

インターネット人口の推計値に関する考察：公表統計数値の問題点について

笹渕，祥一
九州大学大学院芸術工学研究院応用情報部門

古川，淳一
九州大学大学院芸術工学府芸術工学専攻修士課程

<https://doi.org/10.15017/2794896>

出版情報：芸術工学研究. 5, pp.41-52, 2006-03-30. 九州大学大学院芸術工学研究院
バージョン：
権利関係：

インターネット人口の推計値に関する考察：公表統計数値の問題点について

A Consideration on the Estimates of Internet Population : Concerning the Problems with the Announced Statistical Figures

笹渕祥一、古川淳一

SASABUCHI Shoichi, FURUKAWA Jun-ichi

These days more and more people make use of the Internet. Every year there appear several estimates of the Internet using population in Japan, but they are quite different from each other. In this paper, we compare these estimates by using interval estimation and hypothesis testing, which are basic methods of statistical inference. Section 1 is devoted to the introduction and motivation of our research. In Section 2, related research are presented. A review of the statistical inference is given in Section 3. In Section 4, some assumptions are given in order to compare the estimates. In Sections 5 and 6, two of the representative estimates are compared. More estimates are compared in Section 7. We also discuss the causes of the differences between the estimates in Section 8. Concluding remarks are given in Section 9.

1. はじめに

現代社会では、テレビ、新聞、雑誌、専門書、報告書、さらにはインターネット等の様々なメディアにおいて、種々の調査に基づく統計数値が発表されている。それらの中には、その数字をそのまま信じていいのか疑わしいものも含まれているのではないかとと思われる。本稿では、特に「インターネット人口の推計値」に注目して、このような問題について考察する。

数年前から、わが国の「インターネット人口」(インターネット利用者人口)は5,000万人を突破していると言われている。しかし、実は、この「インターネット人口の推計値」は、調査機関によって大きく異なっている。

総務省の編集による「平成16年版 情報通信白書」([1])には、わが国の「インターネット利用人口」は、平成14年(2002年)末で6,942万人(人口普及率54.5%)、平成15年(2003年)末で7,730万人(人口普及率60.6%)と書かれている。

一方、財団法人インターネット協会の監修による「インターネット白書2004」([2])には、わが国の「インターネット利用者数」は、2002年12月調査で5,533.7万人、2003年12月調査で6,229.7万人と書かれている。また、2004年2月調査で6,284.4万人(人口普及率49.25%)とも書かれている。

ほぼ同時期の調査であるにも拘わらず、二つの白書の推計値の差異は非常に大きい。「利用者人口」に関しては、[1]では平成15年(2003年)末で既に7,000万人を大きく上回っていることになっているのに対し、[2]には「2004年末に7,000万人を超えるのは難しそうだ」という見通しが明記されている。また、「人口普及率」に関しても、[1]には平成15年(2003年)末で「人口普及率は60%を超えた」と書かれている

のに対し、[2]では2004年2月調査で「人口普及率は49.25%」となっている。評価や見通しの面でも、両白書の食い違いは、あまりにも大きいと言わざるを得ない。

2000年の「INTERNET Watch(ホームページ)」の「コラム」([3])には、次のように書かれていて、興味深い。

“毎年6月に発表される「通信白書」(郵政省)と「インターネット白書」(日本インターネット協会編、インプレス刊)の中でも特に注目度の高いのが、日本のインターネット人口の推計値である。今年は発表のタイミングがほぼ同時で、「通信白書」が2,706万人、「インターネット白書」が1,938万人と大きな違いが出たために、新聞報道などでは「官民格差」「どちらが本当？」といった言葉が見出しとなったほどだ。”

このように既に5年以上前に二つの白書の推計値の大きな格差が指摘されていたのだが、その後も格差は改まることなく続いている。なお、「通信白書」(郵政省)は現在の「情報通信白書」(総務省)の前身である。

さらに、より多くの推計値を比較するために、「インターネットビジネス白書2002」([4])の「日本のインターネット人口」を見てみると、2000年末から2001年8月の間に行われた調査に基づく推計値として、「情報通信白書」(調査時点2000年末)による4,708.0万人、「Access Media International」(調査時点2001年2月)による3,263.6万人、「ジュピターメディアメトリックス」(調査時点2001年7月)による2,507.2万人、等の多種多様な数字が列挙されている。なお、この詳細は第7節で改めて紹介する。

