票や、確報後の事業所からの修正等は、毎年、1月分確報公表時に前年の12か月分のデータ修正を行う（年間補正）。
- 上記のタイミングで、月次統計の報告を頂けない事業所に対し、決算（年度決算のみならず、半期や四半期決算も考慮）値を参考に報告頂けるように依頼し、月次分割した水準補正を行うことも考慮。
- 年間数値についても報告頂けない場合、事業所を訪問して協力要請するなど、適切な対応を行う事としたい。
- 上記の依頼・要請を最大限実施した上で、報告をいただけない事業所については、LOCFにより欠測値補完を行うが、この中で数年に渡り金額の大きな変動が続いている品目については、当該年間補正作業の中で品目内の伸び率等を乗じてLOCF値の補正を行う。
- 上記の活動と併せ、5年毎に実施する「経済センサス-活動調査」の個票データを用いて（統計法第32条に基づく調査票情報の二次利用）、品目毎に同一従業者規模等無回答事業所と同一の層に属すると思われる事業所集団の傾向を用いてLOCF値の修正を行いたい。

3

経済産業省生産動態統計調査 欠測値補完法の妥当性に関する分析結果報告書 (追加検証)

令和3年4月
総務省 統計研究研修所

© 2021 SRTI, MIC

目次

	ページ
1. 要旨	2
2. 緒言	3 - 4
2.1 背景	
2.2 目的	
3. 分析方法	5 - 16
3.1 使用データ	
3.2 クロスバリデーションに類似した方法によるシミュレーション	
3.3 計算方法等	
4. 結果及び評価	17 - 24
4.1 データ基礎属性	
4.2 長期に渡る「過去の回答結果を使用した単一補完」の影響評価	
4.3 改善案の検討	
5. 結語	25

© 2021 SRTI, MIC

1. 要旨

分析課題の整理

- ・ **統計委員会による指摘**
明確な使用期限なく「過去の回答結果を使用した単一補完」（以下、「LOCF*」。）が行われており、統計精度の悪化に繋がる懸念

↓↓
ポリシーの策定が必要

- ・ **欠測値検査の実施**
シミュレーションを実施し
 - ・ 補完データの使用期限
 - ・ より適切な対応方法を報告

Last Observation Carried Forward

「電子部品」及び「段ボール」に係る調査票を用いて検証

長期LOCFの影響評価

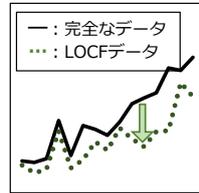


図1.1 LOCFによる乖離のイメージ

完全なデータとの乖離を分析

- ・ シミュレーションを行い、完全なデータと、欠測を発生させLOCFしたデータとを比較した
- ・ 品目によるが、数年であれば顕著な精度の悪化は起きない

精度改善案の提案



図1.2 伸び率によるLOCF値の補正イメージ

「伸び率」による補正

- ・ 直近の回答値をベースに他の事業所の伸び率でLOCF値を補正
- ・ 偏りの抑制・安定化に効果
- ・ 急激な変動や、事業所の追加又は削除による精度悪化のリスクが存在

2.1. 緒言：背景

◆統計委員会での検討と指摘

- ・ **平成28年度統計法施行状況に関する審議結果報告書（統計精度検査編）における指摘事項**

経済産業省生産動態統計調査など、一部の統計調査で「明確な使用期限を定めことなく過去の回答結果を使用した単一補完が行われており、長期にわたってデータを使用し続けた場合、統計精度の悪化につながる懸念がある。」旨が指摘

経済産業省所管の調査については「回収率90%以上を保持していることから、致命的な対応とは判断されないが、平成30年度から総務省の支援を得て、計画的にシミュレーションを行いデータの使用期限も含めた補完方法の検証を行うことが必要である。」



- ・ **平成30年度オプション検査の実施**

総務省統計研究研修所が、上記シミュレーションを実施するとともにより適切な欠測値補完方法を提案

2.2. 緒言：目的

◆主目的：LOCFの継続による結果精度への影響を明らかにする

1. LOCFを長期継続した場合の影響評価

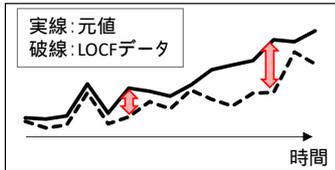


図2.2 LOCFによる乖離のイメージ

有効な回答が得られている事業所の一部に欠測を発生させて、現行の集計方法と同様、LOCFによるシミュレーションを実施した。

図2.2の矢印で示したように、LOCFによる補完値を用いた合計値と、欠測値補完を行っていない本来の値※（以下、「元値」。）との乖離幅を、LOCFを継続した場合の影響として評価する。

シミュレーションのための審査を行い、データクリーニングを行ったため、公表結果とは一致しない。

2. 改善案の検討

上記1. の検討の結果、各品目又は産業ごとの生産（販売）金額の水準変化により、LOCFによる補完値と元値との間に継続的な乖離が発生することが予想される。

本研究では、結果精度を改善する方法についていくつかシミュレーションを行い、その効果を比較検討する。

3.1. 分析方法：使用データ

調査票及び品目の選定

経済産業省生産動態統計調査には生產品目グループ別に109種類の調査票がある。このうち、平成30年度の「機械器具月報(その40)自動車」の追加検証として、以下の理由から「機械器具月報(その35)電子部品」及び「段ボール月報」を選んだ。

- ア 追加検証であるため、機械工業以外の業種であること
- イ 大企業中心の自動車月報（前回の検証対象）と比較して、中小企業も多く含まれること
- ウ 欠測を発生させるため、長期間の有効回答の事業所数が十分にあること

このうち、分析対象期間の時系列データを得ることが出来る、次ページ表3.1に掲げる電子部品8調査品目及び段ボール3調査品目を分析対象とした。

項目

主要な指標である、電子部品については「生産金額」を、段ボールについては「販売金額」をシミュレーションの対象とした。

分析対象期間

リーマンショック前後を含む2003年1月～2017年12月までの15年（180か月）分とした。

表3.1 分析に使用した品目一覧

・電子部品

品目番号	名称
0107	アルミ電解コンデンサ
0109	セラミックコンデンサ
0113	インダクタ (コイルを含む)
0115	フィルタ
0119	プリント基板用コネクタ
0125	両面プリント配線板
0127	多層プリント配線板 (6~8層)
0140	スイッチング電源

・段ボール

品目番号	名称
0101	両面
0102	複両面 (複々両面を含む)
0103	片面

3.2.1 分析方法：クロスバリデーションに類似した方法によるシミュレーション

・用いた手法

標本の性質を分析する目的で、クロスバリデーションに類似した方法でシミュレーションした。

図3.2はビン数 (除外事業所数) が1つでの概要を示している。この方法は、あるデータセットに対して、以下の操作を行い、母数の推定又は交差検証を行うものである。

- ①ローテーションで1事業所を全期間除外した複数(N個)のデータセットを作成
- ②各データセットの平均や総和、分散などの統計量 (θ) を計算
- ③上記②で求めたN個の統計量から母数($\hat{\theta}$)や標本誤差(S_{θ})等を計算

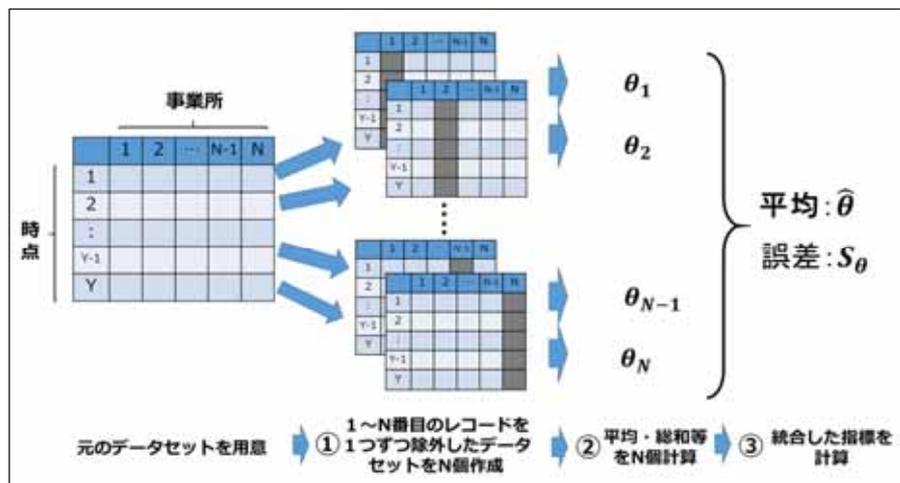


図3.2 シミュレーション方法概要

◆本分析での応用

本分析では、1事業所を全期間欠測させて、その補完を行った。
シミュレーションには、対象期間（2003年1月～2017年12月の180か月）のすべてにおいて、生産（販売）金額（元値）の回答があった事業所のデータを用いた。

1. LOCFを長期継続した場合の影響評価

現行の補完手法であるLOCFを長期間継続した場合の影響を評価した。

(1) 2003年1月の値のLOCFの継続（現行と同様の手法）

2. 改善案の提案

補完法の改善案として、以下に示した平均値代入法及び3通りのLOCFの補正によりシミュレーションを行い、結果を上記1. (1) と比較した。

(2) 平均値代入法※

(3) 年間の伸び率によるLOCFの補正

(4) 3年ごとの伸び率によるLOCFの補正

(5) 前年同月比を使用したLOCFの補正

現行のLOCFと同様に簡易な手法として、比較対象とした。

3.3.1 分析方法：集計結果の計算方法

・販売金額の集計

以下の式により、評価対象として事業所、年、月別の「生産（販売）金額」($S_{i \cdot y \cdot m}$)の年間合計値(X_y)を算出した。ここで求めた X_y は、「元値」に相当する。

$$X_y = \sum_{m=1}^{12} \sum_{i=1}^N S_{i \cdot y \cdot m} \quad \dots \textcircled{1}$$

y : 調査年 m : 調査月, i : 事業所番号 (通番), N : 事業所数

X_y : 年 y における生産（販売）金額の年間合計値, $S_{i \cdot y \cdot m}$: 第 i 番目の事業所の y 年 m 月における生産（販売）金額

・欠測値補完を行ったデータセットの販売金額の集計

以下の式により、欠測値補完を行ったデータセットの生産（販売）金額の年間合計値($x_{j \cdot y}$)及びその平均(\bar{x}_y)を算出した。

$$x_{j \cdot y} = \sum_{m=1}^{12} \left(\sum_{i \neq j} S_{i \cdot y \cdot m} + z_{j \cdot y \cdot m}(q) \right) \quad \dots \textcircled{2}, \quad \bar{x}_y = \frac{1}{N} \sum_{j=1}^N x_{j \cdot y} \quad \dots \textcircled{3}$$

j : 欠測を発生させ、補完を行う事業所番号, $x_{j \cdot y}$: 第 j 事業所を欠測値補完した年 y における生産（販売）金額の年間合計値,
 $z_{j \cdot y \cdot m}$: 第 j 番目の事業所の y 年 m 月における方法 q による補完値

3.3.2 分析方法：補完値の計算方法

欠測とみなした事業所

年月	1	2	...	M-1	M
2003/01					
2003/02					
⋮					
2017/11					
2017/12					

図3.3.1 欠測値の生成イメージ

まず、ローテーションで1事業所のデータを全期間欠測させたものを第1～N番目の事業所までN個作成する(図3.3.1)。次に、(1)～(5)の方法により欠測値を補完して、生産(販売)額の年間合計値を計算し、N個の年間合計値の平均乖離率を確認する。

以下に掲げるイメージ図(図3.3.2.1～3.3.2.5)では、第2番目の事業所を欠測させた事例について示す。

(1) 2003年1月の値のLOCFの継続

以下の式により、LOCFによる補完値 ($z_{j \cdot y \cdot m(LOCF)}$) を算出した。

$$z_{j \cdot y \cdot m(LOCF)} \leftarrow S_{j \cdot 2003 \cdot 01} \quad \dots \quad \textcircled{4}$$

j : 欠測を発生させ、補完を行う事業所番号,

$z_{j \cdot y \cdot m(LOCF)}$: 第 j 番目の事業所に対する y 年 m 月のLOCFによる補完値

年月	1	2	...	N-1	N
2003/01		100			
2003/02		100			
⋮					
2017/11		100			
2017/12		100			

図3.3.2.1 LOCFによる補完イメージ

(2) 平均値代入法

以下の式により、平均値代入による補完値 ($z_{j \cdot y \cdot m(MVS)}$) を算出した。

$$z_{j \cdot y \cdot m(MVS)} \leftarrow \tilde{x}_{j \cdot y \cdot m} = \frac{\sum_{i \neq j} S_{i \cdot y \cdot m}}{N - 1} \quad \dots \quad \textcircled{5}$$

$\tilde{x}_{j \cdot y \cdot m}$: 第 j 番目を除いた事業所の y 年 m 月における生産(販売)金額の平均値

年月	1	2	...	N-1	N
2003/01					
2003/02					
⋮					
2017/11					
2017/12					

事業所 2 を除いた各月の平均値 $\tilde{x}_{2 \cdot 2003 \cdot 01}, \tilde{x}_{2 \cdot 2003 \cdot 02}, \tilde{x}_{2 \cdot 2003 \cdot 03}, \dots$ を求める。

