ウェイティング：WHY, WHEN, and HOW？

Leslie Kish

この文章は次の文献の全訳です：Kish, L. (1990) Weighting: Why, when, and how? *Proceedings of the Survey Research Methods Section, American Statistic Association*, pp.121-130.
原文は次の URL からダウンロードできます：http://www.amstat.org/sections/srms/Proceedings/ .

著者の L. Kish (1910-2000)は高名な統計学者で，その著書"Survey Sampling"(1965)は調査データ解析の古典的教科書として現在でも広く用いられています。Kish はこの短い文章のなかで，調査データのウェイティング(重み付け)について広範でわかりやすい解説をおこなっています。そのためこの文章は，正規の学術論文ではないにも関わらず，頻繁に引用される基礎文献となっています。

もともと原文が平易なのですが，翻訳にあたっては専門的な用語や言い回しを極力避け，さらに平易な訳文になるように心がけました。そのため，正確さを欠いている箇所があるかと思います。また，適宜改行を追加し，(おそらくは講演時のスライドとして用いられた) 冗長な図表，ならびに引用文献リストを割愛しました。また，理解を助けるため注記をつけましたので，参考にしていただければ幸いです。

S.ONO.
2008.1.10
修正 2008.3.11

1.　はじめに

調査データの分析についてのもっともよくある疑問は，ウェイティングについての基本的な疑問でしょう(**表** 1)。私はほとんど毎週のようにこの疑問に遭遇します。しかし，「ウェイティングの過程についてのわかりやすい初心者向け説明は存在しません」(Sharot, 1986) 。調査の実務家は単純なガイダンスを求めているのに，彼らが容易に利用できる便利な説明は見当たらないのです。この論文が目指しているのは，それらの要求に多少なりとも応えることです。

この問題は，調査についての文献のなかで言及されることはあるものの，索引のなかに現れることさえめったにありません。しかし，この問題について 1987 年以降 6 本の論文が発表されていることからみても，この問題への関心は次第に高まっているようです。

多くの調査レポートのなかの分析において，ウェイティングは頻繁に用いられています。レポートのなかでウェイティングについて説明がなされることもありますが，それらは場当たり的なものですし，付録のなかに埋もれているのがふつうです。いっぽう，ウェイティングについて理論的に議論している論文もありますが，それらは層化，事後層化，無回答，分散縮減といった，

表 1. ウェイティングについてのよくある疑問

1. **なぜ**標本データをウェイティング**しなければならないの**か(WHY)？
2. 標本データにウェイティングが**必要**なのは**どのような場合か**(WHEN)？
3. 標本データをウェイティングするのが**適切**なのは**どのような場合か**(WHEN)？
4. データをウェイティングするのが**重要**なのは**どのような場合か**(WHEN)？　それが実際に意味を持つのはどのような場合か？
5. 適切で正確なウェイトを**求める**ためには**どうしたらよいか**(HOW)？
6. ケース，テープ，統計表にウェイトを**適用する**ためには**どうしたらよいか**(HOW)？
7. **公式**やソフトウェアのなかでウェイトを適用するには**どうしたらよいか**(HOW)？

表 2 誤解を招く説明：よくある例

a. 調査データのウェイティングは**単純な**プロセスだ：無回答を調整したうえで，それぞれの反応の抽出確率の逆数を，その反応のウェイトにすればよいのだ。
b. 標本抽出の理論からみると，無回答の問題を別にすれば，ウェイティングすることに理論的な根拠は存在しない。
c. ウェイトなしの標本を手に入れたければ，単純無作為抽出によって標本を得るのが一番だ。
d. 標本のケースにウェイトをつけるのは倫理的にまずい。なぜなら，その過程で誤用が生じ，バイアスのかかった結果が得られてしまうことがあるからだ。

ウェイティングの特定の側面だけに焦点をあてています。誤解を招く説明を目にすることもあります(表 2)。たとえば**表 2b.**のような，さまざまなモデルに基づいてウェイティングを批判する理論的な議論です。しかし，私がここで目指しているのは単純で簡単，かつ一般的で便利な説明ですから，これらの論点については割愛せざるを得ません。単純さと一般性というこの 2 つの基準を満たすためには，深遠さや正確さは犠牲にしないといけないでしょう(単純さ，一般性，深遠さという 3 つの基準のすべてを満たそうとする人は，結局どれについても失敗してしまうにちがいない，と私は思います)。私のこの不十分さをどなたかが乗り越えてくださることを期待しています。

2. ウェイトなしの標本

多くの場合最も好まれるのはウェイトなしの標本です。なぜなら，それらが単純さ・分散の小ささ・頑健性において著しく優れているからです。なんであれ統計理論というものは，なんらかのかたちで，ウェイトなしの標本を強く想定しているものです。

ウェイトなしの標本：原文は self-weighting samples。Kish はこのことばを，標本抽出が EPSEM の性質(後述)を持っているかどうかを問わず，とにかくウェイティングをせずに用いる標本を指して用いている。
分散の小ささ：この文章の中で Kish は「分散」ということばを頻繁に用いている。ここでいう分散とは，「もし標本抽出を繰り返したとしたら，同じような値(データないし統計量)が得られるか，それとも抽出のたびに大きく違う値になってしまうか」ということを指している。したがって分散は小さいほうが望ましい。ある標本に含まれている値のちらばり(標本分散)のことではない点に注意。

