# 乗率の補正による欠測補完に係る指標について


統計数理研究所データ科学研究系教授

土屋　隆裕


## 1　乗率の補正による欠測補完

欠測 (nonresponse) には大きく調査項目単位での欠測 (item nonresponse) と調査対象単位での欠測 (unit nonresponse) とがある。調査項目単位での欠測への対処には一般に補定 (imputation) が用いられるが、調査対象単位での欠測に対しては、補定の他にも乗率の補正がなされることも多い。

乗率の補正は主に次のような三段階を経て行われる (Valliant et al., 2013)。まず、標本抽出デザインを反映した、調査対象$i$の元の乗率（抽出ウェイト）を $v_i$ とすると、調査対象 $i$ の回答確率（response propensity）$\rho_i$ あるいはその推定量 $\hat{\rho}_i$ の逆数を乗じて$v_i/\rho_i$あるいは$v_i/\hat{\rho}_i$ とする。次に母集団の情報をベンチマークとして用いたキャリブレーションを行って、$v_i/\hat{\rho}_i g_i$（$g_i$ はキャリブレーションのためのウェイト）とするのである。

本稿では、この乗率補正に関連した三つの指標を取り上げる。一つは不等加重効果であり、もう一つは R-指標である。

## 2　不等加重効果

回答確率 $\rho_i$の推定やキャリブレーションウェイト$g_i$ の算出に当たって適切な共変量を用いれば、程度の多寡はあれ、乗率の補正によって推定量の偏りは減少すると期待できる。しかし目的とする変数と乗率との関係によっては、推定量の分散は拡大し、偏りは縮小しても結果として平均三乗誤差はむしろ拡大してしまうおそれもある。例えば単純無作為抽出標本などの自己加重標本で、全ての調査対象に一律の抽出ウェイトが与えられる場合と、補正によって一部の調査対象にのみ大きな乗率が与えられる場合とを比較してみる。元の抽出ウェイトでは全ての調査対象が等しく推定値に寄与することになるが、補正後は一部の調査対象の変数値によって推定値が決まることになる。そのためもし補正後の乗率と目的とする変数とが無相関であれば、推定量の分散はかえって拡大してしまう。

そのような、分散の拡大を表す指標として不等加重効果（Unequal Weighting Effect）がある (Kish, 1965)。

$$UWE = n\frac{\sum \omega_i^2}{(\sum \omega i)^2} = 1 + \frac{n-1}{n} CV(\omega i)^2 \approx 1 + CV(\omega i)^2 \qquad (1)$$

$CV(\omega_i)$ は乗率$\omega_i$ の変動係数であり、乗率$\omega_i$ は必ずしも補正後の乗率に限るものではない。不等加重効果は、目的とする変数と乗率とが無相関のときに、乗率が全て等しい場合と比べて（つまり乗率の変動係数が 0 の場合と比べて）推定量の分散がどの程度拡大してしまうのかを表す指標である。そのため様々な場面で用いることができる。

第一の場面は、上記のような欠測補完のための乗率補正を行った結果、得られた乗率の分散が拡大してしまい、推定量の分散が過大とならないかを確認する場面である。得られた乗率の不等加

重効果を確認することで、乗率に不適当な補正がなされていないかを検討したり、キャリブレーションにおける距離関数の選択やパラメータの調整を行ったりすることができる。

例えば図 1 の左は、乗率のキャリブレーションにおいて距離関数をロジット関数とし、パラメータである上限値を様々に変えたときの $g_i$の分布を示したものである（土屋, 2012)。この箱ヒゲ図だけでは上限値をいくつにすべきか決めにくい。図 1 の右は各上限値に対応して得られた乗率の不等加重効果の値である。不等加重効果を基準とするのであれば、上限値は不等加重効果が最小となる 4.5 とすればよいことが分かる。