これらの推計値は、いずれも「権威ある調査機関」による「統計的な調査」「統計的な分析」に基づくもの、ということになっているから、このように全く異なる数字が発表されているということは、一般人の「統計」あるいは「統計学」に対する信用を著しく失うことにつながるのではないだろうか？

実際、[1]における「インターネット利用人口」、あるいは[2]における「インターネット利用者数」等のいずれかを、「わが国のインターネット人口の正しい数字」としてそのまま引用して、わが国の情報通信政策のあり方や情報通信社会の状況について議論しているようなメディアも見られる。調査機関によって数字に大きな食い違いがあるにも拘わらず、一つの発表のみを「うのみ」にして議論されていることは大きな問題である。

そもそも、一体、何故にこのような大きな食い違いが生まれてきたのだろうか？

実は、このようなことは、「インターネット人口」に限ったことではない。例えば、昨年(2005年)の新聞紙上でも、警察の犯罪統計に関して「統計の信用性揺らぐ」(西日本新聞2005年2月15

日朝刊38面)、税関の貿易統計に関して「統計の信頼性揺らぐ」(西日本新聞2005年2月17日夕刊1面)等の見出しで、統計数字の食い違いの問題が大きく報じられている。また、国政選挙の前等に盛んに行われる「世論調査」に基づく「政党支持率」「内閣支持率」等が、発表する新聞社やテレビ局によってかなり異なっていることも、周知の事実であろう。

本稿では、以上で述べてきたような問題意識の下に、各調査機関が発表した「インターネット人口の推計値」に対して、統計的推測理論を用いて数量的に詳細に比較し、検証を行う。ただし、この考察は、「インターネット人口」の「正しい推計値」を求めることが目的ではない。筆者は、様々な「権威ある調査機関」が全くバラバラの数字を発表することによって、一般人の「統計」「統計学」に対する信用が失われることを危惧している。「インターネット人口の推計値」に関する考察を通して、現実に行われている統計的調査分析の問題点を浮彫りにして指摘し、注意を喚起するとともに具体的な対応策を提言することが主な目的である。また、実際に公表されているデータに関しては、分析のために必要な情報が不足していたり、調査機関によって調査時期が異なる等、様々な問題点があるので、これらを補い適切に処理して統計的推測理論を適用できるようにする具体的な方策を与えていく。このことも本稿のもう一つの目的であると言える。

第2節以下の内容は、次の通りである。

2. 関連する研究・文献等について
3. 「全数調査」「標本調査」と「統計的推測」
4. 推計値の比較検討に当たっての仮定と方針
5. 統計的推測手法による二つの推計値の比較(1)
6. 統計的推測手法による二つの推計値の比較(2)
7. 多数の推計値の比較
8. 推計値の「本質的な食い違い」の原因の考察
9. おわりに

2. 関連する研究・文献等について

「インターネット人口の推計値」そのものは、既に述べたように、多くの調査機関によって独立に発表されている。しかし、他の機関の発表との食い違いについて触れるようなことはされていない。各機関は、それぞれが正しいと信ずる調査方法と算出方法に基づいて発表しているのだから、当然のことかも知れない。

前述の[3]では、「通信白書」と「インターネット白書」における推計値の大きな違いに注目し、その原因を「調査方法の違い」「利用者の定義の方法」等にあるのではないかと、かなり詳細な解説を加えている。しかし、本稿のような統計的

推測理論を用いての数量的な考察については書かれていない。

また、前述の[4]においても、「推計値の差異」に触れているが、「利用者の定義や調査方法、対象とする年齢などの違いから生じる差」等と簡単に述べられているのみである。

一方、日本統計学会誌や応用統計学会誌（「応用統計学」）、あるいは、統計学関連学会や日本数学会統計数学分科会の研究発表では、統計学の理論と応用に関する様々な研究成果が発表されている。また、電子情報通信学会論文誌ではインターネットに関する様々な工学技術的研究の成果が発表されている。しかし、いずれにおいても、本稿で扱うテーマに関する発表は見られない。

[5]では、いわゆる「統計のウソ」について、様々な興味深い例が挙げられており、本稿の第 8 節の内容とも深く関連している。この本は、「統計のウソ」に関するバイブル的名著として知られており、40 年近く前（英語版の原著は 50 年以上前）に書かれたにも拘わらず、その内容は全て現代にもあてはまるものと言える。しかし、インターネットが登場する以前に出版された本であり、当然ながらインターネット人口の話などには触れられていない。