さらに次の理由があります。標本抽出において，それぞれのケースが標本として選ばれる確率(抽出確率)がどのケースでも等しいことを，EPSEM(equal probability sampling mechanism)といいます。いま，母集団において調べたい変数が幅広く分布している場合について考えて見ましょう。この場合は，EPSEM が望ましくかつ合理的であるとみなされるのが普通です。その例としてまず思い浮かぶのは投票です。選挙ではすべての大人が投票しますね。投票でなくても，行動・態度・意見が民主的に幅広く分布している場面，すくなくとも大まかにいってそうであるような場面は，ほかにたくさんあります。

表 3 にウェイティングを行う主な理由を 4 つ挙げました(1,2,3,4)。上記の状況について考えてみると，まず理由 1 (層・領域の不均等割付)からは逃れることができるでしょう。運なり技術なりがあれば，理由 2(フレームの問題)も迂回できそうです。無回答率が低ければ，理由 3 も不要

表 3. ウェイティングをおこなう理由

A.　抽出確率の逆数をかけたい
B.　選択バイアスを減らしたい
C.　分散を減らしたい
D.　基になる母集団を変えたい

1. **非比例配分**（A）
　層に対する「最適」配分によって分散を減らしている
　小さな領域について標本を大きめにとっている
　出現率が低い要素があるために多段階抽出を行っている
2. **台帳の不均等性**（A）
　抽出単位のサイズにばらつきがある（家族，組織，会社）
　リストに重複がある
3. **無回答**（AB）
　調査そのものへの無回答（回答拒否，不在）
　コールバック率にちがいがある
　項目への無回答
4. **統計的調整**（ABCD）
　事後層化，比推定量，回帰推定量
　レーク推定量，反復適合，リム・ウェイティング，SPREE
　「最適」ウェイト
5. **標本の結合**（D）
　さまざまな母集団
　定期的な調査を蓄積する
6. **コントロール用統計表**（B）
　無作為か，バイアスを被っているか？
　全体か，小さなセルか？
7. **非確率標本**（D）
　チェックデータを用いたコントロール
　クオータ抽出との類似性

でしょう。そして理由 4 (調整)も，たいていは必要とされません。このように，EPSEM の性質を持っているウェイトなしの標本が得られるという場面は, 調査において珍しいものではありません。

なお, ウェイティングの 4 つの理由のうち 1,2 は EPSEM 抽出かどうかに関わるものですが, 3,4 は無関係です。つまり，EPSEM 標本であってもウェイティングを行うことがあります。また 5 章で述べるように, EPSEM 抽出ではないがウェイティングをする必要がないという場合もあります。したがって EPSEM 抽出はウェイトなしの標本のための必要条件でも十分条件でもありません。もっとも，この 2 つは強く関連しており，なかなか区別できないのですが。

ここで一番大事なのは，ウェイトなしの標本による分析には「頑健性」があるという点です。標本を選び出す操作を通じて EPSEM 抽出に到達する技術こそが，標本抽出の精髄です。その技術のひとつが, 抽出確率をサイズと比例ないし反比例させた複雑な多段階抽出です。そのためには, 抽出台帳における不完全性を賢く扱う必要がありますし, 必要な回答率を得るための熟練と配慮も必要になるでしょう。

3.　ウェイティングする理由

さきほど，ウェイティングの主要な理由を 4 つ挙げました。それらに加え，特定の状況にのみ関わるあまり一般的でない理由が 3 つあります。

以下ではこれらの 7 つの理由についてひとつずつ述べますが，わざわざそうする理由は，通常それらが及ぼす効果が非常に異なっているからであり，またそれらに対して必要な取り扱い方も異なっているからです。

理由 1. 母集団における標本の割合が比例的でない

このような標本抽出が熟慮のうえでなされることがあります。たとえば，データの分散ないし費用を減らすため，層に対して「最適」割り当てを行う場合です。また，特定の層(通常は小さな層)について大きめの標本を抽出しておく，という場合もあります。

こうした配慮に効果を持たせるためには，層 h において抽出された標本が母集団において占めている割合 f_h が，層のあいだで 2 倍から 10 倍くらい異なっている必要があります(100 倍にまで至ることはめったにありません)。多くの場合，f_h は基となる抽出率 f の整数倍にしておくのが便利です(2f, 10f など)。このとき，層を合計した統計量を求める際には，バイアスを防ぐため，倍率の逆数(1/2, 1/10 など)をウェイトとすることによって補正する必要があります。

層：ここでは，年齢，性別，地域など，母集団を構成するグループのこと。層ごとに標本を抽出することを層別抽出という。
最適割り当て：それぞれの層の標本サイズは，母集団におけるその層の大きさに比例させて決めることが多いが，それが常に最良であるわけではない。たとえば，極端に小さな層がある場合は，その層の標本が小さくなりすぎてしまい，都合が悪い。また，一定の予算のなかで効率の良い調査をするためには，ある程度大きな層，内部のばらつきが大きい層，データ収集の費用が低い層の標本を大きくしたほうが良い。こうした点を考慮して標本サイズを決めることを最適割り当てという。
標本割合が比例的でない：のちに Kish(1992)は，この項で説明している状況をさらに次の 2 つに分類している。(1)最適割り当てのために，標本の割合が非比例的になった場合。このとき，fh は層のあいだで大きく異なっており，ウェイティングの必要性が高い。(2)小さな層で大きめな標本を抽出したり，層の大きさを無視して同じサイズの標本を抽出したりした場合。このとき，層のあいだでの fh の差は比較的に小さく，ウェイティングの必要性は低い。

