

# 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 non-response) 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  $h$  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。

## 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 ※前回値が有る個票の場合は、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

# 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。

No	棚卸
001	120
002	28
003	NA
004	90
005	17
006	0
007	100

観測データの平均値

$$\frac{(120 + 28 + 90 + 17 + 0 + 100)}{6} \approx 59$$

を代入

## (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、 $\varepsilon$ 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 $\sigma^2$ 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  $\beta_0$ 、 $\beta_1$  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。

### (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。

$$\tilde{y}_i = rx_i \quad (1)$$

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)]。

$$\hat{r} = \frac{\sum_{k \in \text{obs}} y_i}{\sum_{k \in \text{obs}} x_i} \quad (2)$$

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  $\epsilon_i$  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)]。

$$y_i = rx_i + \epsilon_i, \quad \epsilon_i \sim N(0, \sigma^2 x_i) \quad (3)$$

$x$  と  $y$  の相関が高いほどこの比推定量の推定効率は高いが、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。

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、 $y$ が売上げ高、 $x$ が従業員数だとする。比率 $R$ は従業員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員数に、従業員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	<p>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ことにより補完。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</p>
法人企業統計調査	財務省	標本調査	年次・四半期	<p>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</p>

# 法人企業統計調査における活用事例

未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  
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	規模区分	業種	資本金(億円)
A	9	60	1,350

提出法人(資本金順)

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	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	70	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)に未提出法人の資本金を乗じて、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

## (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 ( $x_1, \dots, x_q$ ) 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。

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

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、「ミニマックス」と称される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 $\gamma_j$ 」を加えた以下の式で与えられる。

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

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

## 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	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スク法により補完

# 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。

No.003企業に最も近い  
企業を探す

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。

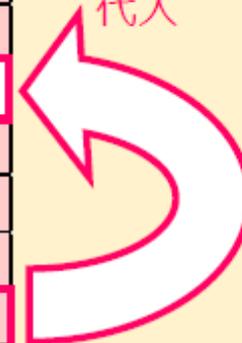
$$D(\mathbf{x}_i) := \sqrt{(\mathbf{x}_i - \hat{\mathbf{x}})^T V^{-1} (\mathbf{x}_i - \hat{\mathbf{x}})}$$

(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

No.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、 $i$  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  $y_{it}$  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。

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

$y_{jt}$ 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

## 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

# 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

母集団名簿

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

過去値×比率※を代入

※<比率算出式>

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

## (7) 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 $i$ 番目のユニット $i$ における標本デザイン上の重み付け（ウェイト） $d_{hi} = \frac{N_h}{n_h}$ ,