広く一般の統計調査や社会調査に関する問題点を扱った文献は、数多く出版されている。前述の[5]もこれに含まれるが、比較的新しいものとして、例えば[6]では、「世論調査」に関する様々な問題点が詳細に論じられている。それらは、本稿の第 3 節や第 8 節の内容と深く関連するものである。しかし、本稿で扱うテーマに関する記述や考察は見られないようである。なお、[6]の巻末の「引用文献リスト」には、世論調査を含む社会調査全般に関する豊富な文献が掲載されている。

本稿における考察は、ある意味で、これまでの様々な研究分野の「隙間」にある考察と言えるのではないかと考えられる。

3. 「全数調査」「標本調査」と「統計的推測」

この節のかなりの部分は、前述の[3]の内容と重複することになるが、第 4 節以降の考察の背景の説明として必要であるので、[3]の内容にさらなる考察を付け加えて改めて述べる。

「インターネット利用者人口」を調べるためには、まず、「インターネット利用者」の「定義」を明確にする必要がある。これに関しては、第 8 節で改めて論じることにする。ここでは、取りあえず、「インターネット利用者」の「定義」は一義的に定められたものとして議論を進める。

「日本のインターネット利用者人口」の正確な数を調べるた

めには、日本国民全員を調査する必要がある。しかし、このような「全数調査」は、膨大な費用と時間と労力を必要とし、現実的ではない。5 年に 1 度の国勢調査は全数調査と言えるが、「インターネット利用者人口」のような日々刻々と変化する数字を調査するには不向きであろう。

全数調査に近い方法で、かつ「現実的な」方法として、個々のプロバイダーの契約者数を、全てのプロバイダーに関して合計すれば、「インターネット利用者人口」となるのではないかと考えられる。しかし、一人で複数のプロバイダーと契約していたり、逆に 1 件の契約で複数の人間が利用したりしているから、この方法では正確な数は出ない。また、大学の施設やいわゆる「インターネット・カフェ」のように大勢の人間が自由に入出入りして利用している場所では、各人の利用頻度も様々であり、また、これらの利用者には個人でもパソコンや携帯電話等をインターネット接続している者も多いであろうから、「重複を許さない正確な利用者数」を得るのは不可能であろう。

全数調査が無理ということになれば、「標本調査」に基づいて推測することになる。これは、日本国民全体の中から一部分を「標本」として抽出して調査し、その結果に基づいて、日本国民全体の姿を「推測」するものである。第 1 節で紹介した各調査機関による推計値は、すべて、この方法による。すなわち、実際には、この方法しかない、ということであろう。

この「統計的推測」を用いる場合には、標本が日本国民全体の姿を的確に反映したものである必要がある。標本が、特定の性質を持つ集団に偏っていたりすると、適切な推測はできない。この問題に関しても、第 8 節で改めて論じることにする。

各調査機関の発表の大きな問題点は、それぞれの発表が、「一つの推計値」すなわち「点推定値」のみであり、「推定の誤差」等について触れられていないことである。これでは、一般人は、「真の値」と「推定値」の誤差がどのくらいあるのかわからないし、各種の推計値の間の食い違いが「単なる偶然誤差の範囲内」であるのか、それとも何らかの「本質的な違い」によるものかさえも判断できず、「一体どれが正しいの？」で終わってしまう。そこで、本稿では、統計的推測理論における基本的手法である「区間推定」「仮説検定」等を用いて検証を行う。

4. 推計値の比較検討に当たっての仮定と方針

本稿のように、自らが現場での調査を実施するのではなく、既に発表されているデータのみに基づいて統計的手法で考察を行う場合、議論のために必要なデータが完全には入手

できないとか、理論上の必要な条件が満たされていないということが、現実問題として、しばしば起こり得る。その場合、入手可能なデータのみで議論を展開できるように適当な仮定を設けたり定式化を行ったりする必要がある。

まず、総人口がわかっているならば、「インターネット人口の推計」と「インターネットの人口普及率の推計」は同等であるから、問題を「母集団における比率に関する推測」の問題として考えてよい。

一般に、「推計値と真値との誤差」あるいは「異なる調査機関による推計値の差異」の原因としては、「標本抽出の際の偶然変動による誤差」と「それ以外の何らかの本質的な原因による誤差」とがある。前者は、あくまで「単なる偶然」によるものであり、標本調査に基づく推定である限り、避けることのできない、やむをえないものである。一方、後者は、調査方法、分析方法、計算方法等に何らかの本質的な問題がある場合であり、その具体的な原因を調べることが重要であると考えられる。

