

外れ値の取扱い方法について

「全国企業短期経済観測調査（短観）」では、売上高等の計数項目を集計する際、業種別・規模別の区分により、統計的手法である「母集団推計」を実施している。2010年12月調査より、統計精度を一段と高める観点から、「標本データの一部が外れ値となった場合における母集団推計値作成上の取扱い」を適用することとした。以下では、外れ値の具体的な取扱いを「検出」と「処理」に分けて説明する。¹

1. 外れ値の検出方法

- (1) 集計区分である6区分（製造業・非製造業×大企業・中堅企業・中小企業）ベースの前年比・修正率に対する各企業の回答値の影響度合い y_{it} を以下の通り定義して、それぞれ算出する。

$$y_{it} = \frac{\text{重み付けされたデータの変化幅}}{\text{前年度(前回)の6区分ベースの母集団推計値}} \times 100$$

$$= \frac{w'_i(x_{it} - x_{it-1})}{|M_{jt-1}|} \times 100$$

- x_{it} : 第*i*番目の企業の時期*t*における回答値
 x_{it-1} : 第*i*番目の企業の時期*t-1*における回答値
 w'_i : x_i にかかる調整後ウェイト（注）
 M_{jt-1} : 時期*t-1*における*j*区分（ $j=1, \dots, 6$ ）の母集団推計値

（注） w'_i は、*t*期および*t-1*期の標本企業数の差異によらず「重み付けされたデータの変化幅」を正確に算出するために、*t*期および*t-1*期の標本企業数を揃えたベースでのウェイト（2時点で何れかが欠測値の場合は、その企業が属する母集団推計層の平均値を代入することで標本企業数を揃える）。

- (2) 次に、 y_{it} を基に外れ度合いの指標 z_{it} を以下の通り定義して、それぞれ算出する。

$$z_{it} = \begin{cases} \frac{y_{it} - d_j^{99}}{D_j} & \text{if } d_j^{99} \leq y_{it} \\ \frac{d_j^1 - y_{it}}{D_j} & \text{if } y_{it} \leq d_j^1 \\ 0 & \text{if } d_j^1 < y_{it} < d_j^{99} \end{cases}$$

外れ度合いの指標 z_{it} が閾値（ $C=50$ ）を超えた場合、当該回答値を外れ値として検出する。閾値は過去データを用いた実証分析を踏まえて設定。

d_j^1 : j 区分の y_{it} の 1 パーセンタイル²
 d_j^{99} : j 区分の y_{it} の 99 パーセンタイル
 D_j : d_j^{99} と d_j^1 の距離 ($d_j^{99} - d_j^1$)

—— 外れ値対応は「売上高」、「経常利益」、「当期純利益」、「設備投資額」、「ソフトウェア投資額」の 5 項目を対象に実施。また、売上高と設備投資額の内訳項目である「輸出」と「土地投資額」については、売上高や設備投資額が外れ値として検出された場合のみ、外れ値対応の処理を行う。

2. 外れ値の処理方法

検出された外れ値は母集団を代表していないデータであり、母集団推計に用いることは不相当であるため、当該項目は欠測値として処理する。欠測値に対しては、現行ルールと同じ手法（当該項目の直近回答値を代入 $\langle x_{it} = x_{it-1} \rangle$ ）を適用し、外れ値の影響を除外する。なお、一旦外れ値となった項目については、次回の標本洗い替えまで当処理を継続する。

以 上

¹ 外れ値対応に関する専門的な論文としては、例えば「[ビジネスサーベイにおける外れ値対応—全国企業短期経済観測調査（短観）のケース—](#)」（2010年7月）を参照。

² 小さい方から順に並べて1パーセント番目にある値のこと。