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。

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

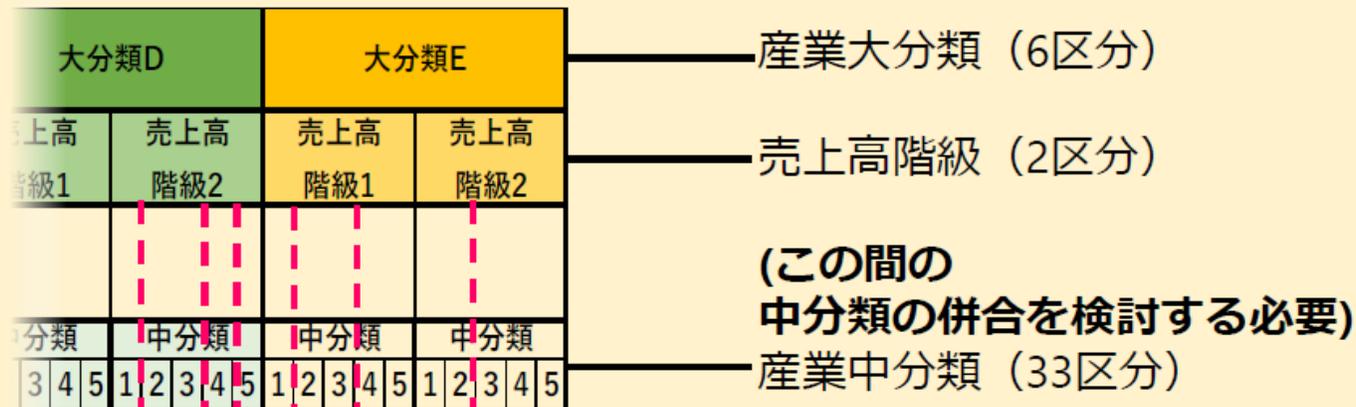
## 2 補完を行う層区分の検討事例

### ○令和元年個人企業経済調査の補完クラスについて

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

- **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を併合を検討する。



# ○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

### 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

# 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  
 $\sigma$  :  $x^{true}$  の標準偏差

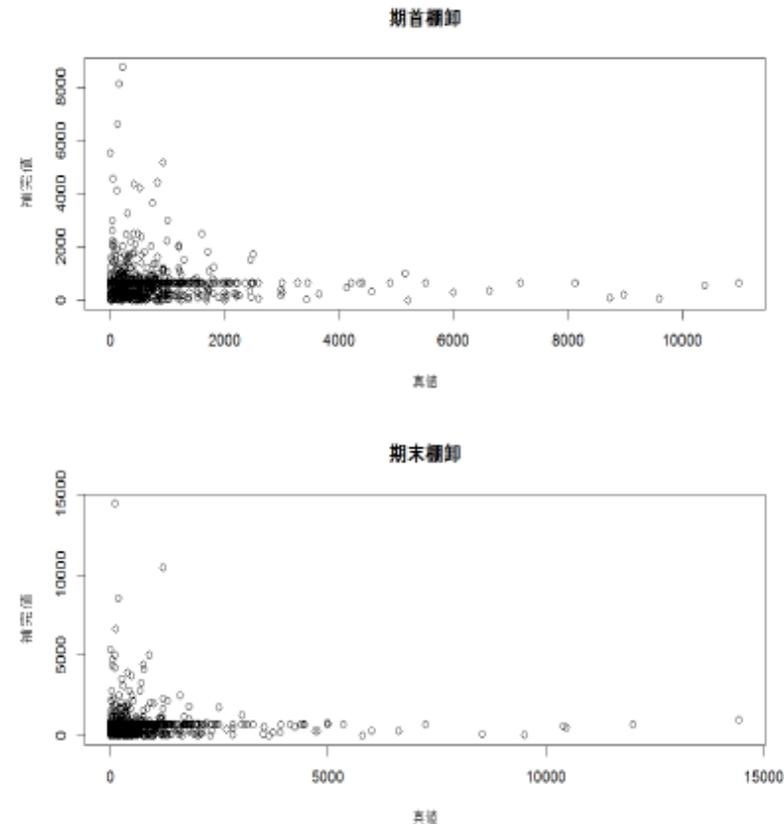
欠測値補完の精度の評価指標の一つで、値が1に近づけば、補完値と真値の差異の分散が真値の分散に近く、0に近づくほど補完値の誤差が小さい。RMSE(二乗平均平方根誤差)を標準化した指標。真値と補完値の乖離について、標準化することにより、データの元々の単位にかかわらず比較が可能。

# [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、**おおむ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	90%以上	1.48	1.61	1.00	1.00
I		1.19	1.27	1.00	1.00
M		1.06	1.12	1.00	1.00
Q		1.31	1.39	1.00	1.00
E	90%未満	1.09	1.12	1.00	1.00
I		0.92	0.98	1.00	1.00
M		0.95	1.04	1.01	1.01
Q		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 「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。