本稿では、「異なる調査機関による推計値の差異が、単なる偶然によって生じる誤差の範囲内にあると言えるか？」という観点に立って、これを統計的推測理論を用いて検証する。

実際に発表されている推計値を比較検証するに当たっては、次のような問題点がある。

- (1) 調査方法や算出方法に不明な部分が多い、あるいは、計算のために必要なデータが記載されていない。
- (2) 調査機関によって調査方法や算出方法が明らかに異なっている、あるいは、恐らく異なっていると考えられる。
- (3) 調査機関によって調査時期が異なっている。

このような状況により、これまで、例えば前述の[3][4]等のように、「調査方法等が異なるのだから、推計値が異なるのは当たり前」として、簡単に片付けられてしまっていたように思われる。

まず、(1)と(2)の問題点を補う方策について論じる。

実際、例えば、[1]と[2]の両白書では、調査と推計値の算出に当たっては、インターネットの使用形態、世帯の構成人数、年代、等の種々の条件を考慮して、かなり複雑な調査と計算を行っていることが窺える。しかし、公表された情報が不足していること、あるいは、計算方法に不明な部分があること等のため、これらの過程を厳密に検証することはできなかった。また、煩雑な計算による推計の検証は、新たな理論的問題を生じてしまうこと、異なる推計値の比較を困難にしてしまうこともあるため、むしろ行う必要はないと考える。実際、[7]では、[8]に記載の調査・計算方法を詳細に検討して、それに基づく「信頼区間」の構成を試みているが、そこには、いくつ

かの解決困難な問題が生じているので、本稿では割愛する。

このような状況下では、実際に各調査機関によって行われた調査や計算を完全に再現して検証を行い、その上で、「異なる調査機関の推計値の差異が単なる偶然による誤差の範囲内にあると言えるか？」を確かめるということは、不可能であろう。そこで、何らかの「単純化」を行うことが必要となる。

例えば、[1]においては、「平成15年末での人口普及率の推計値は60.6%」とされている。これは、仮に[1]の推計値が「当時の全国民から完全無作為抽出された標本」に基づく数値であったならば、標本(調査回答者)の中の丁度60.6%がインターネット利用者であったはずである、ということの意味している。また、もし、「人口普及率60.6%」が正しい数値(真の値)であったならば、当時の全国民から完全に無作為に標本を抽出して調査した場合、平均的には(かつ最も高い可能性をもって)、その中の60.6%がインターネット利用者であることが期待される、ということをも意味している。

そこで、本稿では、各調査機関による「人口普及率の推計値」を、単純に、「その時点での日本国民全体から完全無作為抽出された標本(調査回答者)の中に占めるインターネット利用者の比率」であると仮定して、誤差評価等の考察を行うことにする。このように「単純化」した設定により、推測統計学の理論に従って計算を行って、「異なる調査機関による推計値同士の比較」を行い、「その差異が偶然による誤差の範囲内にあると言えるか？」を調べることが可能となる。これは、あくまで、「問題点(1)(2)のために推計値同士の比較が困難」という状況を回避して、「公表されている推計値同士の間の矛盾の有無」を検証するための一種の「便法」である。

また、次節以降で用いる統計的推測手法はすべて、標本数が多いときの近似的手法である。信頼度や有意水準の値に「(約)」と書いているのは、このためである。実際には、以後現れる標本数の値は十分に大きいので、「(約)」と書かなくてもよいくらい近似の精度は高い。それらの近似も、前段落の「単純化」の仮定の下で理論的に正当であることが保証される。

しかし、このように「単純化」を行ったとしても、考察を進めるためには、「標本数(調査回答者数)」のデータが必要である。標本数が不明の場合、それによって得られた推計値の誤差の評価が全く不可能であり、統計的推測手法によって「差異が偶然誤差の範囲内か？」を調べることはできない。これを補う最も単純な方策の一つは、「推定の誤差はゼロであると仮定すること」であり、これは、「推計値を真値と同一視して扱うこと」に他ならない。もう一つの方法としては、実際の標

