



経済統計における欠測値補完の 評価指標に関する検討

2026年度統計関連学会連合大会
令和8年9月8日（火）

独立行政法人統計センター
村田 一郎

- 公的統計と欠測値補完
- 補完方法を評価する指標の検討
- 補完シミュレーションの実践

公的統計と欠測値補完

公的統計を作成するための統計調査の実施においては、

- ・ 回答期限を過ぎた調査客体への督促
- ・ 調査票記入内容の確認

などを含む調査段階の対応によって、
調査票への記入及び回収の確保が図られている。

しかし、やむを得ず回答が得られなかった場合（＝欠測）は、
結果の有用性を確保するため、集計段階で統計的な補完を
適切に行うなどの対応が考えられる。

- ・総務省 個人企業経済調査の仕入金額等

同一調査の中で補完対象と性質の近い他の事業所の数値により補完

→最近隣ホットデスク補完

- ・経済産業省 商業動態統計調査の商品販売額等

前月及び当月ともに回答のあった事業所のみ集計した合計値の前月比伸び率を補完対象の前月回答値に乗じて補完

→伸び率補完

- ・財務省 法人企業統計調査の計数項目

補完対象の資本金の値と近い前後10社の平均比率（計数項目対資本金）に補完対象の資本金を乗じて算出した値を補完

→比率補完

補完方法を評価する指標の検討

補完方法の検討においては、過去のデータ等を用いたシミュレーションにより適用時の影響を評価することがよく行われている。

シミュレーションの実施例

個人企業経済調査（総務省）における検討（令和元年）

→回答が得られている項目の一部を欠測させて、その値を補完した場合の回答値と補完値の乖離を誤差として評価（誤差二乗和など）

（評価指標の例）

二乗平均平方根誤差

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i^{imp} - x_i^{obs})^2}$$

n : データ数

標準化二乗平均平方根誤差

$$NRMSE = \frac{RMSE}{\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i^{obs} - \bar{x}^{obs})^2}}$$