欠測させた事業所 2 に、求めた各月の平均値を代入する。

図3.3.2.2 平均値代入法による補完イメージ

(3) 年間の伸び率による補正補完値

以下の式により、伸び率によるLOCFの補正を行った補完値 ($z_{j \cdot y \cdot m}(GR1)$) を算出した。

$$g_{j \cdot y}(GR1) = \frac{\sum_{i \neq j} S_{i \cdot y}}{\sum_{i \neq j} S_{i \cdot y-1}} \quad (\text{ただし } g_{j \cdot 2003}(GR1) = 1) \quad \dots \textcircled{6}$$

$$z_{j \cdot y \cdot m}(GR1) = z_{j \cdot y \cdot m}(LOCF) \times \prod_{k=2003}^y g_{j \cdot k}(GR1) \quad \dots \textcircled{7}$$

$g_{j \cdot y}(GR1)$: 年 y における第 j 番目の事業所を除いて求めた前年比 (伸び率) , $S_{i \cdot y}$: 第 i 番目の事業所の y 年における生産 (販売) 金額合計

年	1	2	...	N-1	N	合計金額	伸び率 $g_{2 \cdot y}(GR1)$
2003						$\sum_{i \neq 2} S_{i \cdot 2003}$	1
2004						$\sum_{i \neq 2} S_{i \cdot 2004}$	$\frac{\sum_{i \neq 2} S_{i \cdot 2004}}{\sum_{i \neq 2} S_{i \cdot 2003}}$
:							
2016						$\sum_{i \neq 2} S_{i \cdot 2016}$	$\frac{\sum_{i \neq 2} S_{i \cdot 2016}}{\sum_{i \neq 2} S_{i \cdot 2015}}$
2017						$\sum_{i \neq 2} S_{i \cdot 2017}$	$\frac{\sum_{i \neq 2} S_{i \cdot 2017}}{\sum_{i \neq 2} S_{i \cdot 2016}}$

事業所 2 を除いた各年の合計金額 $\sum_{m \neq 2} S_{m \cdot t}$ から伸び率 (前年比) $g_{2 \cdot t}(1)$ を計算する。

年	1	2	...	N-1	N
2003		100			
2004		$100 \times g_{2 \cdot 2004}(GR1)$			
:					
2016		$100 \times g_{2 \cdot 2004}(GR1) \times g_{2 \cdot 2005}(GR1) \times \dots \times g_{2 \cdot 2016}(GR1)$			
2017		$100 \times g_{2 \cdot 2004}(GR1) \times \dots \times g_{2 \cdot 2017}(GR1)$			

2003年1月値のLOCF値 (例では100) に当該年の年間の伸び率を乗じた値で事業所2を補完する。2005年以降は、前年の補完値に、更に当該年の伸び率を乗じて補完値を算出する。

図3.3.2.3 年間の伸び率によるLOCFの補正イメージ

© 2021 SRTI, MIC

12

(4) 3年ごとの伸び率による補正補完値

以下の式により、伸び率によるLOCFの補正を行った補完値 ($z_{j \cdot y \cdot m}(GR3)$) を算出した。

$$g_{j \cdot y}(GR3) = \frac{\sum_{i \neq j} S_{i \cdot y}}{\sum_{i \neq j} S_{i \cdot y-3}} \quad (\text{ただし } \{g_{j \cdot y}(GR3) \mid y = \left. \begin{matrix} 2004, 2005, & 2007, 2008, \\ 2010, 2011, & 2013, 2014, \\ 2016, 2017 \end{matrix} \right\} = 1) \quad \dots \textcircled{8}$$

$$z_{j \cdot y \cdot m}(GR3) = z_{j \cdot y \cdot m}(LOCF) \times \prod_{k=2003}^y g_{j \cdot k}(GR3) \quad \dots \textcircled{9}$$

$g_{j \cdot y}(GR3)$: 年 y における第 j 番目の事業所を除いて求めた3年前比 (伸び率)

年	1	2	...	N-1	N	合計金額	伸び率 $g_{2 \cdot y}(GR3)$
2003						$\sum_{i \neq 2} S_{i \cdot 2003}$	1
:							
2006						$\sum_{i \neq 2} S_{i \cdot 2006}$	$\frac{\sum_{i \neq 2} S_{i \cdot 2006}}{\sum_{i \neq 2} S_{i \cdot 2003}}$
:							
2009						$\sum_{i \neq 2} S_{i \cdot 2009}$	$\frac{\sum_{i \neq 2} S_{i \cdot 2009}}{\sum_{i \neq 2} S_{i \cdot 2006}}$
:							

2003年を基点とし、(3)同様3年ごとに伸び率を計算する。

年	1	2	...	N-1	N
2003		100			
2004		$100 \times g_{2 \cdot 2004}(GR3) (=1)$			
2005		$100 \times 1 \times g_{2 \cdot 2005}(GR3) (=1)$			
2006		$100 \times 1 \times 1 \times g_{2 \cdot 2006}(GR3)$			

補完値は3年間同じ値を使用する。

図3.3.2.4 3年ごとの伸び率によるLOCFの補正イメージ

© 2021 SRTI, MIC

13

(5) 前年同月比によるLOCFの補正補完値

以下の式により、前年同期比によるLOCF値の補正を行った補完値 ($z_{j \cdot y \cdot m(RPV)}$) を算出した。

$$r_{j \cdot y \cdot m} = \frac{\sum_{i \neq j} S_{i \cdot y \cdot m}}{\sum_{i \neq j} S_{i \cdot y-1 \cdot m}} \quad (\text{ただし } r_{j \cdot 2003 \cdot m} = 1) \quad \dots \quad \textcircled{10}$$

$$z_{j \cdot y \cdot m(RPV)} = z_{j \cdot y \cdot m(LOCF)} \times \prod_{k=2003}^y r_{j \cdot k \cdot m} \quad \dots \quad \textcircled{11}$$

$r_{j \cdot y \cdot m}$: 第 j 番目の事業所を除いた y 年 m 月における前年同月比

年/月	1	2	...	N-1	N	合計金額	前年同月比 $r_{2 \cdot t \cdot q}$
2003/01						$\sum_{i \neq 2} S_{i \cdot 2003 \cdot 01}$	1
2003/02						$\sum_{i \neq 2} S_{i \cdot 2003 \cdot 02}$	1
:							
2004/01						$\sum_{i \neq 2} S_{i \cdot 2004 \cdot 01}$	$\frac{\sum_{i \neq 2} S_{i \cdot 2004 \cdot 01}}{\sum_{i \neq 2} S_{i \cdot 2003 \cdot 01}}$
2004/02						$\sum_{i \neq 2} S_{i \cdot 2004 \cdot 02}$	$\frac{\sum_{i \neq 2} S_{i \cdot 2004 \cdot 02}}{\sum_{i \neq 2} S_{i \cdot 2003 \cdot 02}}$
:							
2005-01						$\sum_{i \neq 2} S_{i \cdot 2005 \cdot 01}$	$\frac{\sum_{i \neq 2} S_{i \cdot 2005 \cdot 01}}{\sum_{i \neq 2} S_{i \cdot 2004 \cdot 01}}$
2005/02						$\sum_{i \neq 2} S_{i \cdot 2005 \cdot 02}$	$\frac{\sum_{i \neq 2} S_{i \cdot 2005 \cdot 02}}{\sum_{i \neq 2} S_{i \cdot 2004 \cdot 02}}$
:							

事業所 2 を除いた合計金額から前年同月比を計算する。

年/月	1	2	...	N-1	N
2003/01		100 × 1			
2003/02		100 × 1			
:					
2004/01		100 × $r_{2 \cdot 2004 \cdot 01}$			
2004/02		100 × $r_{2 \cdot 2004 \cdot 02}$			
:					
2005/01		100 × $r_{2 \cdot 2004 \cdot 01}$ × $r_{2 \cdot 2005 \cdot 01}$			
2005/02		100 × $r_{2 \cdot 2004 \cdot 02}$ × $r_{2 \cdot 2005 \cdot 02}$			
:					
2017/12		100 × $r_{2 \cdot 2004 \cdot 12}$ × ... × $r_{2 \cdot 2017 \cdot 12}$			

LOCF値 (例では100) に前年同月比を乗じた値で事業所 2 を補完する。
2005年以降は、補完値に更に当該年の前年同月比を乗じて補完値を算出する。

© 2021 SRTI, MIC 図3.3.2.5 前年同月比によるLOCFの補正イメージ

3.3.3 分析方法：評価指標の計算方法

以下の式により、 $j = 1 \sim N$ までの N 個全ての事業所を順に欠測させ補完して計算した生産 (販売) 金額の年間合計値 ($x_{j \cdot y}$) ごとに、元値 (X_y) との乖離率 ($e_{j \cdot y}$) を求めた。乖離率は、偏りを評価するための「符号付」(式 - a) と、ばらつきを評価するための「絶対値」(式 - b) の2通り計算した。

そこから、 N 個のデータセットから求めた乖離率を平均し、平均乖離率 (\bar{e}_y) を求め、評価指標とした。

$$e_{j \cdot y} = \frac{x_{j \cdot y} - X_y}{X_y} \quad \dots \quad - a \quad , \quad e_{j \cdot y} = \frac{|x_{j \cdot y} - X_y|}{X_y} \quad \dots \quad - b \quad ,$$

$$\bar{e}_y = \frac{1}{N} \sum_{j=1}^N e_{j \cdot y} \quad \dots \quad \textcircled{13}$$

y : 調査年, N : 事業所数, j : 欠測を発生させ、補完を行う事業所番号,
 $x_{j \cdot y}$: 第 j 事業所を欠測値補完した年 y における生産 (販売) 金額の年間合計値,
 X_y : 年 y における生産 (販売) 金額の年間合計値,
 $e_{j \cdot y}$: 第 j 事業所を欠測値補完した年 y における乖離率,
 \bar{e}_y : 年 y における N 個のデータセットの乖離率の平均

3.3.4 分析方法：ウィンドウ分析

全期間を用いたシミュレーションは、2008年のリーマンショックなど、急激な水準変化による影響を受ける可能性がある。そこで、LOCFの長期継続の影響評価を補強するため、「ウィンドウ分析」を行った。

ウィンドウ分析では、以下の式に示したとおり、LOCFの開始年 $a = \{2003, 2004, 2005, 2006, 2007, 2008\}$ を1年ずつずらした10年間のLOCFのシミュレーションを、2003年開始から2008年開始まで6セット行い、LOCFの開始後の経過年数 b ごとの平均乖離率の平均 (\bar{e}_b) を求めた。

$$z_{j \cdot y \cdot m \cdot a}(\text{LOCF}) = S_{j \cdot a \cdot 01} \dots \textcircled{14}, \quad x_{j \cdot y \cdot a} = \sum_{m=1}^{12} \left(\sum_{i \neq j} S_{i \cdot y \cdot m} + z_{j \cdot y \cdot m \cdot a}(\text{LOCF}) \right) \dots \textcircled{15}$$

$$e_{j \cdot y \cdot a} = \frac{x_{j \cdot y \cdot a} - X_y}{X_y} \quad (\text{又は } e_{j \cdot y \cdot a} = \frac{|x_{j \cdot y \cdot a} - X_y|}{X_y}) \dots \textcircled{16}, \quad \bar{e}_{y \cdot a} = \frac{1}{N} \sum_{j=1}^N e_{j \cdot y \cdot a} \dots \textcircled{17}, \quad \bar{e}_b = \frac{1}{6} \sum_{a=2003}^{2008} \bar{e}_{(a+b) \cdot a} \dots \textcircled{18}$$

j : 欠測を発生させ、補完を行う事業所番号, a : LOCFの開始年,
 $S_{i \cdot y \cdot m}$: 第 i 番目の事業所の y 年 m 月における生産 (販売) 金額
 $z_{j \cdot y \cdot m \cdot a}(\text{LOCF})$: 第 j 番目の事業所の a 年から開始したLOCFによる y 年 m 月の補完値
 $x_{j \cdot y \cdot a}$: 第 j 事業所を a 年から開始したLOCFにより欠測値補完したときの年 y における生産 (販売) 金額の年間合計値,
 $e_{j \cdot y \cdot a}$: 第 j 事業所を a 年から開始したLOCFにより欠測値補完したときの年 y における元値との乖離率,
 $\bar{e}_{y \cdot a}$: 年 y における N 個のデータセットの $e_{j \cdot y \cdot a}$ の平均
 \bar{e}_b : 年 a から b 年経過した時の $a = 2003 \sim 2006$ の6個の平均乖離率の平均値

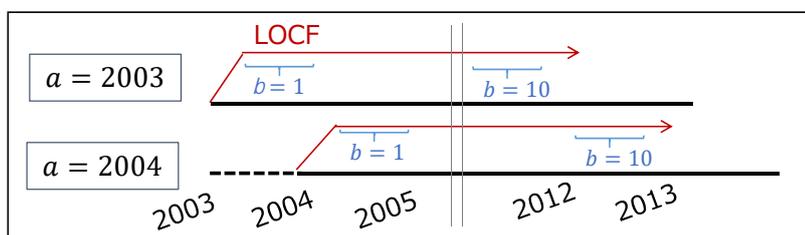


図3.3.3 ウィンドウ分析の模式図

16

4.1. 結果及び評価：データの基礎属性 (電子部品)

「機械器具月報(その35)電子部品」について、以下の特徴を確認した。

- ア 事業所間の生産金額の分布に偏りがみられた。(図4.1.1)
- イ 今回LOCFに用いた1月の季節指数は、0.93である。(図4.1.2及び資料編2章)