本数は不明であるが、「標本数を適当な値に想定すること」である。本稿では、この二つの方法を用いて考察を進める。

次に、問題点(3)を補う方策について述べる。

今後、本稿全体を通して、調査時期の異なる推計の結果を比較するに当たっては、次のように考えるものとする。各調査機関ごとに時系列的に見た推計値は、いずれも、新しい調査時期に基づく推計値ほど、「インターネット人口の推計値」および「人口普及率の推計値」とともに増加している。これは、常識的に考えても当然のことであろう。従って、「真のインターネット人口」および「真の人口普及率」も、時間の経過と共に増加している、と仮定するのが妥当であろう。大前提として、このように仮定しておくことにより、調査時期の異なる推計の結果を比較することが可能となる。

また、本稿では、「異なる調査機関による推計値の差異」に対して、「単なる偶然誤差の範囲内にある」という可能性を、できるだけ排除しないようにする、という方針の下で考察を進める。これは、権威ある調査機関が恐らく大変な費用と時間と労力をかけて実施したであろう調査・分析・計算等に何らかの本質的な問題があると断定することには、慎重であるべきと考えるためである。第7節で具体的に表れるように、標本数や信頼区間の信頼度の設定等を「好意的に」行い、調査時期の違いによる影響等も「好意的に」解釈するのは、このような考えに基づいている。

5. 統計的推測手法による二つの推計値の比較(1)

前節では、推測統計学の理論を用いて議論を行うための前提として様々な仮定を設けたが、実際に異なる調査機関による推計値を比較検討するには、できるだけ同じ時期に行われた調査を取り上げる必要があり、また、やはり調査方法および算出方法に関する詳細な情報があることが望ましい。このような条件を満たすものとして、この節では、第1節で紹介した[1]と[2]の推計値について、詳細な比較を行う。この二つを比較することには、「国(政府)の発表」と「民間の主要調査機関の発表」との間のいわゆる「官民格差」の象徴としての意味もある。

まず、二つの白書の調査概要と推計値を、白書の記載に沿って紹介する。

[1]「平成16年版 情報通信白書」

編集 : 総務省

調査実施時期 : 平成16年(2004年)1月

調査方法 : 郵送調査

サンプリング手法 : 不明

調査対象数 : 不明

有効回答数 : 3,354 世帯 (11,653 人)

有効回答世帯率 : 不明

「平成15年(2003年)末のインターネット利用人口の推計値」 : 7,730 万人

「平成15年(2003年)末のインターネットの人口普及率の推計値」 : 60.6%

[2]「インターネット白書2004」

監修 : 財団法人インターネット協会

調査実施時期 : 2004年2月

調査方法 : 電話調査

サンプリング手法 : RDD

調査対象数 : 100,107 世帯

有効回答数 : 43,027 世帯 (人数は不明)

有効回答世帯率 : 42.98%

「2004年2月調査でのインターネット利用者数の推計値」 : 6,284.4 万人

「2004年2月調査でのインターネットの人口普及率の推計値」 : 49.25%

なお、「RDD」とは、「ランダム・デジット・ダイヤリング」の略で、コンピュータによってランダムに発生させた電話番号に電話して質問するという方法で、現在、多くの電話調査の主流となっている方法である。

[1]では「調査実施時期は平成16年1月」「平成15年末に関する推計値」というように区別されて記述されている。一方、[2]では、そのような区別は記述されていない。代わりに、[2]には、上記の他に、調査方法等の説明は抜きにして「2003年12月調査でのインターネット利用者数の推計値: 6,229.7 万人」という記述もされている。

[1]では「わが国の全人口を1億2,752万人として計算」と明記されている。[2]では、そのような記述は見られないが、「インターネット利用者数推計値」÷「人口普及率推計値」を計算すると、 $6,284.4 \text{ 万人} \div 0.4925 = (\text{約}) 1 \text{ 億 } 2,760 \text{ 万人}$ となる。よって、二つの白書で対象としている母集団はほぼ同じと考えられるので、各々の推計値を比較することは可能であり、また、意味もある。

[1]と[2]の二つの推計値に関しては、調査時期はほぼ同じであり、しかも、[2]の方がわずかながら後なのであるから、前記のような推計値が得られるのは、一見、全く奇妙なことである。しかし、第3節で述べたように、いずれも、全数調査ではなく、標本調査に基づく統計的推定値であるから、「標本抽