x^{imp} : 補完値

x^{obs} : 回答値

平均絶対誤差

$$MAE = \frac{1}{n} \sum_{i=1}^n |x_i^{imp} - x_i^{obs}|$$

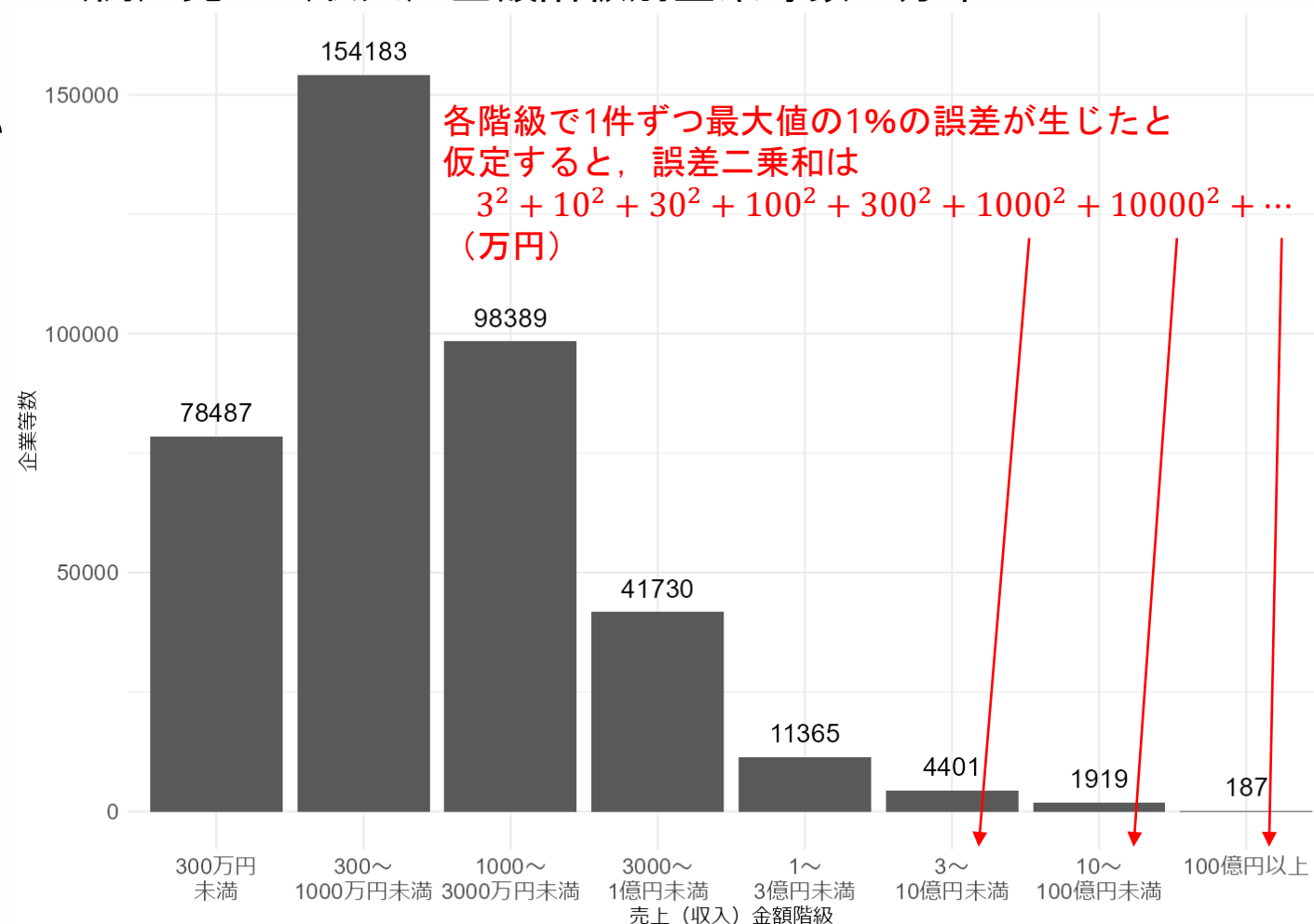
誤差二乗和（ESS）を用いる場合の注意

- ・ 経済統計の調査項目は、
右に裾の長い分布であることが多い

- ・ 分布の裾の方における誤差が、
全体に対して大きな影響を与える

- ・ 分布の左の方に改善があっても
見えにくくなってしまう

（例）売上（収入）金額階級別企業等数の分布



（令和3年経済センサス - 活動調査
企業等に関する集計 産業横断的集計 表8-1から作成）

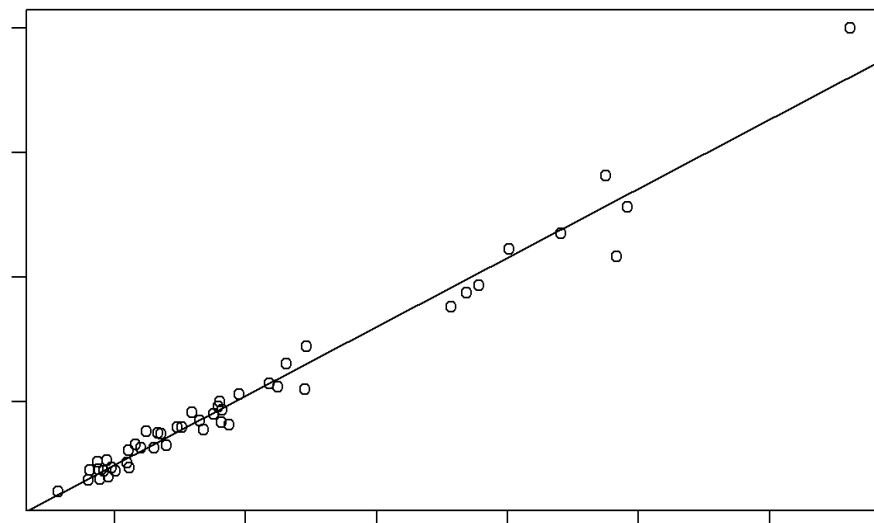
誤差二乗和の考え方・・・個々の点と推定モデルの差を計上

→推定すべきものは点である

一方、公的統計において点（マイクロデータ）は公表の対象ではなく、それらの統計的性質を正しく捉えることが重要

→点で当てる必要はなく、
合計・平均などの統計量の推定を重視

（線形推定モデルのイメージ）



（例）合計が同じでも差の二乗和は積み重なる

データ	回答値	シミュレーション補完値	差分	差の二乗
データ1	100	150	+50	2500
データ2	120	90	-30	900
データ3	240	220	-20	400
計	460	460	0	3800

Andridge et al. (2021) が用いた主な評価指標

- ・ 補完シミュレーションの集計値 Y^* と実調査の補完後集計値 $Y_{disseminated}$ の比較

$$|Y^* - Y_{disseminated}|$$

$$\text{ただし, } Y^* = \sum x_{observed} + \sum x_{imputed}, \quad Y_{disseminated} = \sum x_{processed}$$

- ・ 補完シミュレーションを繰り返したときの集計値 Y^* の分散

$$Variance = \frac{1}{p-1} \sum_i (Y_i^* - \bar{Y}^*)^2, \quad p = \text{繰り返し数}$$

Andridge, R., Bechtel, L. T. and Thompson, K. J. (2021).