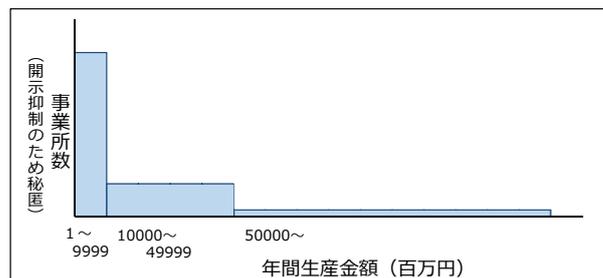


図4.1.1 セラミックコンデンサ(例)の年間生産金額階級別事業所数

今回分析した品目は、生産金額の偏りが大きい。また、LOCFの起点に用いるのは2003年1月だが、1月は12か月で概ね最も小さい値であり、また、2003年の生産金額は、以降の年次と比べて、多くの品目において生産金額の水準が比較的低い。ここから、本分析で行うシミュレーションは、乖離率が比較的大きい値をとりやすい(結果が振れやすい)といった条件の悪いケースであるといえる。

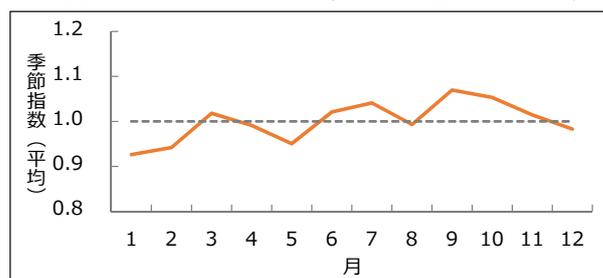


図4.1.2 生産金額の季節指数(2003-2017年の月別平均の全品目平均)
 ※季節指数はX-12-ARIMAにより算出した。条件は資料編2章の参考を参照

17

4.1.結果及び評価（つづき）：データの基礎属性（段ボール）

「段ボール月報」について、以下の特徴を確認した。

- ア 事業所間の販売金額の分布に偏りがみられた。（図4.1.3）
- イ 今回LOCFに用いた1月の季節指数は、0.86である。（図4.1.4及び資料編2章）

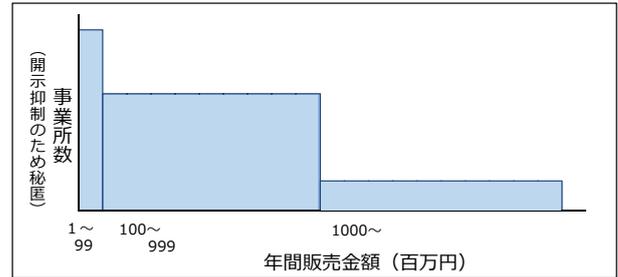


図4.1.3 両面（例）の年間販売金額階級別事業所数

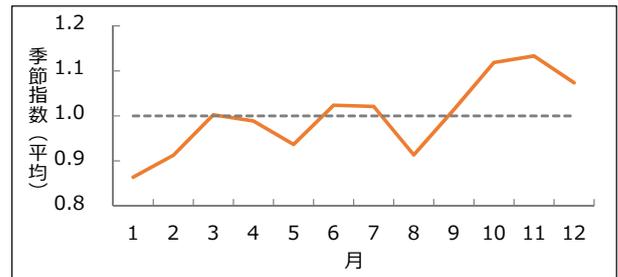


図4.1.4 販売金額の季節指数（2003-2017年の月別平均の全品目平均）
※季節指数はX-12-ARIMAにより算出した。条件は資料編2章の参考を参照

今回分析した品目は、販売金額の偏りが大きい。また、LOCFの起点に用いるのは2003年1月だが、1月は12か月で最も小さい値である。ここから、本分析で行うシミュレーションは、乖離率が比較的大きい値をとりやすい（結果が振れやすい）といった条件の悪いケースであるといえる。

4.2.結果及び評価（つづき）：長期LOCFの影響（電子部品）

現行手法と同様にLOCFを長期間継続した場合のシミュレーション結果の特徴は以下のとおり。（資料編1、4、5章）

- ア 平均乖離率（符号付）は、生産金額に対してほぼ逆の増減を示す（全ての品目でPearsonの相関係数が -0.90 より小さい）。（図4.2.1）
- イ 平均乖離率（符号付）は大きく変動せず（図4.2.1）、平均乖離率（絶対値）も10年程度のLOCFで5%程度。（図4.2.1）

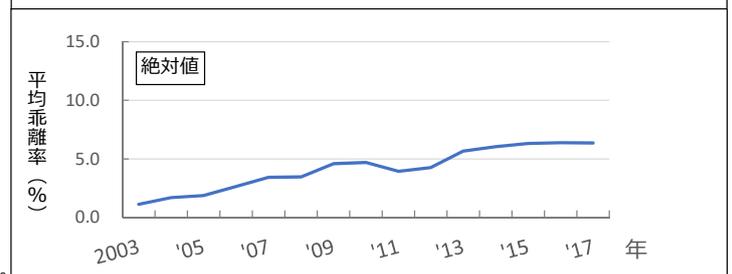
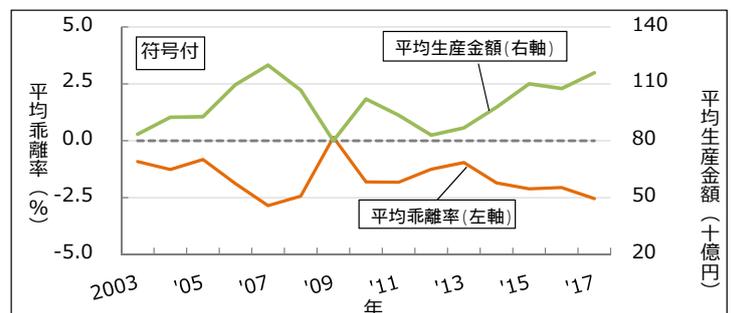


図4.2.1 LOCFの継続による生産金額及び平均乖離率の全品目平均の推移

LOCFの開始時点から、生産金額の水準が大きく離れる品目では、LOCFの補完による乖離が大きくなる傾向がある。

しかし、数年程度のLOCFでは顕著に結果が悪化することはない*。

*「セラミックコンデンサ」や「フィルタ」のように、全体的水準が急激に変化する場合、顕著な乖離が発生しやすい。詳しくは資料編の個別品目を参照。

4.2.結果及び評価（つづき）：長期LOCFの影響 評価（電子部品）

更に、ウィンドウ分析によって得られた結果は以下のとおり。（図4.2.2 及び 資料編6、7章）

- ア 継続期間が10年程度であれば、平均乖離率（符号付）は5%以内（「フィルタ」を除く）。
- イ 平均乖離率（絶対値）は、8品目のうち2品目が5年以内に5%を上回った。



平均乖離率（符号付）の結果から、ほぼすべての品目で、LOCFにより発生する偏りは結果精度に重大な影響を及ぼすことはないと考えられる。

しかし、平均乖離率（絶対値）の結果から、LOCFを長期継続した場合、上記結果がばらつく可能性があり、注意が必要である。特に、「フィルタ」のように、経年により生産金額の水準に極端な違いが生じる場合、LOCFによる補完には限界がある。

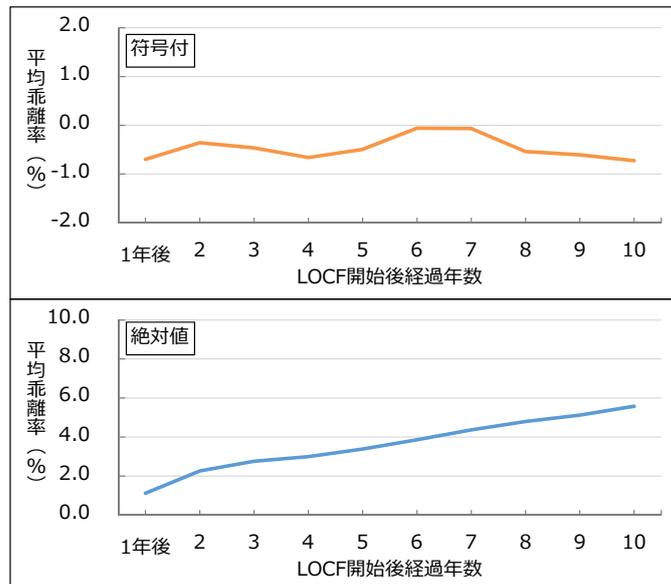


図4.2.2 ウィンドウ分析によるLOCF開始後経過年数による平均乖離率の全品目平均の推移

4.2.結果及び評価（つづき）：長期LOCFの影響 評価（段ボール）

現行手法と同様にLOCFを長期間継続した場合のシミュレーション結果の特徴は以下のとおり。（資料編1、4、5章）

- ア 平均乖離率（符号付）は、販売金額に対してほぼ逆の増減を示す（全ての品目でPearsonの相関係数が -0.90 より小さい）。（図4.2.3）
- イ 平均乖離率（符号付）は大きく変動せず（図4.2.1）、平均乖離率（絶対値）も15年程度のLOCFで5%未満。（図4.2.3）



LOCFの開始時点から、販売金額の水準が大きく離れる品目では、LOCFの補完による乖離が大きくなる傾向がある。

しかし、数年程度のLOCFでは顕著に結果が悪化することはない。

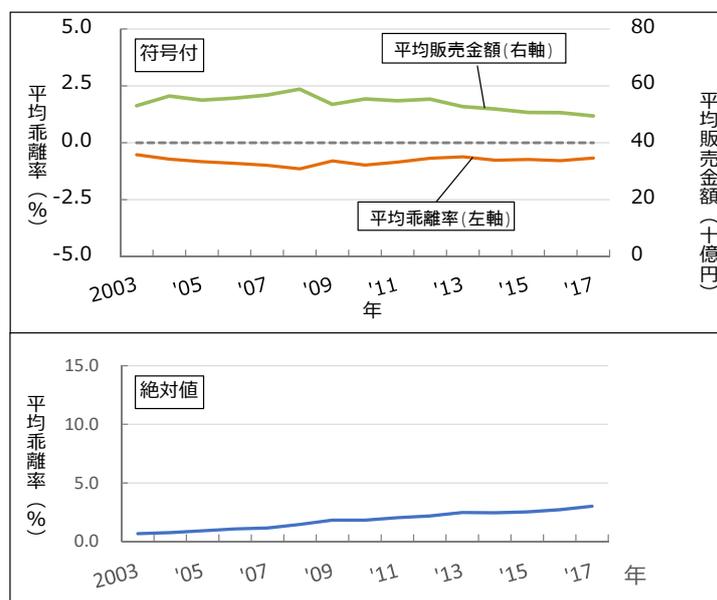


図4.2.3 LOCFの継続による販売金額及び平均乖離率の全品目平均の推移

4.2. 結果及び評価（つづき）：長期LOCFの影響評価（段ボール）

更に、ウィンドウ分析によって得られた結果は以下のとおり。（図4.2.4 及び 資料編6、7章）

ア 継続期間を10年程度としても、平均乖離率（符号付）は、全品目とも5%以内。

イ 平均乖離率（絶対値）は、5年以内で、全品目とも5%未満。

平均乖離率（符号付）の結果から、ほぼすべての品目で、LOCFにより発生する偏りは結果精度に重大な影響を及ぼすことはないと考えられる。

しかし、平均乖離率（絶対値）の結果から、LOCFを長期継続した場合、上記結果がばらつく可能性があり、注意が必要である。特に、「片面」のように、事業所間の販売金額の水準に偏りがみられる場合、LOCFによる補完には限界がある。

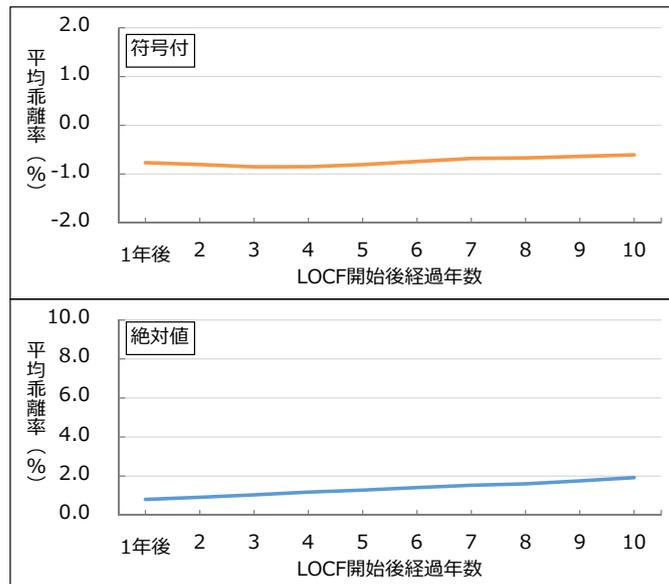


図4.2.4 ウィンドウ分析によるLOCF開始後経過年数による平均乖離率の全品目平均の推移

4.3 結果及び評価（つづき）：改善案の検討（電子部品）

各種改善法を用いたシミュレーションにより得られた示唆は以下のとおり。（図4.3.1 及び 資料編4、5章）。

ア 平均乖離率（符号付）は、3種類の補正方法を用いることでLOCF開始時の水準が維持された。（平均値代入法では計算の性質から0となった。）

イ 平均乖離率（絶対値）は、平均値代入法で高水準となり、他の3種類の補正方法ではLOCFと差が見られなかった。

平均値代入法は、事業所間の偏った生産金額の分布から、極端な値を持つ事業所の数値に影響されやすく、欠測値補完法としては不適切。

伸び率等による補正は、顕著な改善効果は見られないが、「フィルタ」のように、経年により生産金額の水準の極端な変化から大きな乖離が発生する場合において、LOCFの補完と併用することが効果的であると考えられる。