理由 2. 抽出台帳に不均等性がある

ただし，標本抽出においてこれを補正していた場合を除きます。

その不均等性が標本の小さな一部分にしか影響していない場合，あるいはその不均等性が小さなものであった場合には，それがもたらす全体的影響を調べたうえで，無視してよいものかどうかを判断します。無視できないものであった場合には，抽出確率に反比例したウェイトによって補正しなければなりません。もっとも，この場合のウェイトはあまり大きな値にはならず，標本の一部にしか適用されないのがふつうです。

抽出台帳に不均等性がある：Kish(1992)はこの状況をさらに 4 つに分類している。(1)小さなクラスタを抽出し，その大きさが異なる場合。たとえば，世帯や建物を無作為抽出し，各世帯・建物につき 1 名を調査対象者にした場合。世帯・建物の大きさが異なるため，各対象者の抽出確率にばらつきが生じる。ウェイティングの必要性は世帯・建物の大きさによって異なる。(2)リストに重複がある場合。ウェイティングの必要性は高い。(3)リストに空白があり，誤って「空白を抽出してしまったらその次の要素を抽出する」ルールを採用してしまった場合。(4)台帳が不完全であった場合。

理由 1,2 によるウェイトは，単に抽出確率 Π_i が等しくないことを補正するものですから，「Π の逆数」と呼ぶことができます。

理由 3. 無回答の補正

これは理由 1,2 とは異なる問題を引き起こします。無回答の補正は，無回答が標本のケースのなかでランダムに生じると想定できない限り，ただの「Π の逆数」では済みません。無回答のウェイトは，暗黙的なものであれ明示的なものであれ，なんらかのモデルないし想定を必要とします。

そうした想定にたどり着くための方法のひとつが，補正変数によってセルをつくり，それぞれのセルに異なるウェイトを与えることです。ここで補正変数とは，地域，年齢，性別といった，利用可能でありかつ調査変数と関連していると期待される変数です。つまりこの方法のためには，この二つの要請の両方をよく満たしている調整変数を，標本なりチェック用統計表なりから入手しなければならないわけです。それはふつう困難で，めったに実現できません。幸い，無回答が少数である場合には，下位クラスの間での無回答率の差も小さくなりやすく，したがってその影響も小さくなります。

無回答の補正：10000 人を無作為抽出して対象者としたが，そのうち 100 人が無回答であったとしよう。このとき，100 人の無回答の正しい扱い方は，その無回答が生じるメカニズムによって変わる。一般に，このメカニズムは 3 種類に分類される。

- **完全ランダム欠損(MCAR)**: ある人が無回答であるかどうかが，調べたい調査項目とは一切無関係に決まる場合，いわば無回答が偶然によって決まる場合。Kish が述べているように，この場合の無回答は特別な扱いを必要としない。つまり，サイズ 9900 の無作為標本を得たものとみなしてかまわない。ただし，現実には MCAR はきわめてまれである。
- **ランダム欠損(MAR)**: ある人が無回答であるかどうかが，調べたい調査項目によって決まるわけではないものの，別のなんらかの調査項目によって影響される場合。たとえば上記の例で，対象者における男女比は 1:1 なのに，無回答者の多くが男性であった，どうやら男性のほうがより無回答になりやすいようだ，という場合がこれにあたる。この場合は，Kish が述べているように，性別を補正変数にしたウェイティングが有用である。
- **非ランダム欠損(MNAR)**: ある人が無回答であるかどうかが，調べたい調査項目によって影響されている場合である。たとえば，上記の調査がダイエットについての意識調査であったとしよう。自分の体型に劣等感を抱いている女性は，この調査にたいして無回答になりやすいかもしれない。とすると，100 人のなかには体型への劣等感が強い女性が多いものと想像される（しかしそれを確認するのは難しい）。この点を無視して 9900 人から得たデータを集計した結果は，「体型への劣等感が強い女性」の意識を反映していないというバイアスを受けたものになっているかもしれない。こうしたバイアスの補正は，非カバレッジによるバイアスの補正(後述)と同様きわめて困難であり，単なるウェイティングによっては解決できない。

無回答の補正のなかでも，**項目の無回答の補正**は調査そのものへの無回答の補正よりも一般的であり，広く容認されています。項目の無回答を補正する際には，なんらかの回答を代入する(重複させる)のが一般的です。

いっぽう，**非カバレッジの補正**はもっと困難です。カバレッジ率は標本そのものからは得られないからです。

カバレッジ: 母集団と抽出台帳との一致の程度のこと。たとえば，ある会社の従業員満足度を調べるために，その会社の社員名簿を台帳として標本を抽出し調査したとしよう。このとき，「抽出したのにデータがない」のが無回答，「従業員なのに社員名簿に載っていない」のが非カバレッジである。
非カバレッジの補正： これは厳密には理由 2.に属するものと思われる。