"Finding a Flexible Hot-Deck Imputation Method for Multinomial Data".
Journal of Survey Statistics and Methodology, 9(4), 789-809.

補完シミュレーションを繰り返し行ったときに、補完後集計値が変動する場合

- ・ 補完手法自体が確率的である
(例：ランダムホットデック，確率的回帰代入)
- ・ シミュレーションごとに欠測をランダムに仮定する
- ・ シミュレーションごとにブートストラップ法でデータを再構築する

補完後集計値の分散は，その変動要因によって集計値がどのくらいの影響を受けるかを評価する指標となる

補完シミュレーションの実践

調査の目的

我が国の全ての産業の付加価値等の構造を明らかにし、
国民経済計算の精度向上等に資すること

調査周期

毎年（経済センサス - 活動調査の実施年を除く）

調査の対象

- ・ 全ての産業に属する一定規模以上の法人企業（産業横断調査）
- ・ 製造業に属する一定規模以上の法人事業所（製造業事業所調査）

主な調査項目

各産業の売上高や費用内訳等

今回のシミュレーション対象

総務省統計局のWebサイトの内容に基づき作成

製造業事業所の「製造品出荷額計」を補完するシミュレーションを行った。

※ 製造品出荷額・・・その事業所が製造したものの工場出荷金額など

補完手法：前年からの伸び率による補完

各事業所の前年の「製造品出荷額計」×伸び率 ＝ 各事業所の今年の補完値

評価方法

- ・元の集計値との差の割合を，伸び率の作成方法を変えて比較
- ・補完後の集計値の変動係数を，伸び率の作成方法を変えて比較

補完シミュレーションでは次の2年分のデータを用いた。

- ・2022年経済構造実態調査（製造業事業所調査）
- ・2021年（令和3年）経済センサス - 活動調査の製造業以降，「今年」は2022年，「前年」は2021年を指す

製造業事業所を区分別に集計して、前年からの伸び率を作成

$$\text{区分別伸び率} = \frac{\text{今年の区分別集計値}}{\text{前年の区分別集計値}}$$

区分は次の3とおりを使用

- ・ 産業大分類～細分類（製造業に属する726分類）
- ・ 上の産業×従業者規模（1～29, 30～99, 100人以上の3区分）
- ・ 上の産業×都道府県