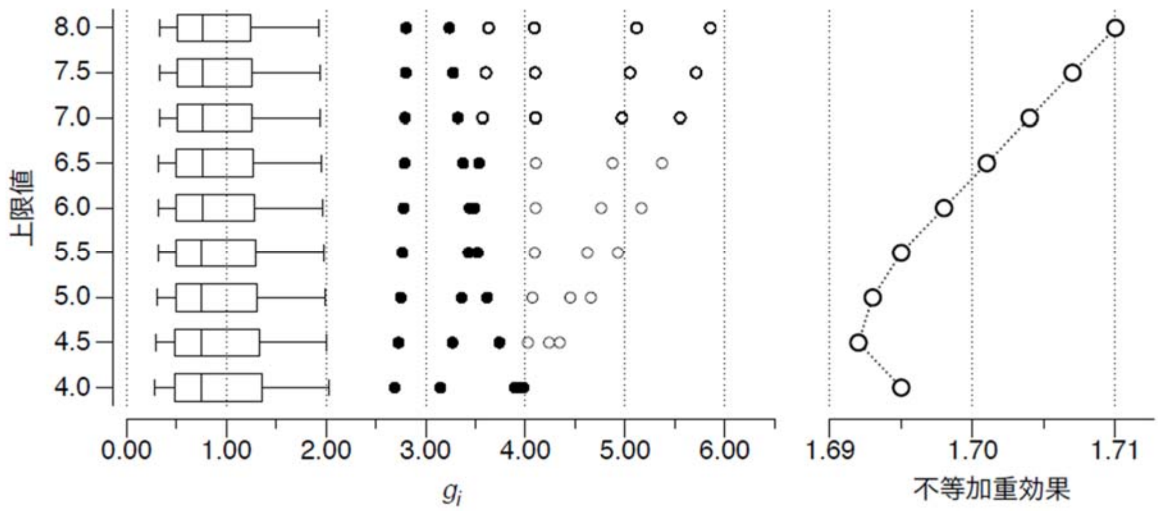


図 1: キャリブレーションにおける上限値パラメータと$g_i$の分布、不等加重効果

第三の場面は、欠測補完という文脈ではないが、適切な標本抽出デザインを設計する場面である。層化抽出法や多段抽出法を用いると、抽出ウェイトは調査対象の間で一定とならないことがある。例えば規模比例確率抽出標本のように、抽出ウェイトと目的とする変数とが反比例していれば、推定量の分散は単純無作為抽出法のときよりも小さくなることが知られている。しかし目的とする変数と抽出ウェイトとが無相関となったり、逆に比例してしまうと、推定量の分散は単純無作為抽出法のときよりも大きくなってしまう。調査では目的とする変数は一般に多数あるため、抽出ウェイトと変数との相関関係は一律ではない。そこで、標本抽出デザイン自体の精度（の 低さ） を表す一つの指標として (1) 式の不等加重効果が用いられることがある。仮に抽出ウェイトと無相関の変数があった場合に、その変数について推定量の分散がどの程度拡大するのかという不等加重効果の値をもって、標本抽出デザインを評価するのである。そのため不等加重効果をデザイン効果と表現する文献もある。

第三は複数の調査データを統合する場面である。近年、調査対象の特性は多様化が進み、同じ目 標母集団の中でも調査対象によって適切な調査モードは異なることがある。そのため異なるモードで独立に調査したデータを統合し、単一の推定値を得ようとする試みは増えている。例えば従来型の調査員調査や郵送調査に加えてインターネット調査を実施し、両者を統合しようとする場面や、同じ電話調査であっても固定電話に対する調査と携帯電話に対する調査とを統合しようとする場面などである。先述のとおり、不等加重効果はデータの精度（の低さ） を表す指標であるため、その逆数で各調査モードのデータを重みづけて統合したり、あるいは不等加重効果が大きくならないよう、データの統合に当たってのパラメータを定めるといった利用がなされている。