出の際の偶然変動による誤差が生じるのは、当然であり、やむを得ないことである。そこで、「二つの推計値の食い違い」が「偶然変動による誤差の範囲内にある」と考えられるのか否かを検証する必要がある。これを、統計的推測の基本的手法である「区間推定」「仮説検定」を用いて検証する。

前節でも述べたように、実際には、両白書とも、かなり複雑な調査と計算を行っていることが窺える。しかし、これも前節で述べた理由によって、各白書における「人口普及率の推計値」を、単純に、「その時点での日本国民全体から完全無作為抽出された標本(調査回答者)の中に占めるインターネット利用者の比率」であると仮定して、考察を進める。

このように「単純化」を行っても、「誤差の評価」を行うためには、「標本数(調査回答者数)」のデータが必要である。[1]では「有効回答数:3,354 世帯(11,653 人)」というデータが明示されているが、[2]では「有効回答数:43,027 世帯」とあるのみで、「人数」のデータが明示されていない。また、二つの白書の調査時期は、わずかではあるが、異なっている。これらの事情により、二つの推計値を「全く対等に」扱って「有意差の有無の検定」等の統計的推測手法を適用することは不可能であり、かつ不適切である。

そこで、前節でも述べたように、まず、本節では、[2]の推計値については「推定誤差はゼロである」と仮定して、[1]の推計値と比較する。これは、[2]の「2004年2月調査での人口普及率の推計値」を暫定的に「その時点における人口普及率の真の値」と見なして取り扱うということである。

以上のような設定の下で、次のように記号を定義する。

- p : 2003 年末のインターネットの人口普及率の真の値
- n : 標本数(調査回答者数)、ここでは、 $n = 11,653$ (人)
- \hat{p} : $n = 11,653$ (人)の標本に基づく p の推定値、
ここでは、 $\hat{p} = 0.606$
- p_0 : \hat{p} と比較する値、ここでは、 $p_0 = 0.4925$

ここで、 p は「真の値」であるが、未知である。一方、 \hat{p} は特定の 11,653 人のデータから求めた「(p の)推定値」に過ぎず、「誤差」を含んでいる、ということに注意されたい。

前節で述べた方針に従って、以後、 \hat{p} を「2003 年末の日本国民全体から完全無作為抽出された標本(調査回答者)の中に占めるインターネット利用者の比率」であると仮定して、考察を進める。

(1) \hat{p} に基づく p の区間推定(信頼区間の構成)

一般に、標本数 n が大きいとき、 p の信頼度(信頼率、信

頼係数)(約)99%の信頼区間は、次の公式で与えられる。なお、この公式については、例えば、[9]等を参照のこと。

$$\hat{p} - 2.58\sqrt{\hat{p}(1-\hat{p})/n} < p < \hat{p} + 2.58\sqrt{\hat{p}(1-\hat{p})/n}$$

また、上式において 2.58 を 1.96 に変えれば、信頼度(約)95%の信頼区間となる。直感的にも当然であろうが、この構成方法から明らかなように、信頼度(約)99%の信頼区間は信頼度(約)95%の信頼区間を包含している。さらに、 \hat{p} の値を固定すると、 n が増加すれば信頼区間は狭まっていくこともわかる。今後、これらの性質を必要に応じて用いる。

今、具体的な値として $n = 11,653$ と $\hat{p} = 0.606$ を代入して計算すると、 p に関する信頼度(約)95%、99%の信頼区間として、各々、次の区間を得る。

信頼度(約)95% : 0.5971 ~ 0.6149

信頼度(約)99% : 0.5943 ~ 0.6177

これは、[1]による推計値 \hat{p} に基づいて計算した限りでは、 p の真の値は、信頼度(約)95% あるいは 99%をもって、上記の区間の中に入っているはずである、ということの意味する。

これらの結果から、次のことがわかる。

(i) $p_0 = 0.4925$ という値は、上記のいずれの区間にも含まれていない。つまり、[2]において与えられている「人口普及率」は、[1]の推計値に基づいて計算した「 p が入っているはずの区間」の中に入っていない。

(ii) $p_0 = 0.4925$ という値は、上記のいずれの区間の下限値(0.5971 および 0.5943) よりもはるかに小さい。つまり、[2]において与えられている「人口普及率」は、[1]の推計値に基づいて計算した「 p が入っているはずの区間の下限値」よりも、はるかに小さい。

なお、ここで求めた p に関する信頼区間に「1 億 2,752 万」を乗じれば、[1]の推計値に基づく「2003 年末のインターネット人口」の信頼区間が得られる。計算すると、次のようになる。