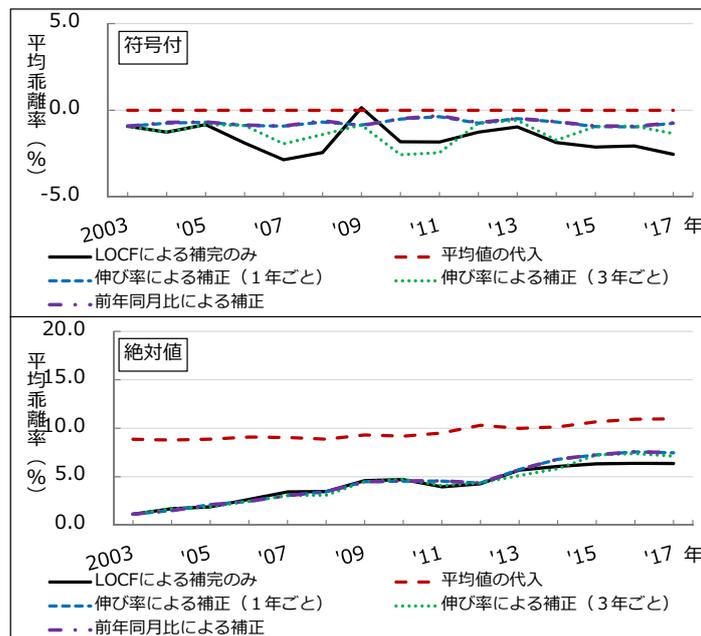


図4.3.1 各種補完法による平均乖離率の全品目平均

4.3 結果及び評価（つづき）：改善案の検討（段ボール）

各種改善法を用いたシミュレーションにより得られた示唆は以下のとおり。（図4.3.2 及び 資料編4, 5章）。

- ア 平均乖離率（符号付）は、3種類の補正方法を用いることでLOCF開始時の水準が維持された。（平均値代入法では計算の性質から0となった。）
- イ 平均乖離率（絶対値）は、平均値代入法で比較的高水準となり、他の3種類の補正方法ではLOCFと差が見られなかった。

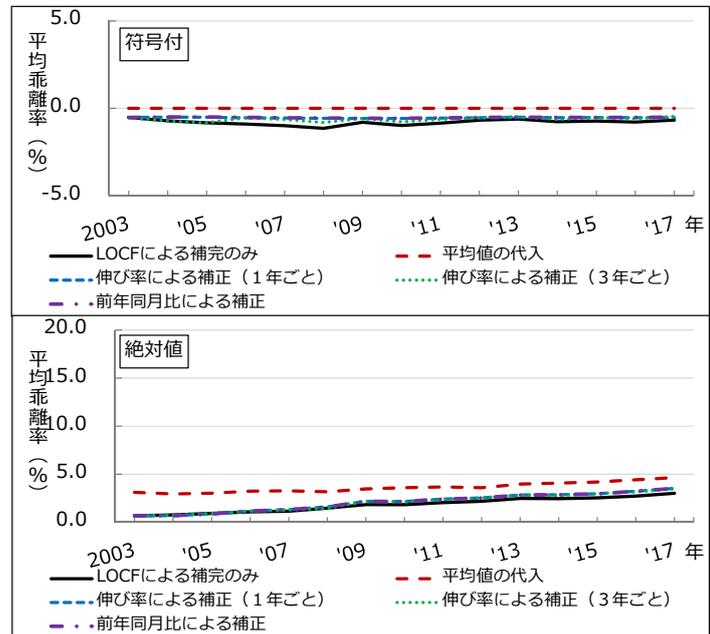


図4.3.2 各種補正法による平均乖離率の全品目平均

24

平均値代入法は、事業所間の偏った販売金額の分布から、極端な値を持つ事業所の数値に影響されやすく、欠測値補完法としては不適切。

事業所数が十分に確保できる場合は、一つの事業所の欠測に対してLOCFを継続しても大きな乖離は発生しない。

伸び率等による補正は、顕著な改善効果は見られないが、「片面」のように、事業所数が少ないことや事業所間の販売金額の水準に偏りがみられることから、大きな乖離が発生する場合において、LOCFの補完と併用することが効果的であると考えられる。

© 2021 SRTI, MIC

5. 結語

（1）LOCFの長期継続について

- ア 長期に渡り継続しても必ずしも精度が下がる訳ではない。
- イ 生産（販売）金額の水準が経年により大きく変化する場合、補完値は元値から大きく乖離する傾向がある。
- ウ 数年程度LOCFを継続した場合、生産（販売）金額の水準の変化を確認するとともに、補正の要否を検討することが望ましい。

（2）改善案の提案について

伸び率等で、金額の水準の変化に合わせてLOCF値の補正を行うことで、真値からの乖離を一定程度に抑えることができることが分かった。特に、経年により、生産（販売）金額の水準が大きく変化した場合、この補正は効果的である。

© 2021 SRTI, MIC

25

欠測値の補完に係る主な方法等について

令和3年6月23日

総務省政策統括官(統計基準担当)

目次

はじめに

1 欠測値の補完に当たっての考え方

- (1) 欠測による影響
- (2) 補完に関する視点

2 補完の処理の主な手順

3 補完を行うための主な方法

< 項目の欠測への対応 >

- (1) (層化) 平均値代入 (Mean Imputation)
- (2) 回帰代入 (Regression Imputation)
- (3) 比率補完 (Ratio Imputation)
- (4) ホットデッキ法 (Hot Deck Methods)
- (5) LOCF (Last Observation Carried Forward)
- (6) その他

< 客体単位での欠測への対応 >

ウェイトを調整する方法 (Weighting Adjustments)

おわりに

はじめに

公的統計を作成するための統計調査において、調査の実施段階においては調査客体への確認や督促なども含め調査票への記入及び回収の確保に努めているところであるが、一部の調査項目が無回答である場合や、回答そのものが得られずにすべての調査項目に欠測値が発生した場合には、結果の有用性を確保するため、統計的な補完を適切に行うなどの対応が考えられる。

一方で、これらの欠測値については、統計調査ごとに欠測の発生状況や補完に利用できるデータなどに違いがあり、特定の補完方法の適用など一律の対応は困難な面がある。

このため、統計調査ごとの状況を踏まえた適切な対応が重要となるが、欠測値に関するこれまでの評価分科会における審議や各府省における取組状況等も踏まえて、補完を行うに当たっての主な方法・手順や利用上の注意点など、実務上参考となる事項を整理してまとめて示すことは、公的統計の正確性や有用性の確保を促進する上で有意義と考えられる。

2

1 欠測値の補完に当たっての考え方

(1) 欠測による影響

統計調査においては、調査項目の一部について無回答である場合や調査客体から回答が得られない場合などの欠測が発生した際には、そのまま集計すると母集団としての代表性が損なわれ、平均などの結果に偏りが発生するおそれがあり、結果利用上の有用性に影響が生じる。

(2) 補完に関する視点

欠測に対しては、基本的には調査票の回収段階においてできる限り発生しないようにすることが必要である。

その上で、最終的に発生した欠測値に対しては、統計的な処理として、可能な限り補完を行うことにより統計表及び平均などの結果の有用性を確保する。

また、補完により欠測のないデータを作成することは、マイクロデータ利用による分析などの面でも意義を有することと考えられる。

3

1 欠測値の補完に当たっての考え方

(2) 補完に関する視点(続き)

欠測への対応を考える上では、欠測が何に依存して発生しているか(欠測データメカニズム)も考慮して適切な対応をとることが必要である。

単純に欠測のない客体のみを用いて母平均を推定する場合、欠測のメカニズムが完全にランダムな場合以外は結果に偏りが生じる。

完全にランダムな欠測(MCAR)とは、変数の欠測する確率が当該変数の値及び他の観測されている変数の値に依存しない場合である。

参考: 欠測のない客体のみで平均を推定した場合の偏りについて

項目(y)に欠測が発生しているとして、母集団をyについて回答する層と回答しない層に分けると仮定し、それぞれの層の平均を μ_c 、 μ_{ic} 、対応する構成割合を c 、 $1-c$ とすると、母集団全体の平均は、

$$\mu = c\mu_c + (1-c)\mu_{ic}$$

と書ける。ここで、欠測のない客体のみから平均を推定した場合は μ_c となることから、偏りは、

$$\mu_c - \mu = (1-c)(\mu_c - \mu_{ic}) \quad (*)$$

となる。欠測の発生が完全にランダムな場合には、 $\mu_c = \mu_{ic}$ となるため、(*)の偏りは0となるが、それ以外の場合には0とはならない。

4

1 欠測値の補完に当たっての考え方

(2) 補完に関する視点(続き)

これに対し、補完を行う場合には、ランダムな欠測を仮定できるならば、観測されている他の項目(補完の対象となる項目と関連を有し、欠測のし易さとも関連する項目)を適切な補助変数として利用して補完を行うことにより、平均などの結果の偏りを緩和することができる。

ランダムな欠測(MAR)とは、変数の欠測する確率が当該変数の観測された値及び他の観測されている変数の値には依存するが、当該変数の欠測となった値には依存しない場合である。

補完のための方法としては、いくつかの方法が一般的に用いられているが、統計調査ごとに、欠測の発生状況や利用可能な補助変数の状況、実務上の適用可能性なども考慮し、適切な方法を選択することが必要である。

以下に、補完の処理の主な手順を概観した上で(6~7ページ参照)、欠測値の補完に関して利用される主な方法として、項目の欠測の場合(8~14ページ参照)及び未回収など客体単位での欠測の場合に分けて示す(15~17ページを参照)。

5

2 補完の処理の主な手順

欠測の発生状況の確認

欠測が発生しており、補完の対象となる項目の確認

欠測値が生じている客体について、欠測の発生状況や、他の項目・特定の属性等との関係が欠測の発生のし易さに影響していないかなどの特徴を把握

分布を確認することや、全ての客体において観察されている適当な項目で標本を層化し、層ごとの回収率を確認するなど



補完に利用可能な補助変数等の検討

の確認結果を踏まえ、欠測している変数と関係の強い変数や欠測のし易さに関連していると見られる項目(補助変数)などの利用可能性を検討

その他、欠測の内容に応じて他の情報(当該調査の前回等の結果や外部の関連情報等)の適切な利用可能性についても検討



6

2 補完の処理の主な手順

適切な補完方法の検討

により利用可能な補助変数等も考慮し、適切な補完の方法について検討(8~17ページを参照。なお、それぞれの方法の算式等詳細については参考資料も参照)

- ・まず、演繹的な補完に加え、過去の結果から経年で安定的なものや他の統計調査結果・行政記録情報等で適切なものがあれば利用を検討
- ・さらに、補助変数を基に欠測値を適切に予測できそうな場合は回帰補完や比率補完、ホットデック法等の検討。他には、層化平均値代入、LOCF(時点調整を含む)等の検討
- ・客体単位での欠測の場合はウェイト調整法を検討

補完を行うに際して客体の層化を行う場合、適切な層区分の方法についても検討(欠測している項目と関連し、欠測のし易さにも関連する項目(補助変数)で層化を行うのがよい)

調査項目ごとに異なる複数の補完方法を用いる場合は、補完の手順等を検討

適用する補完方法間の比較を行うには、以下のような方法がある

- ・観測されている項目の一部を欠測させるなどのシミュレーションを行い、推定値の真値からの乖離を表わす指標(平均平方誤差(RMSE)など)を利用する方法や、推定結果をセンサス調査等の情報源と比較する方法など



補完方法の選択

実務上の実行性等も勘案し、適切な補完方法を決定

補完を行った場合、その方法等について適切な情報提供にも努める

7

3 補完を行うための主な方法

< 項目の欠測 (Item Nonresponse) への対応 >

(1) (層化) 平均値代入 (Mean Imputation)

手法の概要

- ・欠測値に対し、観測された客体の値の平均値を代入する。層化平均値代入は、すべての客体を適切に層化した上で、その層内の平均値を代入に用いる。

手順

- ・層化平均値代入では、すべての客体について観測されている適切な項目 (補助変数) に基づき、欠測のある客体を含むすべての客体を層化する。
- ・各層内で、欠測値に対し、観測されている客体の平均値を代入する。

利用上の注意点等

- ・平均値代入は簡易な方法であるが、欠測が完全にランダムに発生している場合以外は、母平均の推定値には偏りが発生する。その改善として、適切な項目 (補助変数) に基づきすべての客体を層化した上で代入を行うことにより、偏りを緩和することができる。
- ・補助変数として利用する項目には、欠測している項目と関連を有し、欠測のし易さとも関連する項目を使うのがよい。
- ・なお、平均値の補完に伴い、標本分散は過小に評価される可能性がある。

8

3 補完を行うための主な方法

(2) 回帰代入 (Regression Imputation)

手法の概要

- ・欠測値に対し、回帰モデルに基づく推定値を代入する。

手順

- ・欠測が生じていない客体を用いて、欠測している項目を従属変数とし、観測されている項目を説明変数とする回帰モデルを推定する。
- ・当該回帰モデルにより推定した値 (回帰直線上の理論値) を代入値とする。