理由 4. 統計的調整

これは，事後層化ウェイティング，比推定量，回帰推定量によって行われます。

> **統計的調整**：いま母集団における男女比が 5:5 であるとしよう。調査を行ったあとで，標本における男女比がたまたま 6:4 となっていることを発見したとする。
> たとえ無作為標本であっても，運悪くこのような結果が得られることはあり得るから，このこと自体はウェイティングを行なう理由にはならない。しかし，調べたい変数の分布が男女で大きく違っていることがわかっている場合には，標本の男女比が母集団の男女比と大きくかけ離れていることを知りつつ，あえてウェイトなしの集計を行う(そしてその結果，男性の性質をより強く反映した統計表を得てしまう)のは，もったいないような気がするだろう。
> こうした場合，標本を男性と女性にわけ(これを事後層化という)，男性に 5/6=0.83，女性に 5/4=1.25 のウェイトをかけることがある。これに類似した方法として，比推定量・回帰推定量がある。

技術的文献では，標本抽出の過程で用いられなかったコントロールによって，分散を縮減することに関心が向けられています。しかし，こうしたコントロールや方法が実務的に重要になるのは，無回答や非カバレッジによるバイアスを縮減するためです。

> **分散の縮減・バイアスの縮減**：上記注で紹介した例では，標本抽出のプロセスにおいて偶然に生じるばらつきに集計結果が左右されてしまうことを防ぐために事後層化ウェイティングを行っている。Kish はこれを指して「分散の縮減」と表現している。
> いっぽう，標本の男女差と母集団の男女差のちがいが偶然によって生じたのではなく，非カバレッジや無回答によって生じていたとしても，同じ手法を用いて事後層化ウェイティングを行うことができる。この場合のウェイティングは上記の例とは異なり，むしろ標本抽出における歪み(バイアス)を集計結果から取り除くための手法であるとみなすことができる。Kish はこれを「バイアスの縮減」と表現している。

こうしたバイアス縮減は，利用可能ななんらかのチェックデータを用いたウェイティングによって行われます。原理的にいえばこの方法は，理由 6 で述べる標準化と同種のものです。十分な調整のためには，データとソフトウェアの両方を吟味する必要があります。

分類が複数である場合，個別の調整セルのウェイトは n が小さいために不安定になったり，データの欠落のために未知になったりします。しかし，レイキング，マージナル，リム，SPREE などと呼ばれる反復推定量を用いれば，複数の周辺統計量を用いることができます。現代のコンピュータと大量のデータファイルのおかげで，こうした方法も次第に便利なものとなっています。

文献のなかには，あるケースのウェイトをその正確性($1/\sigma_i^2$)に比例させる方法が登場しますが，これは理論的には魅力があるものの，現実への応用可能性は高くありません。

理由 5. 標本の結合

近年では結合に利用できる標本の数が増えているため，この方法の人気・重要性・現実性が増大しつつあります。

標本の結合のためには，なんらかの形でのウェイティングが常に必要です。ウェイトがなんであるかをはっきりさせておき，ウェイトの測定のしかたが標本のあいだで異なっている可能性について注意しておく必要があります。さらに注意しておきたいのは，どんなときでも(地域なり社会的・人口学的階級なりといった)さまざまな領域があって，調べたい変数の分布はそれらの領域によって異なっているのに，国レベルの標本はそれらを結合してしまっているということです。

昨今では，いくつかの国から得た国レベルの標本を標準化し結合することさえあります。空間的な統合と同様, 時期の異なる標本をより長いタイムスパンのローリング標本へと結合することもあるでしょう。

統計量の結合における成長分野のひとつがメタ分析です。メタ分析の兆しは 1924 年にすでにあらわれていました。個別事例の累積も，標本の結合の単純な特殊ケースであるということができます。

理由 6. コントロールに一致させるための調整

これが行われる際の事情はさまざまです。理由 4 では，標本抽出におけるばらつきを減らすことを主眼において，事後層化や比推定量を用いる場合について述べました。いっぽう理由 6 が主に指しているのは，ある枠母集団(台帳)から得た標本を，別のなんらかの目標母集団(標準母集団)へと調整することです。たとえば，ある地域から得た標本を国レベルの母集団と一致させるためにウェイティングしなおすような場合です。または，ある国ないしある時期の標本を，別の国ないし時期に対してウェイティングしなおす場合もあるでしょう。

一般に, 標本のある下位クラスを目標母集団のある領域と対応づけてウェイティングしなおすことになりますから，こうしたコントロールのためのデータが，標本においても目標母集団においても利用可能でなければなりません。セル数が多すぎる場合には，コントロール・データは利用できなくなり，標本のケース数が少なくなりすぎて不安定になってしまいます。この場合は，周辺調整と反復適合が用いられます。

理由 7. 非無作為標本の調整

これはデータをチェックするために行われるものです。当該の標本について，(操作的には定義されていないにせよ)ある母集団「台帳」が存在するのだとみなせば，これは理由 6 の特殊事例であるとみることができます。