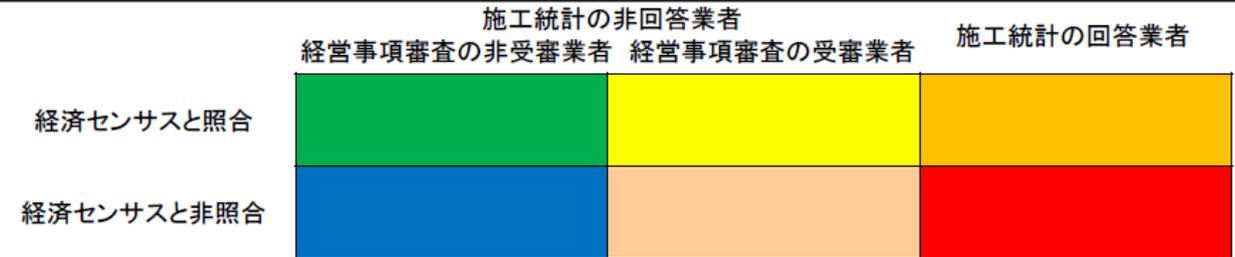
- ・ 平均値補完を基本として、事業従事者規模500人以上の層について、比補完によって総和を推定した場合、業種によっては平均値補完のみを用いる場合と比較して真値に近い推定が可能となり、より精度の高い推定が期待出来るが、その差はわずか。
- ・ ただし、「500人以上」の層に該当する事業所（企業）が少ない等の業種では、比補完を行う場合の方が、ばらつきが大きくなる場合がある。

# 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

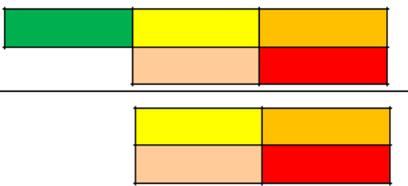
○建設工事施工統計調査における、平成28年経済センサス一活動調査等の情報を用いたウェイトの調整による未回答への対応について、平成28年経済センサス一活動調査の結果との比較により検証

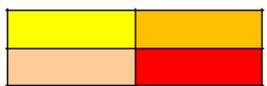
第8回評価分科会（令和2年10月30日）資料3 建設工事施工統計調査における欠測値補完の見直しについて（案）より

<「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、抽出層別にウェイト調整を行う。



↓ ウェイト調整を抽出層別(業種別・資本金階層別)に実施

ウェイト  $W =$    $\times$  1  
層別抽出率

回答  $X :$   の回答

現在使用しているウェイト

推定値  $\sum W \cdot X$

○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 現行 B	施工統計 補完 C	C/B
45,280	45,280	100%
153,244	191,488	125%
1,929	2,212	115%
70,427	77,355	110%
48,310	51,990	108%
14,915	15,713	105%
11,952	13,469	113%
9,890	10,990	111%
2,056	2,474	120%
18,755	19,904	106%
13,403	14,092	105%
5,352	5,812	109%
2,687	2,904	108%
22,118	25,364	115%
7,738	8,591	111%
3,446	4,208	122%
9,206	10,607	115%
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
-	-	-
175,350	87%	109%
2,348	82%	94%
78,225	90%	99%
53,547	90%	97%
16,077	93%	98%
10,371	115%	130%
-	-	-
-	-	-
19,681	95%	101%
-	-	-
-	-	-
4,016	67%	72%
24,489	90%	104%
8,754	88%	98%
3,063	113%	137%
9,070	101%	117%
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。

# (参照文献)

○Handbook on Methodology of Modern Business Statistics (2017;Eurostat)

[https://ec.europa.eu/eurostat/cros/content/handbook-methodology-modern-business-statistics\\_en](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](https://www.soumu.go.jp/main_content/000533099.pdf)

○第5回評価分科会

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](https://www.soumu.go.jp/main_content/000666127.pdf)

○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](https://www.soumu.go.jp/main_content/000670702.pdf)

○個人企業経済統計研究会（第9回（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/shiryu3.pdf>

○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](https://www.soumu.go.jp/main_content/000670702.pdf)

資料6 「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](https://www.soumu.go.jp/main_content/000670703.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](https://www.soumu.go.jp/main_content/000715085.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>

○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>