利用上の注意点等

- ・回帰モデルは、欠測値に対しよい予測値を与える可能性があるが、そのためには適切なモデリングが必要となる。
- ・説明変数に用いる変数には、連続値のほか、カテゴリカル変数などがある。なお、説明変数を一定の層への所属を表わすダミー変数とした場合には、層化平均値代入と同じものを表わす。
- ・回帰モデルによる理論値の代入に伴い、標本分散は過小に評価される可能性がある。欠測値のばらつきを考慮して、予測値に誤差項 (乱数) を加える方法は確率的回帰代入と呼ばれる。

9

3 補完を行うための主な方法

(3) 比率補完 (Ratio Imputation)

手法の概要

- ・欠測が発生している項目と他の項目との比率を利用して、代入値を算出する。

手順

- ・欠測が発生していない客体を用いて、補完の対象とする項目 (y) と他の項目 (x) との比率 (r) を計算する。
- ・欠測が生じている客体において観測されている項目 (x) に当該比率 (r) を乗じることによって得られた値を欠測値への代入値とする。
- ・比率の算出は、観測されている項目を利用して適切な層区分を設定し、それら層区分ごとに行う。

利用上の注意点等

- ・比率を算出する際に利用する項目としては、欠測が生じている項目に対して比例関係にある項目を利用するのがよい。

10

3 補完を行うための主な方法

(4) ホットデッキ法 (Hot Deck Methods)

手法の概要

- ・欠測値に対し、同じデータセットの中で、欠測が生じている客体と類似した客体 (ドナー) を探し出し、ドナーの観測値を欠測値の代わりとして代入する。
- ・客体間の距離を定義し、欠測がある客体に近い客体をドナーとする。

手順

- ・欠測が生じている客体と欠測が生じていない客体について、共通して観測されている項目 (補助変数) の値を基に一定の距離を計算し、最も距離の近い標本の観測値を欠測値に代入する。

利用上の注意点等

- ・回帰代入のようなモデルの仮定を要しないが、類似した客体を探し出すための作業が必要となる。
- ・用いる距離としては、客体に関する補助変数のベクトルに関するユークリッド距離や、マハラノビス距離などがある。マハラノビス距離は、変数間の相関を考慮した距離である。

11

3 補完を行うための主な方法

- ・また、客体のすべてについて傾向スコア^(注)を推定し、傾向スコアを距離としてその値が最も近い(差の絶対値が最小となる)客体の観測値を代入値とする方法(傾向スコアによるマッチング)もある。

(注)傾向スコア: 客体ごとの補助変数の値に応じて客体が回答する確率を表わし、ロジットモデルなどにより推定される。傾向スコアを利用する際は、欠測の発生がランダムであり、モデルが正確であることが仮定されている。

- ・他にも、観測されている項目に基づきすべての客体をセルに分類し、欠測のある客体と同じセル内に存在する欠測のない客体からランダムに選んで、その観測値を代入値とする方法などがある。
- ・なお、ドナーを同一のデータセットではなく、過去の調査結果など別のデータセットから探す場合はコールドデックと呼ばれる。過去の調査結果を利用する場合、利用するデータが経年で安定的なものであることなどが必要と考えられる。

12

3 補完を行うための主な方法

(5) LOCF (Last Observation Carried Forward)

手法の概要

- ・同一の客体を複数時点にわたって調査するパネルデータの場合において、欠測が発生した以降の各時点の値として、直近の観測値を代入値とする。
- ・欠測の発生以降、長期に適用する場合は、経時による変化等を反映させるため、何らかの調整を行うことが考えられる。

手順

- ・欠測が発生している客体について、直近の観測値を欠測値に代入する。
- ・経時による調整の例としては、欠測が生じている項目について、直近の観測値からの伸び率を欠測のない客体を用いて算出し、欠測が発生している客体の直近の観測値に乗じた値を代入値とする方法がある。

利用上の注意点等

- ・欠測が発生した以降、当該項目の値は変化しないとみなすため、補完の対象とする項目によっては長期に固定して用いた場合、妥当な推計とならない可能性がある。

13

3 補完を行うための主な方法

(6) その他

演繹的補完 (Deductive Imputation)

- ・欠測が生じている客体において、観測されている項目間の関係から、欠測している項目の値を論理的に定めることができる場合、その値により補完する。

(例)費用合計の回答があり、内訳の一つにのみ欠測が生じていた場合、引き算で算出した欠測値を補完するなど

- ・補完に際して、一番初めに取り組むべき方法と考えられる。

他の統計調査の結果、公開情報、行政記録情報等の活用

- ・欠測が生じている客体について、他の情報(他の統計調査、公開情報、行政記録情報、事業所母集団データベースの情報等)を用いて補完する。

ただし、情報の把握時点の違いや、統計上用いている定義との違いなどに注意する必要がある。

14

3 補完を行うための主な方法

< 客体単位での欠測 (Unit Nonresponse) への対応 >

未回収など調査票の項目のすべてについて回答が得られない場合の対応としては、ウェイトを調整する方法 (Weighting Adjustments) がある。

手法の概要

- ・標本設計に基づく通常のウェイトについて、個々の標本が回答する確率(回答確率)を考慮して調整した上で、推定を行う。(資料5の参考「1 欠測値補完に係る主な方法(6) ウェイト調整」を参照)

手順

- ・客体(i)ごとの回答確率(p_i)を求め、通常のウェイト(w_i : 抽出率の逆数に相当)に回答確率の逆数をかけてウェイトを調整し($w_i \times 1/p_i$)、当該調整したウェイトを用いて推定を行う。
- ・回答確率(p_i)は、客体が回答する確率の推定値であり、未回収を含むすべての客体において観測されている項目(補助変数)¹に基づき客体をクラス(Weighting Class)に分けた場合、当該客体が入るクラスにおける回答した客体の割合(回収された標本の数 / 配分された標本の数)を用いる。

1 未回収の場合は調査項目の回答は得られないので、ここでは主に調査の設計などに用いた変数となる。

15

3 補完を行うための主な方法

- ・当該客体の区分は、欠測している項目と関連を有し、欠測のし易さとも関連する項目により行うのがよい。
- ・調整したウェイトを用いて推計を行う場合は、当初配分された標本についてではなく、実際に回収された標本について合計をとることとなる。

利用上の注意点等

- ・複数の属性などの情報を利用する場合には、客体をクラスに分けるための組み合わせが増えてしまうが、複数の補助変数の情報を回答確率という観点から一つの値に集約するため、傾向スコアを利用する方法もある。その場合、この傾向スコアの値を用いて客体をグループに分ける（傾向スコアによるクラス分け）。

(注)傾向スコアを回答確率の推定値として直接用いることもできるが、その場合、傾向スコアの値が非常に小さいと推定結果への影響が過度に大きくなるため、調整後のウェイトが極端になっていないかなどの確認も必要である。

- ・なお、回答確率により修正したウェイトについて、補助変数に関する母集団総計の情報が別途把握できている場合には、それを利用してウェイト($w_i \times 1/\hat{p}_i$)を更に調整することもできる²。これは、回答確率を求めるクラスを事後層と見た事後層化推定となっている。また、複数の項目(補助変数)によってクラスを構成し、同様の操作を行う場合はレイキングと呼ばれる。

² 回答確率を求めるクラスごとに補助変数に関する母集団総計が別途把握できている場合、その補助変数に関するクラスでの総計の推計結果が当該母集団総計に一致するようにウェイトを調整するものであり、 W_i/\hat{p}_i に更に比率(クラスごとの補助変数の母集団総計/標本から求めた補助変数の総計の推定値)を乗じたものを調整後のウェイトとする。補助変数を1とした場合は、母集団のサイズに関する情報を利用することになる。

3 補完を行うための主な方法

欠測の種類	主な補完方法	補助変数の利用
項目の欠測	層化平均値代入	客体の層化
	回帰代入	説明変数
	比率補完	比率の計算
	ホットデック法	客体間の距離の計算、マッチングのための傾向スコアの算出等
	LOCF	伸び率の計算
客体単位での欠測	ウェイト調整	回答確率を計算するクラスの設定、クラス分け等のための傾向スコアの算出

今回、統計調査の実施に伴い発生する欠測値への対応として、統計作成の実務において利用が考えられる主な補完方法について概括的に整理して示した。

適用する補完の方法は、統計調査ごとの欠測の状況等を踏まえ、適切に選択することが必要であるが、これらの補完方法全体に通じることとして、欠測による結果の偏りを緩和するためには、他の観測されている項目を適切な補助変数として利用して補完を行うことが重要である。

しかしながら、統計調査ごとに利用できる補助変数の種類や内容などは異なっており、適切な補助変数が常に十分に利用できるとは限らない。統計調査の実施の段階で欠測をできるだけ発生させないようにすることが何より重要である。

(参考1) 欠測データメカニズムについて

データの欠測がどのようなメカニズムで発生しているかについて、欠測の発生が何に依存するかという視点から、以下の3つの分類が用いられる。

完全にランダムな欠測(MCAR):

- ・変数の欠測する確率が当該変数の値及び他の観測されている変数の値に依存しない場合である

ランダムな欠測(MAR):

- ・変数の欠測する確率が当該変数の観測された値及び他の観測されている変数の値には依存するが、当該変数の欠測となった値には依存しない場合である

ランダムでない欠測(MNAR):

- ・変数の欠測する確率が当該変数自体の値に依存する場合である

(参考2) 多重代入法について

平均値代入や回帰代入など単一の値を代入する補完法の場合、欠測値に平均的な値が代入されることにより、標本分散については過小に評価される可能性があるが、これに対し、補完に関わる不確実性を考慮し、補完による精度の評価を可能とする方法として多重代入法がある。これは、一定の統計的方法により欠測値への補完値を複数発生させ、それらを代入することにより複数の補完されたデータセットを作成し、それらを基に平均や分散などを推定する方法である。

参考文献

阿部貴行（2016）「欠測データの統計解析」朝倉書店

岩崎学（2015）「統計的因果推論」朝倉書店

高井啓二・星野崇宏・野間久史（2016）「欠測データの統計科学」岩波書店

高橋将宜・渡辺美智子（2017）「欠測データ処理」共立出版

高橋将宜、阿部穂日、野呂竜夫（2015）「公的統計における欠測値補定の研究：多重代入法と単一代入法」、製表技術参考資料

土屋隆裕（2009）「概説 統計調査法」朝倉書店

内閣府経済社会総合研究所（2017）「欠測値補完に関する調査研究報告書」

星野崇宏（2009）「観察データの統計科学－因果推論・選択バイアス・データ融合」岩波書店

Handbook on Methodology of Modern Business Statistics. (2017), European Commission, Eurostat.

(https://ec.europa.eu/eurostat/cros/content/handbook-methodology-modern-business-statistics_en)

Kalton, G and Kasprzyk, D. (1986), The Treatment of Missing Survey Data, *Survey Methodology* Vol.12, No.1, pp.1-16

Rebecca R.Andridge and Roderick J. A. Little. (2010), A Review of Hot Deck Imputation for Survey Non-response, *International Statistical Review* 78, pp.40-64.

Roderick J. A. Little, and Donald B. Rubin. (2020), *Statistical Analysis with Missing Data*. Third edition, John Wiley & Sons.

欠測値補完に係る主な方法等参考資料

1 欠測値補完に係る主な方法

< 項目の欠測 (Item nonresponse) への対応 >

- (1) 平均値代入 (mean imputations)
- (2) 回帰代入 (regression imputation)
- (3) 比率補完 (ratio imputation)
- (4) 最近隣ホットデック (nearest neighbor imputation)
- (5) LOCF (Last observation Carried Forward) 法のうち伸び率補完

< 客体単位での欠測 (Unit nonresponse) への対応 >

- (6) ウェイティング

2 補完を行う層区分の検討事例

3 シミュレーションによる補完方法の検討事例

4 経済センサス調査等他の情報源により判明している結果と比較した補完方法の検討事例

5 各府省の統計作成支援のための業務相談窓口について

1 欠測値補完に係る主な方法

<項目の欠測 (Item nonresponse) への対応>

(1) 平均値代入 (mean imputations)

計算式

平均値補完においては、当該項目についてのすべての回答の平均値を欠測値に代入する。 y_i を欠測が生じている*i*番目のユニットの補完の対象となる項目*y*の値であるとすると、これらのすべての欠測値に代入される値は以下の通り

$$\tilde{y}_i = \bar{y}_{obs} = \frac{\sum_{k \in obs} y_k}{n_{obs}}, \quad (1)$$

obs は、項目*y*についての項目回答者の集合を示している。

平均値補完は、補完の対象となっている項目の値がある程度均質的になる補完クラスを設定し、その平均値を代入することもできる。その場合、式(1)は以下の通り置き換えられる。

$$\tilde{y}_{hi} = \bar{y}_{h,obs} = \frac{\sum_{k \in h \cap obs} y_{hk}}{n_{h,obs}},$$

y_{hi} は補完クラス*h*の中の*i*番目のユニットの値であり、 $n_{h,obs}$ は補完クラス*h*における項目*y*についての項目回答者の数である。この拡張は、時にグループ平均値代入法と呼ばれる。

Handbook on Methodology of Modern Business Statistics(2017;Eurostat); Theme: Model-Based Imputation p.3,4