## 3 R-指標

調査データの質を表す指標としては、主に回収率が用いられることが多い。回収率が高いときには欠測（unit nonresponse）はあまり問題とされず、欠測補完が課題となるのは、もちろん回収率が低く多数の欠測が生じた場合であろう。しかしよく知られるように、回収率の低さは必ずしも推定量の偏りの大きさを示すものではない。例えば自己加重標本において母集団平均を標本平均で推定するときの偏りは、目的とする変数 $y$ と回答確率 $\rho$ との共分散を $\sigma_{yp}$ とし、回答確率の平均を $\mu_p$ とすると次式となる。

$$Bias\,(\bar{y}) = \frac{\sigma_{yp}}{\mu_p} \tag{2}$$

回答確率の平均 $\mu_\rho$ あるいはその推定量としての回収率が低くとも、$\sigma_{yp} = 0$ であれば偏りは生じない。

当然 $\sigma_{yp}$ は変数 $y$ に応じて異なるため、(2) 式を調査データ全体の質を表す指標としては用いにくい。しかし回答確率 $\rho_i$ が調査対象の間で一定であれば、どのような変数 $y$ であっても $\sigma_{yp}$ は 0 に近いであろう。つまり回答確率が一定であれば、回収標本のサイズは小さくなり、推定量の分散は拡大するものの、標本の代表性は保たれているために、欠測補完の必要性は薄れると考えてよい。

そのような、調査データの代表性あるいは質を表す指標の一つとして R-指標（R-indicator）がある（Schouten et al., 2009)。

$$R = 1 - 2\sqrt{\frac{1}{N-1}\sum_{i \in U}\left(\rho_i - \mu_\rho\right)^2} \tag{3}$$

(3) 式は母集団 $U$ における R-指標であり、標本 $S$ によるその推定量は次式となる。

$$\hat{R} = 1 - 2\sqrt{\frac{1}{N-1}\sum_{i \in S} v_i\left(\hat{\rho}_i - \hat{\mu}_\rho\right)^2} \tag{4}$$

R-指標は 0 と 1 の間の値をとり、値が 1 に近いほど回答確率 $\rho_i$ の分散は小さく、回収標本の代表性は保たれていると言える。推定量の偏りを表す (2) 式の上限は、R-指標 $R$ を用いると

$$|Bias(y)| \leq \frac{1-R}{2\mu_\rho}\sigma_y^2 \tag{5}$$

と表すことができ、R-指標が 1 に近ければ、(5) 式の分母にある $\mu_\rho$ あるいはその推定量としての回収率が低くとも、偏りの上限は低く抑えられるからである。そして回答確率 $\rho_i$ の分散が小さいと、$v_i/\rho_i$ によって乗率の補正を行っても、乗率の変動係数の大きさは補正前後であまり変わらず、不等加重効果を拡大することにはつながらない。

### 4　二つの指標の実際

欠測の補完を行うに当たっては、結局は欠測が存在しない真の状態が分からないために、いくつかの指標を用いながら複数の観点で補完の適切さ、あるいは欠測補完をして得られた推定値の適切さを評価していくことが必要である。本稿では、そのときに役立つであろう三つの指標を紹介した。

しかしこれら三つの指標を現実の欠測補完の際に利用しようとするとき最も悩ましいのは、回答確率$\rho_i$　が知られておらず、推定しなければならないという点である。回答確率 $\rho_i$ を精度高く推定しようとしたり、回答確率によって乗率を補正し、数多くの目的変数のいずれに関しても偏りを小さくしたりするには、一般に適切な共変量をより多く用いるのがよいと考えられる。しかしより多くの共変量を用いるほど、得られた $\hat{\rho}_i$の分散は一般に大きくなり、補正後の乗率の不等加 重効果を拡大したり、R-指標の値を引き下げてしまうことにつながる。したがって共変量の選択 基準として、本稿で紹介した指標を用いるのは難しい。利用する共変量を固定した上で、乗率の補正方法を比較したり、複数の調査データの質を比較するときなどに利用するのが適切であろう。

## 文献

Kish, L. (1965). *Survey Sampling*, John Wiley & Sons, New York.

Schouten, B., Cobben, F. and Bethlehem, J. (2009). Indicators for the representativeness of survey response. *Survey Methodology*, **35**, 101–113.

土屋　隆裕 (2012). 明治末期における小学生の理想人物調査―キャリブレーション手法の比較― 『統計　数理』, **60**, 219–234.

Valliant, R., Dever, J.A. and Kreuter, F. (2013). *Practical Tools for Designing and Weighting Survey Samples*, Springer, New York.