信頼度(約)95% : 7,614.2 万人 ~ 7,841.2 万人

信頼度(約)99% : 7,578.5 万人 ~ 7,876.9 万人

信頼度(約)99%の信頼区間には、約 300 万人の幅がある。これは、標本数(回答者数)が 1 万人程度の場合には、推計値の 300 万人程度までの差異は「偶然変動に伴う誤差」によ

るものである可能性も否定できない、ということを示唆している。

(2) \hat{p} と p_0 の有意差に関する仮説検定

帰無仮説を「 $p \leq p_0$ 」とし、対立仮説を「 $p > p_0$ 」とする統計的仮説検定問題を考える。通常は「有意差の有無の検定」と言えば帰無仮説を「 $p = p_0$ 」とし対立仮説を「 $p \neq p_0$ 」とすることが多いが、今の場合、上記のように帰無仮説と対立仮説を共に「片側仮説」に設定するのは、調査時期のズレを考慮して実際に意味のある検定結果を導くためである。

この仮説検定問題に対する有意水準(約)5% (または 1%) の検定は、以下の方法で与えられる。これについても、例えば、[9]等を参照のこと。ただし、[9]では、帰無仮説を「 $p = p_0$ 」として書かれている。帰無仮説を「 $p \leq p_0$ 」という「片側複合仮説」に変えても、同じ方法が有意水準(約)5% (または 1%) の検定となることは、仮説検定の理論により示すことができる。

$$(\hat{p} - p_0) / \sqrt{p_0(1 - p_0) / n} > 1.65 \text{ (または } 2.33)$$

となれば、帰無仮説「 $p \leq p_0$ 」を棄却して、対立仮説「 $p > p_0$ 」を正しいと判定する。

今の場合、 $n = 11,653$ 、 $\hat{p} = 0.606$ 、 $p_0 = 0.4925$ を代入して計算すると、

$$(\hat{p} - p_0) / \sqrt{p_0(1 - p_0) / n} = 24.535$$

となるので、有意水準(約)5%,1%のいずれでも、帰無仮説「 $p \leq p_0$ 」は棄却されて、対立仮説「 $p > p_0$ 」が正しいと判定される。

つまり、[1]の推計値の誤差を考慮して計算すると、「2003年末の人口普及率の真の値」は、[2]において与えられている「2004年2月の人口普及率」よりも大きいと判定される、という結果になる。

以上の(1)と(2)の方法は、[1]の推計値に関する「偶然変動による誤差」を考慮に入れて、[2]の推計値と比較したものである。二つの白書の調査時期がほぼ同じであり、しかも、[1]よりも[2]の方がわずかながら後であることを考慮すると、上記の「区間推定」「仮説検定」のいずれの結果においても、明らかな「矛盾」が生じていることになる。これは、[1]の推計値の「誤差」を考慮に入れても「説明の付かない矛盾」が生じているということである。

なお、本節で行ったような比較方法を用いれば、複数の推

計値が与えられたとき、そのうち一つだけが標本数がわかっていたら、他の推計値は単に「推計値」のみしかわかっていなくても、さらに、調査時期にズレがあっても、誤差を考慮に入れた統計的推測手法による比較を行うことが可能となる。

6. 統計的推測手法による二つの推計値の比較(2)

前節の推論においては、[1]による推計値 \hat{p} と比較する値である $p_0 = 0.4925$ を、あたかも「2004年2月の人口普及率の真の値」であるかのように扱っている。しかし、実際には、この値も、[2]における「推計値」であるから、「偶然変動による誤差」を含んでいるはずである。これを、あたかも「真の値」であるかのように扱ったのは、前述のように、[2]においては「標本数(調査回答者数)」が不明であるので、 $p_0 = 0.4925$ という推計値の誤差の評価等が不可能であるためであった。

本節では、この弱点を補うために、次のような手法を用いて、さらなる推測を行う。

目標は、何らかの方法で、[2]における推計値の「誤差の評価」を行うことである。今、「調査回答世帯数」が 43,027 世帯であるから、「標本数(調査回答者数)」は 43,027 人以上となる。また、1世帯あたりの人数の平均は、多くとも 10 人以下であると考えてよいであろう。そこで、[2]における「2004年2月調査でのインターネットの人口普及率の推計値： $p_0 = 0.4925$ 」を、「標本数 43,027 人に基づく推計値」または「標本数 430,270 人に基づく推計値」と仮定して、「人口普及率の真の値」の信頼区間を構成してみる。標本数が与えられれば、信頼区間の構成方法は、前節と同様である。計算の結果、次のようになる。