1

活用事例

統計調査名	府省	全数・標本調査の別	調査周期	活用事例
個人企業経済調査	総務省	標本調査	年次	期首・期末棚卸高について、層化平均値により補完
特定サービス産業実態調査	経済産業省	標本調査	年次	「主たる業務」の年間売上高の業務種類別割合、契約先産業別割合について、前回個票をもとに、規模別等のグループに分け(例:主業事業従事者数、売上高など)し、グループ毎の内訳項目の構成比を算出し、補完を行う個票の該当する規模別等のグループの構成比により補完 ※前回は有る個票の場合は、前回値により算出。
経済産業省生産動態統計調査	経済産業省	一定規模以上全数調査	月次	生産金額の回答があり、数量が無回答の場合、双方の回答がある企業の平均単価を計算し、回答金額を当該平均単価で割り戻して補完

2

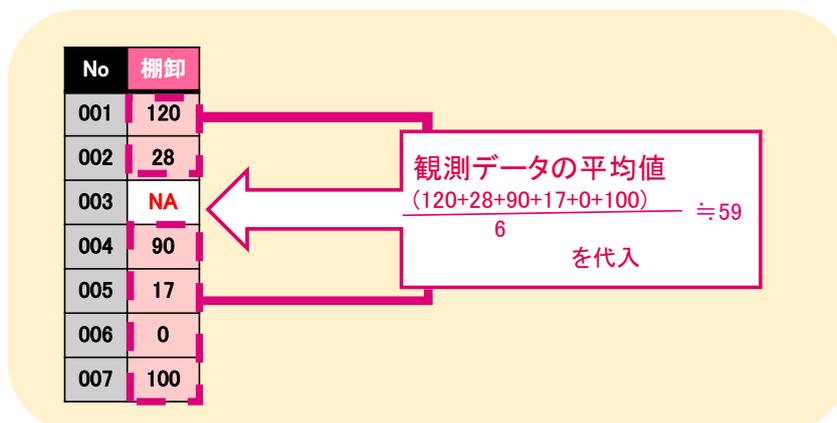
個人企業経済調査における活用事例

平均値補完

・期首棚卸高、期末棚卸高に対しては補完クラスを考慮した上で平均値代入法を用いる。

平均値代入法

観測されているデータの平均値を代入する方法。



第6回評価分科会(令和2年2月19日)

資料5 令和元年度個人企業経済調査～欠測値の補完について～(独立行政法人統計センター技術研究開発課)より

3

(2) 回帰代入 (regression imputation)

計算式

回帰補完は、仮定された回帰モデルにより、全てのユニットについて観測されている項目(補助変数) x_1, \dots, x_q から得た予測値 y を補完代入するもので、平均値補完や比率補完が一般化されたもの。多くの場合では線形回帰モデルが用いられる。

$$y = \alpha + \beta_1 x_1 + \dots + \beta_q x_q + \varepsilon, \quad (3)$$

ここで、 $\alpha, \beta_1, \dots, \beta_q$ は未知のパラメーター、 ε は誤差項である。すべてのユニットの誤差項は独立的に、平均が0で分散が σ^2 である同じ正規分布に従うと仮定されている。

モデル式(3)は項目 y の値と補助変数が双方とも観測されているレコードから通常は最小二乗法により推定される。これにより、項目 y の予測値は補助変数から以下の通り与えられる。

$$\hat{y} = a + b_1 x_1 + \dots + b_q x_q, \quad (4)$$

ここで、 a, b_1, \dots, b_q は最小二乗法により与えられた $\alpha, \beta_1, \dots, \beta_q$ の推定値である。補助変数は全て観測されたものであり、項目 y の予測値は、無回答の項目に対しても、回答者の項目に対しても与えられる。

活用事例

統計調査名	全数・標本調査の別	調査周期	活用事例
サービス産業動向調査 (2012年12月調査まで回帰代入による補完を適用)	標本調査	月次	売上高及び事業従事者数について、1か月目に限り、産業分類別に事業従事者数を説明変数とした対数回帰モデルにより推定した値により補完

サービス産業動向調査における活用事例(2012年12月調査まで回帰代入による補完を適用)

売上高及び事業従事者数についての補完
(1か月目)

産業分類別に事業従事者数を説明変数とした対数回帰モデルにより推定した値を用いる。

$$\log(y_i) = \beta_0 + \beta_1 \times \log(x_i)$$

y_i : 当月の売上高(当月の事業従事者数)

x_i : 母集団事業従事者数

ただし、回帰係数 β_0 , β_1 は、回答があった事業所のうち、 x_i 又は y_i が 0 又はマイナスのものは除外して計算する。

(2か月目以降)

産業分類別の事業従事者規模別に前月からの変化率の平均値を算出し、前月の売上高(事業従事者数)を乗じて推定する。

平成24年サービス産業動向調査年報「付録3 調査対象事業所の抽出方法、結果の推定方法及び推定値の標本誤差」より

5

(3) 比率補完 (ratio imputation)

計算式

統計実務において、比率補完は、補完を行う目的変数 y が、ある単変量の説明変数 x との比がほぼ定数になる場合に使用される。
 r を y と x の比とすると、欠測値 y_i は次のような推定値により補完される。

$$\tilde{y}_i = r x_i$$

一般に、 r は未知なので、 x と y がともに欠測のない観測値により以下のように推定する [De Waal et al.(2011)].

$$r = \frac{\sum_{k \text{ obs}} y_i}{\sum_{k \text{ obs}} x_i}$$

ここで、“obs” は欠測のない観測値のレコードを示す。この補完のモデルを次のように表すと、誤差項 ϵ_i は平均 0 で分散が x に比例する独立でランダムな変数である [Rao(1996)].

$$y_i = r x_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2 x_i)$$

x と y の相関が高いほどこの比推定量の推定効率が高いが、回帰と同様に外れ値に弱いことが知られている。
[e.g. Farrell and Barrera(2007)]. また、それぞれの変数の値を合算して比をとるという性質上、この推定量は特に数値の大きい観測値の影響を受けやすい。

平成28年3月10日 第4回経済センサス活動調査研究会
参考資料 ロバスト比率補完、CART、U検定について より

6

比率補完においては、補完対象の項目 y とある程度比例関係にある、全て観測されている(又は補完されている)一つの項目 x を補助変数とする。最初に、 y と x が両方とも観測されているユニットから、項目 y と x の間の未知の比率 R を推定する。

$$\hat{R} = \frac{\sum_{k \in obs} y_k}{\sum_{k \in obs} x_k} .$$

次に、欠測が生じている i 番目のユニットの項目 y_i に、同じユニットの観測されている項目 x_i にこの比率をかけた値を代入して補完する。

$$\tilde{y}_i = \hat{R}x_i = \frac{\sum_{k \in obs} y_k}{\sum_{k \in obs} x_k} x_i . \quad (2)$$

補完値は、回答者の値から推定された比率が、無回答者の値にも正確に当てはまるということを前提として、導き出されている。

例えば、 y が売上高、 x が従業者数だとする。比率 R は従業者当たりの平均売上高を示す。式(2)により、 i 番目のユニットについて観測されている従業者数に、従業者当たり平均売上高の推定値をかけることにより得られた i 番目のユニットの売上高推定値を導きだし、この推定値を補完に使用する。

活用事例

統計調査名	府省	全数・標本調査の別	調査周期	活用事例
経済センサス活動調査(平成28年度)	総務省	全数	5年	主要な経理項目(売上(収入)金額、費用総額及び給与総額)について、比率補完の手法に基づき、24年活動調査において回答された経理項目の層区分(産業分類、経営組織、国内常用雇用者数)ごとの集計値の比率(売上(収入)金額と費用総額、費用総額と給与総額の比率)を値の得られた経理項目に乗じるにより補完。層区分については「回帰木(CART: Classification And Regression Tree)」の手法を活用して設定し、比率値については「繰返し加重最小二乗法(IRLS: Iterative Reweighted Least Squares)」の手法を活用して算出
法人企業統計調査	財務省	標本調査	年次・四半期	全部無回答の全ての調査項目(計数項目)について、無回答法人の資本金前後10社の平均調査項目対資本金比率に資本金を乗じて算出した値を補完

法人企業統計調査における活用事例

未回答前後10社法人の資本金前後10社の平均調査項目対資本金比率に資本金を乗じて算出
 欠測企業の業種・規模に応じた欠測値が補完される設計となっている。

※全ての計数項目が補完の対象

例: 情報通信業(業種コード60)の設備投資

未提出法人(資本金順)				提出法人(資本金順)					
提出法人	規模区分	業種	資本金(億円)	提出法人	規模区分	業種	資本金(億円)	設備投資(億円)	設備投資/資本金
A	9	60	1,350	a	9	60	9,000	38	0.004
				b	9	60	3,200	600	0.188
				c	9	60	3,100	610	0.197
				d	9	60	2,100	300	0.143
				e	9	60	2,000	550	0.275
				f	9	60	1,400	240	0.171
				g(基準法人)	9	60	1,050	90	0.086
				h	9	60	230	22	0.096
				j	9	60	200	62	0.310
				k	9	60	140	0.2	0.001
				m	9	60	100	0	0.000
				o	9	60	72	8	0.111
				p	9	60	71	1	0.014
				q	9	60	71	2	0.029

1. 未提出法人の資本金上位から提出法人の資本金順リストを降順に検索、未提出法人の資本金以下となる基準位置法人を決める。

2. 基準位置法人の前後10社(上位4社、下位5社)の調査項目(例では設備投資)対資本金比率を算出。

10社の比率の単純平均
0.139 A

3. 未回答法人の資本金額の資本金に近い前後10社の調査項目対資本金比率(単純平均)に未提出の資本金を乗じて、補完値を作成。

例 A社のケース

$$1350 \times (0.197 + 0.143 + 0.275 + 0.171 + 0.086 + 0.096 + 0.310 + 0.001 + 0.000 + 0.111) / 10 = 187.65 \text{ 億円 (補完値)}$$

第5回評価分科会(令和2年2月19日)資料4 法人企業統計調査の欠測値補完について(財務省財務総合政策研究所調査統計部)より

9

(4) 最近隣ホットデック(nearest neighbor imputation)

計算式

最近隣ホットデックにおいては、補助変数は、補完対象の*i*番目のユニットとドナー候補である*k*番目のユニットの間の距離関数を定義するのに用いられる。*i*番目のユニットの最近隣ユニットは、距離関数が最小となる回答者ユニット*d*と定義される。式としては、

$$d = \arg \min_{k \in obs} D(i, k), \quad (2)$$

ここで *obs* は項目 *y* が観測されているユニットの集合を示している、すなわちドナー候補の集合である。

補完方法の説明に入る前に、上記の式(2)における距離関数の選択肢について簡単に説明しておく。補助変数 (x_{j1}, \dots, x_{jq}) は全て量的変数であると仮定すると、良く用いられるのは、以下の式により与えられる距離関数である。

$$D_z(i, k) = \left(\sum_{j=1}^q |x_{ji} - x_{jk}|^z \right)^{1/z} \quad (3)$$

10

ここで z は0より大きい値である。 $z=2$ であるとき、式(3)は良く知られているユークリッド距離を与える式となる。 $z=1$ であるときは、絶対値 $|x_{ji}-x_{jk}|$ の合計となる、これはしばしば、「シティブロック距離」、又は「マンハッタン距離」と称される。 z がさらに大きくなると、式(3)は、個々の補助変数の差分の大きさにより大きなペナルティが与えられることになる。実際のところ、式(3)において z を無限大に近づけると、以下の距離関数が得られる。

$$D_{\infty}(i, k) = \max_{j=1, \dots, q} |x_{ji} - x_{jk}|. \quad (4)$$

式(4)による距離関数を適用すると、最近隣のドナーは、全ての補助変数について、補完対象から大きく離れてはならないことになる。実用的には、距離関数式(3)式において $z=1$ 、 $z=2$ 、又は z を無限大に近づけるといった以外の選択肢が用いられることは希である。

式(3)をさらに一般化した式は、それぞれの補助変数が持つ「正確な補完値を探し出すために必要な重要性を表すウェイト γ_j 」を加えた以下の式で与えられる。

$$D_{z, \gamma}(i, k) = \left(\sum_{j=1}^q \gamma_j |x_{ji} - x_{jk}|^z \right)^{1/z}. \quad (5)$$

11

補助変数を異なる尺度で測ったとき、(3)式又は(5)式に対する補助変数の寄与率に隠れたウェイト付けがされてしまうことに留意すべきである。例えば、 x_1 が昨年のヨーロッパにおける売上高を示し、 x_2 が従業者数であるとする、マンハッタン距離 $D_1(i, k) = |x_{1i} - x_{1k}| + |x_{2i} - x_{2k}|$ は、ほとんど絶対的に前者に依存する。こうしたことを避けるためには、それぞれの補助変数の分散が等しく1になるように補助変数を最初に標準化すべきである。あるいは、変数間の相関を考慮に入れたマハラノビス距離と呼ばれる距離関数が使われることもある(例えば、Little and Rubin, 2002参照)、この距離はユークリッド距離 $D_2(i, k)$ を一般化したものと見なすこともできる。

Handbook on Methodology of Modern Business Statistics(2017;Eurostat); Theme; Donor Imputation p.4,5

12

統計調査名	府省	全数・標本調査の別	調査周期	活用事例
個人企業経済調査	総務省	標本調査	年次	仕入金額、経費計、経費計のうち給料賃金について、同一調査年の他のユニットの数値により最近隣ホットデック法により補完