非無作為標本の調整：調査協力者を知人・友人から募って予備調査を行ったところ(つまり非無作為標本), ある意見への賛成率が 60%であった。いっぽう，無作為標本による本調査では賛成率は 70%であった。本来，この２つの標本は比較できない。しかし，データをチェックするための便宜的な方法として，予備調査の標本における(たとえば)男女比が，本調査の標本における男女比に等しくなるように，前者にウェイティングを行うことにする。
ここで，本調査における男女比を，母集団における男女比を示すチェックデータであるとみなせば，この状況は理由 6 の状況と同じだとみなすことができる。

理論的にいえば，非無作為標本の調整は「クオータ」標本抽出と似ています。クオータ抽出では，年齢，性別，地域，職業などの割合が母集団における割合と一致するようにクオータの大きさを決めた上で，各クオータに対して非無作為抽出を行います。いっぽう，いま述べている非無作為標本の調整においては, クオータごとの抽出のかわりにウェイティングを行っているわけです。

なお，クオータを定義する変数が複数ある場合は周辺適合を用いればよいでしょう。

4.　ウェイティングの基本的手法

ウェイティングの主な手続きは 4 つあります。それらは技術的にみても異なり，分散に及ぼす影響も異なるので，それぞれについて注意を向ける必要があります(**表 4**)。

手法 1. 個別ケースへのウェイト(ICW)

これはもっとも一般的で，単純で，実用的で，そして柔軟な手続きです。近年では ICW を扱うことができるコンピュータとプログラムがありますので，まさにこの形容があてはまります。他の手続きについて考える際，その比較の対象は ICW です。どの手法も，ICW と比べると統計量の分散が大きくなる傾向があります。

複雑な統計的多段階抽出の結果，標本の要素 j (j=1,…,n)の抽出確率が p_j，(カバレッジ率を含む)回答率が r_j となったとしましょう。標本要素へのウェイトは，この 2 つの積に反比例する値 $w_j = 1 / p_j r_j$ となります。無作為標本であれば，すべての標本要素について p_j と r_j を求めることができるはずです。

統計量のうち，平均はウェイトの和 Σw_j によって「規準化」(標準化)する必要があります。たとえば

$$(y\text{ の平均}) = (\Sigma w_j y_j) / (\Sigma w_j)$$
$$(y^2\text{ の平均}) = (\Sigma w_j y_j^2) / (\Sigma w_j)$$
$$(xy\text{ の平均}) = (\Sigma w_j y_j^2) / (\Sigma w_j)$$

となります。この規準化のおかげで，ウェイトは「真の」ウェイト w_j と比例している非負値であればなんでもよいことになります。たとえばウェイトは $w_j' = fw_j$ でもよいわけです。

EPSEM 抽出においては，真のウェイトは $w_j = 1/f = N/n$，その合計は $\Sigma w_j = N$ となり，ウェイト $w_j' = fw_j = f/f = 1$，その合計は $\Sigma w_j' = n$ となります。

ウェイトの規準化：男 10000 人，女 10000 人からなる母集団から，男 200 人，女 100 人を層別無作為抽出した。話を単純にするために，カバレッジ率も回答率も 100%であったとしよう。

標本割合は，男では 200/10000=0.02，女では 100/10000=0.01 である。層のなかでは無作為抽出がおこなわれているので，抽出確率 p_j も，男で 0.02，女で 0.01 であるとみなすことができる。

- 本文中でいう**「真の」ウェイト**は，男では 1/0.02 = 50，女では 1/0.01=100 である。つまり，本文中の「真の」ウェイトとは，＜その合計が母集団サイズに一致するウェイト＞である(男では 50×200 人=10000; 女では 100×100 人=10000)。このウェイトは，ふつう非常に大きな値となってしまい，不便である。
- **実際によく用いられているウェイト**は，「真の」ウェイトに標本割合(全層の母集団において標本が占めている割合)を掛けたものである。上の例では，標本割合は 300/20000=0.015 なので，男のウェイトは 50*0.015=0.75，女のウェイトは 100*0.015=1.5 となる。もしカバレッジが 100%であれば，このウェイトは＜その合計が標本サイズに一致するウェイト＞である。このウェイトは 1 に近い値になるので扱いやすい。さらに次のような長所がある。今度は，男 200 人，女 200 人を抽出した場合について考えよう。この場合はウェイティングが不要なのだが，あえてウェイトを求めてみる。「真の」ウェイトは男女ともに 1/(0.02) = 50 である。いっぽう，それに標本割合 400/20000=0.02 を掛けたウェイトは 50*0.02=1 となる。つまりこのウェイトは，＜ウェイティングがいらないときに 1 となるウェイト＞であり，とてもわかりやすい。

なお，文献のなかでは拡大合計 $\Sigma w_j y_j$ が登場することがありますが，実際に用いられることはめったにありません。

平均に関して，また他の複雑な統計量に関して，このウェイティング手続きは「整合的」な推

定量を提供します(厳密に不偏な推定量を提供するとはいえません)。しかしそれらの分散は，ウェイトのなかの不均等性によって増大していきます(5.2 で後述)。

なお，標準化された相対ケースウェイト $W_j = w_j / \Sigma w_j$ を用いれば

$$(y\text{の平均}) = \Sigma W_j y_j$$

となります。しかし，多くの複雑な調査分析においては，単一の標準化係数ではうまくいきません。ベースが変わってしまうからです。

標準化された相対ケースウェイト：上述の例で，真のウェイト(男 50，女 100)のかわりに，それをその合計(つまり母集団サイズ 20000)で割った値(男 0.0025，女 0.005)をウェイトとして用いる，という考え方もできる。こうしておけば，ある変数について全体平均を求めるとき，その変数の個々の値にウェイトを掛けて合計した値を求めるだけで済むので，その点では便利である。しかし実際には，特定のグループにおける平均(たとえば，ある年齢層だけについての平均)を求めることもあるだろう。その場合はかえって不便になる。