次の2つの集計値を用いた伸び率を作成

- ・ 製造品出荷額等合計（加工賃とその他の収入も含めた合計）
- ・ 製造品出荷額計のみ

※この伸び率算出に用いる事業所は次の条件を満たすものに限る

- ・ 前年から産業分類が変わっていない
- ・ 事業所単体の前年からの伸び率が0.5～2.0の間

シミュレーションでは、前年・今年ともに製造品出荷額計が得られているデータを用いる。

（前年か今年のいずれかにだけ存在する事業所はあらかじめ除く）

このデータに対して、今年の製造品出荷額計をランダムに欠測とする。
欠測割合は約40%とした

※ 産業細分類ごとにこの割合で欠測とした

手順

- 1 欠測とする事業所をランダムに選ぶ
- 2 欠測としなかった事業所のみで伸び率を作成
- 3 欠測とした事業所の製造品出荷額計を補完
- 4 それら全ての事業所の製造品出荷額計を産業細分類別に集計

上の1～4を300回繰り返す、集計値の平均と変動係数を算出

元の集計値との差の割合 (%)

$$= \frac{|\text{補完シミュレーション後の集計値の平均} - \text{元の集計値}|}{\text{元の集計値}} \times 100$$

- ・ 産業によりスケールが大きく異なるため、差そのものではなく割合を使用

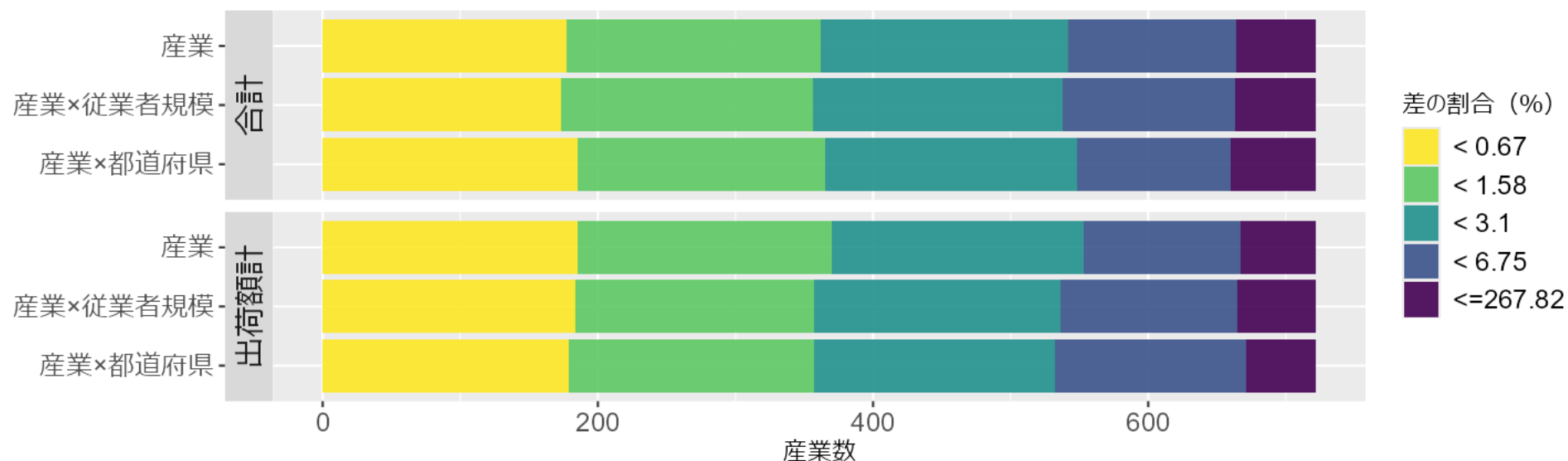
補完シミュレーション後集計値の変動係数

$$= \frac{\text{補完後集計値の標準偏差}}{\text{補完後集計値の平均}}$$