個人企業経済調査における活用事例

最近隣ホットデック補完

・仕入金額、経費計、給料賃金に対しては補完クラスを考慮した上で最近隣ホットデック法を用いる。

最近隣ホットデック法
性質の近いデータをドナーとして選び、そのドナーの値を代入する方法。

Np.003企業に最も近い企業を探す

今回は欠測変数以外の多変数の線形関係も考慮し、マハラノビス距離を用いる。

$$D(x_i) := \sqrt{(x_i - \hat{x})^T V^{-1} (x_i - \hat{x})}$$

(i = 1, ..., n)

(補完対象項目のうち欠測していない変数を用いる。)

No	売上	仕入	経費	給料
001	5000	150	3050	0
002	6325	273	4797	1238
003	4173	225	2759	NA
004	2500	135	882	0
005	985	0	358	0
006	6610	970	3628	900
007	3908	175	1323	0

Np.007の給料の値を代入

(5) LOCF (Last observation Carried Forward) 法のうち伸び率補完

計算式

補完対象である項目 y が一つの補助変数 x と強い相関関係に有り、例えば $y = Rx$ の様に、係数 R が項目 y と x の関係を表していることを考えてみよう。縦断的データにおいては、このような関係が最も一般的に考えられるのは x とその同じ項目の過去データ y の対応である、ある時点 t における観測データがその前の $t-1$ 時点に観測されたデータと比例関係にあると仮定することは合理的である。過去の値の現時点 t のデータへの更新には、回答者のデータにおいて観測された伸び率が用いられる。補完値がこのような形式で決定されることとなると、項目 y の値は同じ項目の前年の同じ月（もしくは同じ四半期）と比例関係になる、従って、この選択肢は、前年 $t-12$ の値又は四半期前 $t-4$ の過去値を参照していることにな観測値を増やしていき（これは時間が経緯していった場合の例である）。従って、欠測値は、前回の、回答者ユニットの調査時点 t 時点とその前の $t-1$ 時点の間の観測された値の比率と同じ比率を、欠測する直前の観測された値にかけることにより推定ができることになる。

言うなれば、 j 番目のユニットの過去 $t-1$ 時点の項目 y の値が、調査時点 t 時点の同じユニットの項目 y_{jt} の補完値の情報として使用され、一定の比率 R が二つの値の経時的関係値として使用されるということである。比率 R は未知の数字であるが、 t 時点及びその前の時点 $t-1$ どちらにおいても観測されているユニットのみから推定される。

$$\tilde{y}_{it} = \tilde{R}_t y_{it-1} = \frac{\sum_{j \in obs} y_{jt}}{\sum_{j \in obs} y_{jt-1}} y_{it-1} \quad (5)$$

15

y_{jt} は j 番目の回答者の t 時点において観測された項目 y の値、 obs は観測された回答者のサンプルである。前者の式によると、一定の比率 R は、 t 時点及びその前の $t-1$ 時点双方において回答のあったユニットを用いて計算された t 時点及び $t-1$ 時点それぞれの平均値の間の比率と同じである。

Handbook on Methodology of Modern Business Statistics(2017;Eurostat); Theme: Imputation for Longitudinal Data p.7

16

活用事例

統計調査名	府省	全数・標本調査の別	調査周期	活用事例
個人企業経済調査	総務省	標本調査	年次	売上金額について、同一ユニットの過去データを時点調整(回答を得られているユニットの過去からの変化率を乗じる)した数値により補完
商業動態統計調査	経産省	標本調査	月次	全部無回答者の、商品販売額、販売先別商品販売額、商品別手持額について、前月及び当月ともに回答のあった事業所の集計値合計の前月比伸び率を当月無回答者の前月回答値に乗じて補完

17

個人企業経済調査における活用事例

時点調整済LOCF法

売上金額に対しては**時点調整済Last Observation Carried Forward(LOCF)法**によって補完を行う。

LOCF法

同一企業の前回の調査データを据え置きする方法。
(時点調整:過去からの変化率(比率)を加味した調整。)

調査データ

No	売上
001	5000
002	6325
003	NA
004	2500
005	985
006	6610
007	3908

← 過去値 × 比率 を代入

< 比率算出式 >

$$\frac{\text{調査データの売上金額の合計}}{\text{欠測を除く事業所の事業所母集団DBの売上金額の合計}}$$

母集団名簿

No	売上
001	5300
002	6705
003	4131
004	2650
005	1044
006	7007
007	4142

第6回評価分科会(令和2年2月19日)

資料5 令和元年度個人企業経済調査～欠測値の補完について～(独立行政法人統計センター技術研究開発課)より

18

(6) ウェイティング

計算式

標本デザイン上サンプルには既に重み付けがされている(ウェイトが付与されている)が、回答率が全ての層において決定された後に、新たな重み付け(ウェイト)は標本設計上のウェイトと回答率の逆数の産物として定められる。新たな重み付け(ウェイト)は、推定値を求めるために、回答のあった標本単位の重み付けを行う(ウェイティング)際に用いられる。

簡潔な説明のために以下の通り仮定しよう、標本デザインは層化されており、サブグループ(又は事後層)は標本デザイン上の層と一致している、層*h*における回答率は以下の通り評価される。

$$r_h = \frac{n_{rh}}{n_h} .$$

次に、層の中の番目のユニットにおける標本デザイン上の重み付け(ウェイト) $d_{hi} = \frac{N_h}{n_{rh}}$,

は新たな重み付け(ウェイト) $w_{hi} = \frac{d_{hi}}{r_{hi}} = \frac{N_h}{n_{rh}}$ に置き換えられる、一般的にHT推定量は以下の通り与えられる。

$$\hat{Y}_{HT} = \frac{1}{N} \sum_{h=1}^H \frac{N_h}{n_{rh}} \sum_{i=1}^{n_{rh}} y_i = \sum_{h=1}^H \frac{N_h}{N} \left(\frac{1}{n_{rh}} \sum_{i=1}^{n_{rh}} y_i \right) = \sum_{h=1}^H \frac{N_h}{N} \bar{y}_{rh} .$$

【事後層化によるウェイトの調整】

回答確率を求めるサブグループ(クラス)ごとに、補助変数(X)に関する母集団の総計が別途把握できている場合には、クラスを事後層とみなして、このXに関する情報を利用することでウェイトを更に調整することもできる。

この場合、項目(Y)に関する母集団総計の推定量は、

$$\sum_{d=1}^D \sum_{sr,d} w_i \times 1/\phi_i \times \left[\text{補助変数(X)に関するクラス(d)での母集団総計} / \sum_{sr,d} w_i \times 1/\phi_i \times X_i \right] \times Y_i$$

※ w_i は抽出ウェイト、 ϕ_i は客体での回答確率を表す

※ $\sum_{sr,d}$ は、クラス(d)での回収された客体についての和

(なお、 $X_i=1$ とする場合は、母集団のサイズに関する情報を利用することになる。)

また、二つ以上の層化変数で個別に事後層を構成し、いずれの事後層についても、補助変数に関する推定値が母集団の値への一致を図るようにウェイトを調整する方法としてレイキング(raking)がある。これは、回答確率で調整したウェイトについて、更に事後層における層化変数ごとに、補助変数に関する比率(補助変数の母集団総計/補助変数の母集団総計の推定値)を交互に繰り返し乗じて調整する方法である。

参考: レイキングのイメージ

		事後層e				
		e1	e2	e3	合計	真の母集団サイズ
事後層d	d1	\tilde{N}_{11}	\tilde{N}_{12}	\tilde{N}_{13}	\tilde{N}_{d1}	N_{d1}
	d2	\tilde{N}_{21}	\tilde{N}_{22}	\tilde{N}_{23}	\tilde{N}_{d2}	N_{d2}
合計		\tilde{N}_{e1}	\tilde{N}_{e2}	\tilde{N}_{e3}		

真の母集団サイズ	N_{e1}	N_{e2}	N_{e3}
----------	----------	----------	----------

回答確率で調整したウェイトを $w_{i(0)} (= w_i / p_i)$ とし、ここでは二つの事後層化のための変数(dにより2つの層、eにより3つの層に分かれるとする)による場合とする。

$\tilde{N}_{11} \sim \tilde{N}_{23}$ は $w_{i(0)}$ を用いたサイズの推定値を表わす。

例えば、 $d1 \times e1$ 層に属している客体(i)については、以下のようにウェイトを調整する。

(1) 第1の層化変数(d)による事後層を用いて $w_{i(0)}$ を調整し、 $w_{i(1)}$ とする。

$$w_{i(1)} = w_{i(0)} \times g_{i(d(0))}, \quad g_{i(d(0))} = N_{d1} / \tilde{N}_{d1}, \quad \tilde{N}_{d1} \text{は } w_{i(0)} \text{により推計。}$$

(2) 第2の層化変数(e)による事後層を用いて $w_{i(1)}$ を調整し、 $w_{i(2)}$ とする。

$$w_{i(2)} = w_{i(1)} \times g_{i(e(1))}, \quad g_{i(e(1))} = N_{e1} / \tilde{N}_{e1}, \quad \tilde{N}_{e1} \text{は } w_{i(1)} \text{により推計。}$$

(3) 再び第1の層化変数(d)による事後層を用いて $w_{i(2)}$ を調整し、 $w_{i(3)}$ とする。

$$w_{i(3)} = w_{i(2)} \times g_{i(d(2))}, \quad g_{i(d(2))} = N_{d1} / \tilde{N}_{d1}, \quad \tilde{N}_{d1} \text{は } w_{i(2)} \text{により推計。}$$

(4) 調整後のウェイトの変化がなくなるまで、二つの事後層を交互に用いて調整を繰り返す。最終的に得られるウェイト $w_{i(c)}$ は、

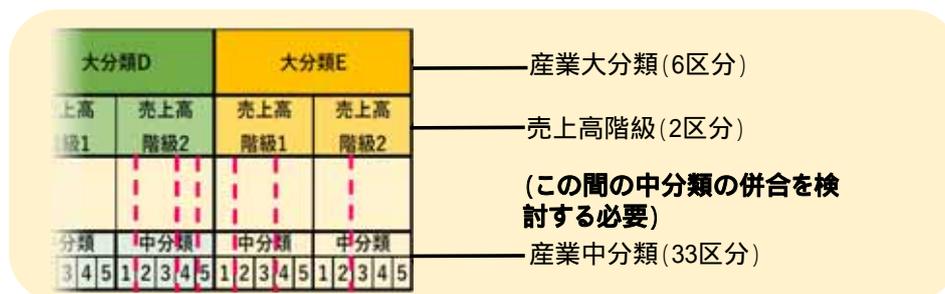
$$w_{i(c)} = w_{i(0)} \times (g_{i(d(0))} \times g_{i(d(2))} \times g_{i(d(4))} \dots) \times (g_{i(e(1))} \times g_{i(e(3))} \times g_{i(e(5))} \dots) \text{ となる。}$$

土屋 隆裕 (2009) 概説 標本調査法を参考とした 21

2 補完を行う層区分の検討事例

令和元年個人企業経済調査の補完クラスについて

- ・ **補完クラス**とは、欠測値の補完処理の際に参照するクラスである。できる限り欠測データと類似した補完クラスを決定する必要がある。
- ・ 補完クラスを「売上高階級(2区分)」×「産業中分類(33区分)」としたいが、補完クラスのメンバー数が極端に少ない場合、特徴の似た中分類同士の併合を検討する。



第6回評価分科会(令和2年2月19日)

資料5 令和元年度個人企業経済調査～欠測値の補完について～(独立行政法人統計センター技術研究開発課)より

令和元年度個人企業経済調査の欠測値の補完を検討した際の、補完クラス(層区分)の考え方

第9回個人企業経済統計研究会(平成29年10月7日) 資料3 令和元年個人企業経済調査～欠測値の補完について～ より

- ・ 無回答により起こる結果数値のバイアスを減らし、推定精度を上げるために設定する
- ・ クラスは全ての調査単位について欠測のない変数により、クラス内の補完対象ができるだけ均質で、クラス間の差異は大きくなるように設定する
- ・ クラスを構築する変数は、回答・無回答の傾向に関係するものを使用する

例) 大企業と中小企業で、回答する企業の割合が異なる場合
補完クラスを規模に関する変数、例えば従業員数等を用いて設定し、クラス内の回答割合をできるだけ均一にする

23

3 シミュレーションによる補完方法の検討事例

令和元年個人企業経済調査の期首・期末棚卸の欠測値に対する補完に関して、観察されている項目の一部を欠測させるシミュレーションを行い、NRMSEを用いて比率ホットデックと平均値補完を比較。

第9回個人企業経済統計研究会(平成30年10月7日) 資料3令和元年個人企業経済調査～欠測値の補完について～
第6回評価分科会(令和2年2月19日) 資料5令和元年個人企業経済調査～欠測値の補完について～ より

シミュレーション方法*

H14～29個人企業経済調査構造編データを産業大分類、売上高90%点別で補完クラスを設定し、ランダムに一定割合(20%)を欠測とみなし、欠測パターン別に補完する

データサイズ

産業大分類	集計対象企業のデータサイズ		20%欠測とした場合	
	売上高階級		売上高階級	
	90%以上	90%未満	90%以上	90%未満
E 製造業	1079	9944	215	1988
I 卸売業、小売業	1922	17631	384	3526
M 宿泊業、飲食サービス業	1092	9815	218	1963
Q サービス業	1402	12400	280	2480