ところで，層 h における標本割合 n_h/N_h の逆数 N_h/n_h をその層のウェイトとする，という考え方もできます。しかしこれは，抽出確率 p_j が層のなかで等しいことが前提です。このように抽出確率と標本割合をごっちゃにするのは誤解を招きます（ないし，それ自体が誤解の産物です)。たとえば，層 h においてサイズ N_h の母集団からサイズ n_h の標本が抽出されたとしましょう。標本割合は n_h / N_h ですが，その層における抽出確率が n_h / N_h であるとは限りません。抽出確率は実際の確率操作に照らして根拠付ける必要があります。さもなくば，私たちの標本抽出には主観的な側面が含まれているということ，なんらかのモデルに基づいていることになってしまいます。また，母集団サイズ N_h について正確であてになる値が手に入らないが，抽出確率はわかっている，という場合もあります。

手法 2. 統計量のウェイティング

下位母集団 h での y の統計量が $\overline{y}_h$ であるとします。それぞれの下位母集団に，適切な相対ウェイト W_h を与えることができれば($\Sigma W_h=1$ とします)，下位母集団を結合し，全体での y の統計量 $\overline{y}_w = \sum W_h \overline{y}_h$ を求めることができます。

この方法は ICW よりも優れているのは以下のような場合です：

- 個別のケースが利用可能でなく，統計表を結合する場合
- 抽出手続きが異なるいくつかの層を結合する場合
- 統計量が比較的に単純である場合(例，平均，合計)。

いっぽう，この方法はある単一の調査について複雑な分析をする場合には有用でありません。

この方法のためには，信頼できるウェイト W_h を確かなソースから手に入れる必要があります。もし手にいれられるのなら，それは ICW のためにも使えます。

手法 3. ケースの重複

下位クラスのセルのなかでケースを重複させる方法です。この方法は，使い勝手の観点からウ

ェイトなしのデータ行列を用意しなければならない，という状況において利用されることがあります。特に項目の無回答に対しては便利な方法ですし，また複雑な分析的統計量を求める場合にも便利です。

ケースを重複させ，下位クラス間を「ぴったり」マッチングさせれば，バイアスは除去できるでしょうが，無作為抽出ではなくなります。ふつうは，無作為抽出と「ぴったり」の中間の妥協的な方法が用いられます。たとえば，セル h における回答率が r_h であるとき，$(1-r_h)$の割合のケースを重複させて擬似ケースをつくります。重複させるケースは，$(1-r_h)/r_h$ の確率で無作為に選択したり，「ぴったり」マッチしているケースを探しだしたりします。

ケースを重複させる方法では，個人にウェイティングする方法よりも分散が大きくなります。しかし，重複させるケースの割合が小さければ大した問題にはなりません。さらに，分散の増大を抑える「多重反復」という方法もあります。

次の点に注意してください。真のケースが n 個で，ケースを重複させた結果合計で m 個のケースを得たとします。計算プログラムはケース数が m 個だと出力しますが，これをそのまま受け入れてしまうと，大変な誤りとなります。あとでケース数を数えるために，真のケース n 個に「タグ」をつけておいてもよいのですが，「有効ケース数」はもっと少なくなり，$n'=n/(1+L)$ となってしまいます(5.2 で後述)。

手法 4. ケースの削除

情報を投げ捨ててしまうことは，統計学的にみて致命的な過ちであるように思われます。にもかかわらず，次のような状況ではケースの削除も許されるかもしれません。

- 大標本を得たが，いくつかの地域において抽出割合が異なっていた。国レベルで複雑な分析をしたいので，標本割合が最低である地域にあわせて EPSEM 標本をつくりなおす。
- ほかの分析のために，ある小さな領域だけ非常に大きな標本を得た。すべての領域をあわせて複雑な分析を行いたい。標本をそのまま用いるのは困難であり，またそうすることによって取り立てて正確さが増すともいえない。そこで，上の領域の標本の一部に「タグ」をつけ，比例標本をつくりなおす。
- 層間の無回答の差異を調整するための妥協案として，ケースのごく一部を除去する。

直観に反するのですが，ケースのごく一部(5%程度以下）を除去した際の分散の増大は，同程度の率のケース重複による分散の増大とたいして変わりません。

5.　ウェイティングに反対する理由

理由 1. 厄介である

たとえ優れた計算プログラムを用いていても，ウェイティングはしばしば厄介なものです。計算プログラムのインターフェイス上の欠陥のせいで，分析がより複雑になることもあります。さらに，複雑で分析的な統計量や，有意性検定のような推測統計量については，ウェイティングの下での推定量や標本誤差が知られていない場合もあります。