- ・ 産業によりスケールが大きく異なるため、分散ではなく変動係数を使用

全726産業の集計値を個別に評価することは避け、
産業別の差の割合を相対的に比較

※ 図の色分けは図中の全割合を四分位範囲で便宜的に分けたもの

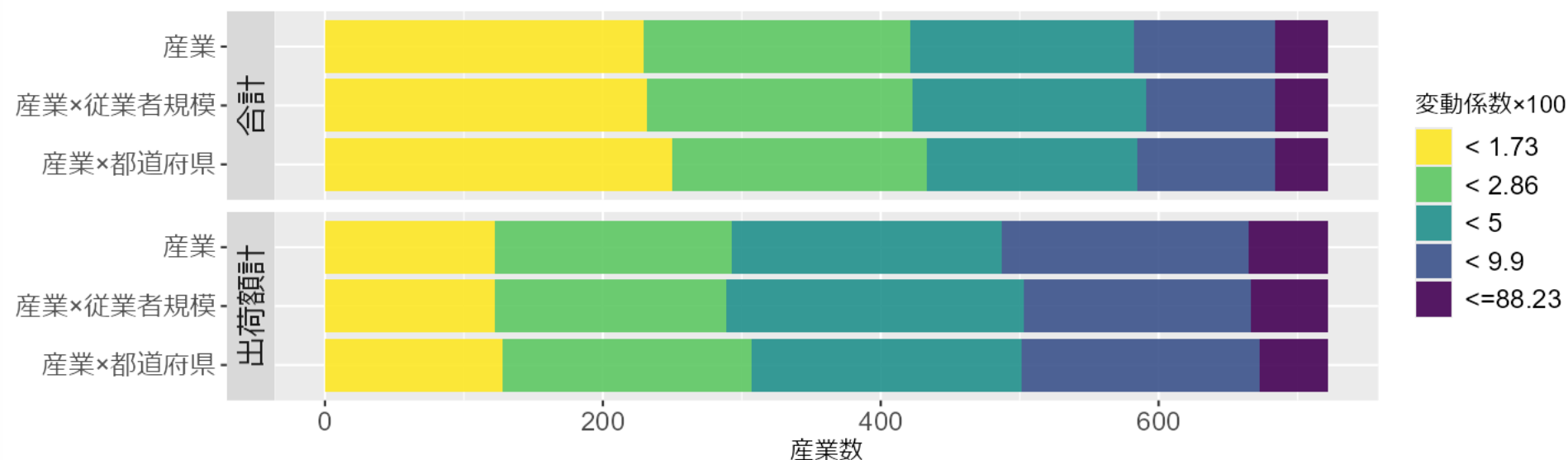


マイクロデータを基に発表者が集計・作成

どの伸び率を用いた場合も大きな差はないが、
合計よりも製造品出荷額計のみから伸び率を産業別に算出した場合の
差の割合が比較的小さい

同様に，変動係数を相対的に比較

※ 図の色分けは図中の全変動係数を四分位範囲で便宜的に分けたもの



ミクロデータを基に発表者が集計・作成

合計から伸び率を算出した方が変動係数が小さい

さらに，産業×都道府県別に伸び率を算出した方が変動係数が小さい

経済構造実態調査（製造業事業所調査）の製造品出荷額計について、
前年からの伸び率による補完をシミュレーションした

補完の結果について

- ・元の集計値との差の割合
- ・補完後の集計値の変動係数

を見て、伸び率の作成方法による違いを評価した

ミクロデータ単位の補完の誤差の大小ではなく、
補完の全体の特徴を捉えた評価を行うことができた

欠測値補完方法を評価する際に、「真値からの誤差」を見ることは必ずしも実用的でない場合がある

公的統計では、合計や平均などのデータ全体の統計的性質を正しく推定することが重要である

集計値単位で見たときの差や、分散などを評価することで、データ全体としての妥当性や、欠測の不確かさに対する補完方法の安定性などを検討することが可能となる