24

結果の評価方法

標準平均平方誤差 NRMSE*
(Normalized Root Mean Square Error)

$$\text{NRMSE} = \sqrt{\frac{\sum \left(\frac{x^{true} - x^{imp}}{\sigma} \right)^2}{n}}$$

x^{true} : 真値
 x^{imp} : 補完値
 σ : x^{true} の標準偏差

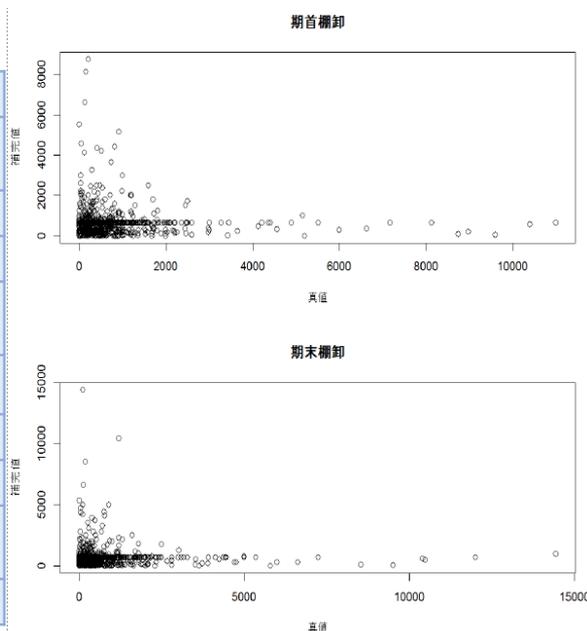
欠測値補完の精度の評価指標の一つで、値が1に近づけば、補完値と真値の差異の分散が真値の分散に近く、0に近づくほど補完値の誤差が小さい。RMSE(二乗平均平方根誤差)を標準化した指標。真値と補完値の乖離について、標準化することにより、データの元々の単位にかかわらず比較が可能。

[07期首棚卸高]と[09期末棚卸高]が欠測の場合

期末棚卸高は期首棚卸高との相関が高いが、他の項目との相関関係が低いため、両者が欠測すると誤差が大きくなる傾向がみられ、概ね平均値補完の方が良い推定となる

補完結果《NRMSE*》

産業	売上高	比率ホットデック		平均値	
		期首棚卸高	期末棚卸高	期首棚卸高	期末棚卸高
E 製造業	90%以上	1.48	1.61	1.00	1.00
I 卸売業、小売業		1.19	1.27	1.00	1.00
M 宿泊業、飲食サービス業		1.06	1.12	1.00	1.00
Q サービス業		1.31	1.39	1.00	1.00
E 製造業	90%未満	1.09	1.12	1.00	1.00
I 卸売業、小売業		0.92	0.98	1.00	1.00
M 宿泊業、飲食サービス業		0.95	1.04	1.00	1.00
Q サービス業		1.22	1.31	1.00	1.00



特定サービス産業実態調査についての補完方法の検討

第6回評価分科会(令和2年2月19日)

資料6 平成21年度サービス産業統計等経済産業統計の基盤整備事業「特定サービス産業実態調査等における推計手法の確立に関する調査研究」についてより

- ・ 特定サービス産業実態調査について、4つの推計方法について比較検証。
a) 横置き補完 b) 平均値補完 c) 比補完 d) 平均値補完 + 比補完
- ・ 検討に際しては、母集団の全ての事業所・企業について、主要な調査項目を上記の4つの方法で補完し、真値との乖離(誤差)を評価。
- ・ 具体的には、各調査の直近2年分の調査結果を使用し、パネル化できた事業所・企業の集団を擬似的に「母集団」と見なして評価。

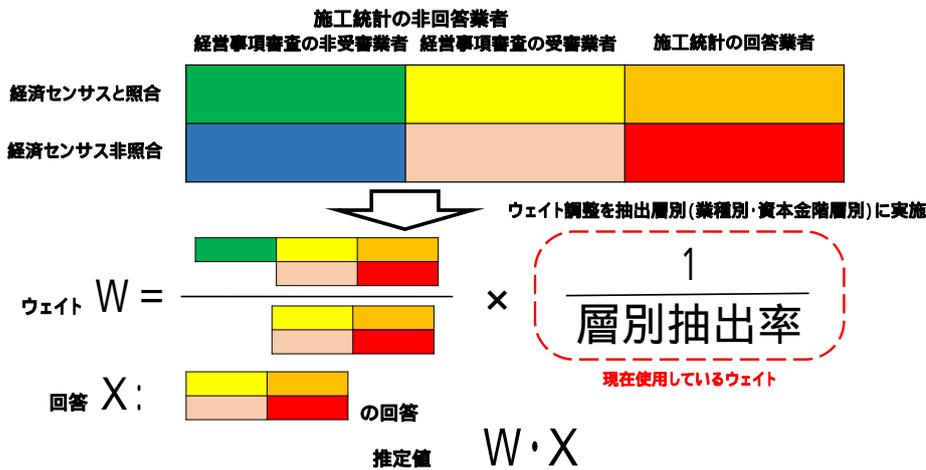


- ・ 平均値補完を基本として、事業従事者規模500人以上の層について、比補完によって総和を推定した場合、業種によっては平均値補完のみを用いる場合と比較して真値に近い推定が可能となり、より精度の高い推定が期待出来るが、その差はわずか。
- ・ ただし、「500人以上」の層に該当する事業所(企業)が少ない等の業種では、比補完を行う場合の方が、ばらつきが大きくなる場合がある。

4 経済センサス調査等他の情報源により判明している結果と比較した補完方法の検討事例

建設工事施工統計調査における、平成28年経済センサスー活動調査等の情報を用いたウェイトの調整による未回答への対応について、平成28年経済センサスー活動調査の結果との比較により検証

<「ウェイト調整法」による補完>
 建設工事施工統計の非回答業者かつ経営事項審査の非受審業者のうち、経済センサスと照合できた者(緑色部分)が施工統計の回答業者及び経営事項審査の受審業者(橙色部分 + 赤色部分 + 黄色部分 + 桃色部分)と同じ分布であると仮定し、抽出層別にウェイト調整を行う。



平成28年経済センサス-活動調査及び経営事項審査結果を活用した欠測値補完の試行結果は、平成28年経済センサス-活動調査の結果とほぼ合致。



建設工事施工統計調査の 部分の結果 (公表結果)		建設工事施工統計調査の 部分の結果 (補完試行結果)			経済センサス-活動調査の 部分の結果 (A表集計結果)		
	施工統計 現行 A	施工統計 現行 B	施工統計 補完 C	C/B	センサス 現行 D	B/D	C/D
集計対象サンプルサイズ	67,047	45,280	45,280	100%	-	-	-
事業者数	217,566	153,244	191,488	125%	175,350	87%	109%
従業者数(千人)	2,788	1,929	2,212	115%	2,348	82%	94%
完工高(10億円)	88,247	70,427	77,355	110%	78,225	90%	99%
元請完工高	56,413	48,310	51,990	108%	53,547	90%	97%
土木工事	14,598	14,915	15,713	105%	16,077	93%	98%
建設工事・建設設備工事(住宅)	15,540	11,952	13,469	113%	10,371	115%	130%
新設	12,290	9,890	10,990	111%	-	-	-
維持・修繕	3,250	2,056	2,474	120%	-	-	-
建築工事・建築設備工事(非住宅)	19,805	18,755	19,904	106%	19,681	95%	101%
新設	13,512	13,403	14,092	105%	-	-	-
維持・修繕	6,293	5,352	5,812	109%	-	-	-
機械装置等工事	6,467	2,687	2,904	108%	4,016	67%	72%
下請完工高	31,833	22,118	25,364	115%	24,489	90%	104%
土木工事	9,212	7,738	8,591	111%	8,754	88%	98%
建設工事・建設設備工事(住宅)	9,598	3,446	4,208	122%	3,063	113%	137%
建築工事・建築設備工事(非住宅)	13,294	9,206	10,607	115%	9,070	101%	117%
機械装置等工事	3,368	1,728	1,959	113%	2,909	59%	67%

建設工事施工統計調査の従業者数は、建設業の従業者数 + 建設業以外の部門の常雇数。

5 各府省の統計作成支援のための業務相談窓口について (令和元年8月～)

業務

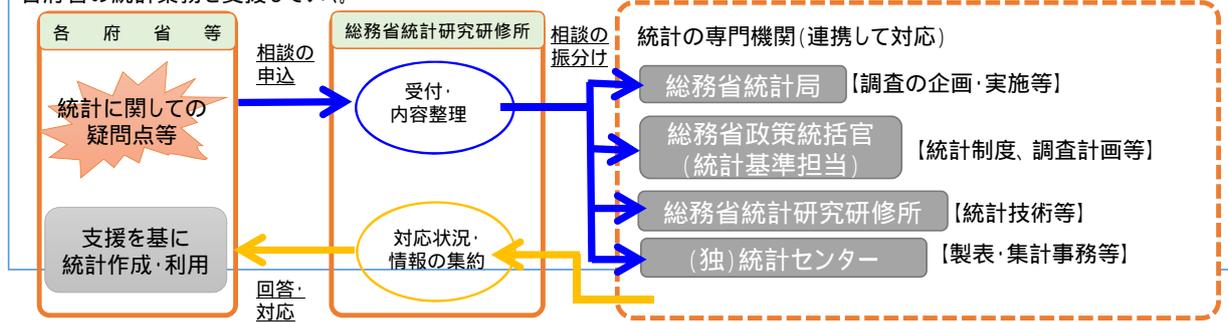
統計の作成や利用に対する各府省からの相談を受け、支援を行う。

背景・必要性

一連の統計不正問題を受けて、不適切事案の再発防止や政府統計の品質向上が喫緊の課題。このためには、**統計担当者の統計リテラシーの向上が不可欠**であり、統計専門機関の一つである総務省統計研究研修所が中心となって、**各府省等に対して統計作成・利用の支援を行い、正確な統計の作成、統計の品質向上を推進していくことが必要。**

相談業務の流れ

総務省統計研究研修所は、総務省統計局、総務省政策統括官(統計基準担当)及び(独)統計センターと連携し、各府省の統計業務を支援していく。



回答の集約

(参考文献)

Handbook on Methodology of Modern Business Statistics (2017;Eurostat)
https://ec.europa.eu/eurostat/cros/content/handbook-methodology-modern-business-statistics_en

旧横断的課題検討部会
第6回統計の精度向上及び推計方法改善ワーキンググループ会合(平成30年2月14日)
資料1-4 欠測値及び外れ値検査 調査別検査結果
https://www.soumu.go.jp/main_content/000533099.pdf

第4回経済センサス 活動調査研究会(平成28年)
参考 ロバスト比率補定、CART、U検定について
<http://www.stat.go.jp/info/kenkyu/e-census/katsuken/pdf/kk040201.pdf>

第8回個人企業経済統計研究会(平成30年度第1回)
資料2 平成31年個人企業経済調査 ~ 経理項目の補完に向けて ~
<https://www.stat.go.jp/info/kenkyu/kojinke/09/pdf/shiryo2.pdf>

第9回個人企業経済統計研究会(令和元年度第1回)
資料3 令和元年個人企業経済調査 ~ 欠測値の補完について ~
<https://www.stat.go.jp/info/kenkyu/kojinke/10/pdf/shiryo3.pdf>

第5回評価分科会(令和2年1月)
資料4 法人企業統計調査の欠測値補完について
https://www.soumu.go.jp/main_content/000666127.pdf

資料10 各府省の統計作成支援のための業務相談窓口について
https://www.soumu.go.jp/main_content/000666134.pdf

31

第6回評価分科会(令和2年2月)
資料5 令和元年度個人企業経済調査 ~ 欠測値の補完について ~
https://www.soumu.go.jp/main_content/000670702.pdf

資料6 平成21年度サービス産業統計等経済産業統計の基盤整備事業「特定サービス産業実態調査等における推計手法の確立に関する調査研究」について
https://www.soumu.go.jp/main_content/000670703.pdf

参考1 「特定サービス産業実態調査等における推計手法の確立に関する調査研究」調査報告書
https://www.soumu.go.jp/main_content/000670995.pdf

第8回評価分科会(令和2年10月)
資料3 建設工事施工統計調査における欠測値補完の見直しについて(案)[修正版]
https://www.soumu.go.jp/main_content/000715085.pdf

平成28年経済センサス 活動調査
欠測値等の取扱いについて
(統計局ホームページ掲載)
<https://www.stat.go.jp/data/e-census/2016/kekka/pdf/hotei.pdf>

「個人企業経済調査 未回答項目の補完について」(PDF:143KB)
統計局ホームページ掲載
<https://www.stat.go.jp/data/kojinke/pdf/hokan.pdf>

平成24年サービス産業動向調査年報 付録3 調査対象事業所の抽出方法、結果の推定方法及び推定値の標本誤差
<https://www.stat.go.jp/data/mssi/report/2012/pdf/ap03.pdf>

土屋隆裕(2009)概説標本調査法.朝倉書店

De Waal, T., Pannekoek, J., Scholtus, S. (2011) *Handbook on Statistical Data Editing and Imputation*, Wiley handbooks in survey methodology. Hoboken, New Jersey:John Wiley & Sons.

Rao, J. N. K. (1996) On Variance Estimation With Imputed Survey Data, *Journal of the American Statistical Association*, Vol.91, pp.499-506.

Farrell, P. J. and Saliban-Berrera, M. (2006) A Comparison of Several Robust Estimators for a Finite Population Mean, *Journal of Statistical Studies*, Vol.26,pp.29-43. 32