理由 2. 分散が大きくなる

抽出確率の差異がまったく「最適」ではなく，ランダムないし無計画(不規則)なものであるとき，ウェイティングによって分散が増大します。抽出確率の不均等性が台帳や無回答のせいで生じている場合は，概してこれに相当します。さらに，ウェイティングによる分散の増大は，クラスタリングによる分散の増大とは異なり，下位クラスにおいても消えずに残りますし，あらゆる統計量に影響します。たとえるなら，母分散がσ^2から$(1+L)\sigma^2$へと増えてしまうようなもの，ないしケース数が n から n/(1+L)に減ってしまうようなものです。

分散が大きくなる：ここで Kish が想定している状況は，本節の冒頭で述べられているように，抽出確率が無計画に決まっており，ウェイト値と調べたい変数とのあいだに強い関係がない状況である。このとき，ウェイティングは統計量の分散を増大させる。
いっぽう，ウェイト値と調べたい変数が強く関係している状況もある。この場合，統計量の分散はウェイティングによってむしろ減少することもある。こうした場合には，ウェイティングは抽出確率の際によるバイアスを取り除くという点で望ましいだけではなく，分散の縮減という観点からみても望ましいことになる。この点を踏まえ，L の推定式を修正する考え方も提案されている(Little&Vartivarian, 2003)。
この点について，Kish(1992)は次のように注記している：たとえば，収入・資産の平均を調べるために，最適割り当てによって標本を得たとしよう。しかしその標本は，購買行動に関しては比例割り当ての標本より非効率かもしれないし，収入・資産の中央値について調べる際でさえも，比例割り当て標本より非効率かもしれない。従って，ウェイティングが多目的な調査においてなにをもたらすかを調べるためには，指標として（1+L）を用いるのがもっとも良い。

このような分散の増大は，調査計画の段階において以下の式で推定できます。調査変数の分散が層のあいだで大体同じだと仮定すれば，層 h の母集団におけるサイズを W_h，ウェイトを k_h とすると

$$(1+L) = (\Sigma W_h k_h)(\Sigma W_h/k_h)$$

式の形をみるとわかりますが，k_hのかわりに，ウェイトの逆数 $1/k_h$，ウェイトに比例している数 ck_h, ないし c/k_h を用いてもかまいません。

ICW ウェイトなら次のようになります。層 h の標本サイズを n_hとすると

$$(1+L) = \Sigma n_h \Sigma n_h k_h^2 / (\Sigma n_h k_h)^2$$

標本のケース数を $n=\Sigma n_j$，ケース j のウェイトを k_j とすると

$$(1+L) = n \Sigma k_j^2 / (\Sigma k_j)^2$$

さてここで，L は分散の相対的変化を表しています。分散/(平均の二乗)を**相対分散** cv^2 と呼びますが，L は相対的ウェイト k_j の相対分散 cv^2 に一致します。つまり，$(1+cv^2) = (1+L)$はケースのウェイトのちらばりによって決まるわけです。

L は cv2 に一致する：なぜなら

$$L = n\Sigma k_j^2/(\Sigma k_j)^2 - 1$$
$$= \{n\Sigma k_j^2 - (\Sigma k_j)^2\} / (\Sigma k_j)^2$$

kj の平均は$\Sigma k_j/n$，分散は$\{n\Sigma k_j^2 - (\Sigma k_j)^2\}/n^2$ なので

$$= (k_j\text{の分散}\times n^2)/(k_j\text{の平均}\times n)^2$$
$$= cv^2$$

ですから，不規則なウェイトを用いることによって分散がどのくらい増大するかを事前に見積もるためには，ウェイトの**累積分布**をプリントアウトすると良いでしょう。

理由 3. バイアスより分散のほうが大きい

バイアスを B^2, 分散を S^2 としましょう。ある統計量の平均平方誤差は次のように表されます:

$$MSE = S^2 + B^2 = S^2\ (1+B^2/S^2)$$

平均平方誤差(MSE)：推定量と真値の差の二乗の期待値。たとえば，標本平均と母平均の間には常にある程度のずれがある。このずれの大きさを表す概念が「平均の MSE」である。上で Kish が述べているのは，「標本平均と母平均のあいだのずれは，バイアスに由来する部分と，データのばらつきに由来する部分とにわけられる」というごく一般的な考え方である。

さて，B^2 と S^2 を比べると B^2 のほうが小さいという場合があります。ウェイティングによって取り除かれるバイアスの比 B/S は，ウェイティングした平均とウェイティングしていない平均の差 $\left(\bar{y} - \bar{y}_w\right) / SE(\bar{y})$ によって推定できます。

バイアス比 B/S (そして B^2/S^2, B, S)は，それぞれの統計量が個別に持っており，あるひとつの調査のなかでも，統計量によって大きく異なっている可能性があります(そして調査というものは，たいてい複数の目的を持っているものです)。特に注意すべきなのは，下位クラスにおける S^2 はかなり大きい(したがって B^2/S^2 はかなり小さい)という点です。

にもかかわらず, ウェイティングについての判断はある調査のなかで一律に行ったほうが便利だ，とみなされることが多いように思います。ウェイティングに際して，私たちは 2 つの平均平方誤差を比較しなければなりません。すなわち，ウェイティングされておらずバイアスを含んでいる平均平方誤差 $MSE = S^2(1+B^2/S^2)$と，ウェイティングされバイアスが取り除かれた平均平方誤差 $MSE=(1+L)S^2$ です。

ひとつの合理的な妥協策として用いられているのが**切り詰め（トリミング)**です。つまり，相対ウェイトにおける極端に大きな値を切り詰めてしまうわけです。では，どのウェイトをどのように切り詰めればよいでしょうか？ その理論と方法は，最近になってようやく手に入るようになりました。

理由 4. モデルに基づく批判

モデルに基づく分析という観点に立つ論者たちは, 標本抽出におけるバイアスを補正する必要はないのだという批判をおこなっています(2b を参照)。

理由 5. 広報上の理由，ないし倫理的な理由

はっきり目立つかたちでのウェイティングは, 広報上の理由や倫理的理由から避けられることがあります。ウェイティングが誤用されれば，主観的に望ましい，偏見をうけた結果がつくられてしまう可能性があるからです。バイアスを受けた標本抽出が往々にして人々の目をすり抜けるのに対し，明示的なウェイトはデータを公衆の面前にさらします。たとえば，たとえウェイト W_h が極端な値であっても, あらゆる層の平均 $\bar{y}_h$ をつかって, 結合された平均 $\bar{y}_w = \sum W_h \bar{y}_h$ を求めてしまうことができるでしょう。(ジャーナリストがよくやるやり方ですね。異なる価格

システムを持っている経済のあいだで生活費を比較するとき，彼らはその指標として，いっぽうでは自動車とテレビの値段を用い，他方では家賃と医療費を使ったりします。）

6. 統計量が異なれば影響も異なる

ウェイティングがもたらす影響は統計量によって異なります。多くの人がこの点について誤解しています。おそらくは，”THE Bias”といった言い回しのせいでしょう。

バイアスの大きさやバイアス比 B/S は，変数によっても，ベースとなる下位クラスによっても大きく異なりますし，推定される統計的関数によっても異なります(例，中央値の B/S と平均の B/S は異なります)。

- **拡張合計** $\Sigma y/f$ は，ウェイトに由来するバイアスにもっとも敏感です。たとえ無回答が一様かつランダムに生じていても，拡張合計は過小な推定となります。$\Sigma N_h(y_h/n_h)$のような拡張合計も，N_h におけるバイアスに敏感です。ただし，縦断的研究において時点間で合計の差ないし比を求める場合は，バイアスの影響をあまり受けません。
- **平均**は，通常は合計にくらべてあまり影響を受けません。アメリカではいまや無回答率がおそろしく高いのが当たり前になっていますが，それでも標本調査が生き延びています。その理由は，たいていの調査変数において，回答者と無回答者のあいだにさほど大きな差がないからです。バイアスが大きくなるのは，下位クラスのなかで，ウェイトの差と調査変数の差が連動しているときだけです。ウェイトと調査変数のどちらかが，下位クラスを通じて一様であれば，バイアスの合計は小さくなる傾向があります。
- **下位クラスの平均**についていえば，標本ベースの減少と比例して分散が増大します。つまり，標本ベースの減少とともにバイアス比は減少します。**下位クラスの平均の差**についていえば，バイアスは多くの場合同じ方向を向いているので，バイアス比は劇的に下がる傾向があります(Kish 1987, 2.7)。
- **分析的統計量**(例，回帰係数)のウェイティングした推定は，計算的・方法論的・哲学的問題を抱えています。この問題について一般論を述べることは不可能だと私は思います。もっとも，表 2b のようなウェイティング嫌いに対して私はいささか疑念をもっています。
- **標本誤差**，推測統計量，有意性検定もまた，そのウェイティングした推定は計算的・方法論的・解釈的に深刻な問題を抱えています。

7. 結論：ウェイティングすべきか，せざるべきか

抽出確率の不均等性(とくに無回答や台帳の問題)に対処しバイアスを取り除くために，ウェイティングが**常に**必要だ，という意見はあまりに極端です。同時に，(モデル論者のように)ウェイティングは**常に**不要だ，という意見も極端です。残念なことに，これらの極端な意見を信奉する統計学者がおり，彼らに従う人が大勢います。

私が信じるのは，バランスであり，トレードオフです。ウェイティングされた推定は，バイアスの減少と同時に分散の増大を招く，ということを覚えておられますか。$S^2(1+B^2/S^2)$と$(1+L)S^2$のあいだのバランスが大事なのです。

ですから，こう考えましょう。望ましいのは，バイアスと推定の分散の増大とのバランスが保たれることです。その際の主たる基準は，平均平方誤差(MSE)を減少させることです(あるいは，私たちの推測の誤りを減らすためのなんらかの公式が基準となるかもしれません)。

バランスのとりかたは変数によって異なります。変数が異なれば，また統計量が異なれば，B^2 と S^2 も大きく異なるからです。バイアスは，標本全体について調べているときは重要ですが，下位クラスについて調べているときはさほど重要でないでしょうし，特に下位クラスの平均を比べているときにはさらに重要でなくなります。

とはいえ，ある調査においては整合的で一様なやり方を採らなければならないかもしれません。切り詰め(トリミング)を用いた妥協案が最善である場合もあるでしょう。

最後に一言。私のこの概観のなかには，私には答えられない問いが登場しています。この概観をきっかけにして，どなたがそれらの問いへの答えを探してくださるのを，私は願っています。

(終)