$p_0 = 0.4925$ を「標本数 43,027 人に基づく推計値」と仮定した場合の「人口普及率の真の値」の信頼区間

信頼度(約)95%	: 0.4878	~	0.4972
信頼度(約)99%	: 0.4863	~	0.4987

$p_0 = 0.4925$ を「標本数 430,270 人に基づく推計値」と仮定した場合の「人口普及率の真の値」の信頼区間

信頼度(約)95%	: 0.4911	~	0.4939
信頼度(約)99%	: 0.4907	~	0.4943

これらを、前節で計算した[1]の推計値に基づく p の信頼区間と比較してみると、次のことがわかる。

(i) [2]に基づいて構成した「人口普及率」の信頼区間と、[1]に基づいて構成した「人口普及率」の信頼区間とは、全く重なっていない。

(ii) [2]に基づく「人口普及率」の信頼区間の上限値は、[1]に基づく「人口普及率」の信頼区間の下限値よりも、はるかに小さい。

前節と同様に、二つの白書の調査時期がほぼ同じであり、しかも、[2]の方がわずかながら後であることを考慮すると、やはり、明確な「矛盾」が生じている。これらの信頼区間は、二つの白書における推計値に基づいて計算したものであり、しかも各々の「誤差」をも考慮したものであるから、これらの推計値の食い違いは、「偶然変動によって生じた誤差によるもの」では到底説明できない、「本質的な食い違い」であると判断せざるを得ない。

なお、前節の(1)で信頼区間の公式を与えた際に述べた標本数 n と信頼区間との関係により、信頼度を揃えた場合には、「標本数 43,027 人に基づく信頼区間」は「標本数 430,270 人に基づく信頼区間」を包含する。従って、前節で求めた信頼区間について、前者との「重なり」がなければ後者との「重なり」もないので、今の場合は結果的には後者は必要なかったことになる。しかし、ここでは、「実際の標本数が不明な場合の信頼区間の構成法」を示すために、両者を求めた。つまり、「実際の標本数に基づく信頼区間」は、「43,027 人に基づく信頼区間」と「430,270 人に基づく信頼区間」との中間の区間となることがわかる。

前節では、[1]の推計値 \hat{p} には「偶然変動による誤差」があることを考慮し、[2]の推計値 p_0 の方は「(2004 年 2 月の)真の値」(定数)と見なして、仮説検定も行って、やはり「矛盾」を検出した。本節のように、[2]における「標本数」を適当に設定すれば、両方の推計値に対して「偶然変動による誤差」を考慮に入れた仮説検定を行うことも可能である。詳細は省略するが、これまでの諸結果からも予測されるように、やはり、「矛盾」が検出されて、両方の推計値に対して「偶然変動による誤差」があることを考慮しても「説明の付かない食い違い」が存在する、ということが示される。

なお、前節同様に、本節で求めた「人口普及率」に関する信頼区間に「1 億 2,760 万」を乗じれば、[2]の推計値に基づく「2004 年 2 月のインターネット人口」の信頼区間が得られる。計算すると、次のようになる。

$p_0 = 0.4925$ を「標本数 43,027 人に基づく推計値」と仮定した場合の「インターネット人口」の信頼区間

信頼度(約)95% : 6,224.3 万人 ~ 6,344.3 万人

信頼度(約)99% : 6,205.2 万人 ~ 6,363.4 万人

$p_0 = 0.4925$ を「標本数 430,270 人に基づく推計値」と仮定

した場合の「インターネット人口」の信頼区間

信頼度(約)95% : 6,266.4 万人 ~ 6,302.2 万人

信頼度(約)99% : 6,261.3 万人 ~ 6,307.3 万人

前節の(1)の最後で求めた信頼区間に比べて、区間の幅がずっと狭くなっている。これは、前節の場合よりも標本数が大きいために「偶然変動による誤差」が小さくなったことを意味している。

7. 多数の推計値の比較

前節までは[1]と[2]における推計値の比較を行ってきたが、この節では、多数の調査機関による推計値を同時に比較する。できるだけ多くの推計値の比較を試みるために、少し古い[4]のデータを用いる。[4]に記載の「インターネット人口の推計値」は次の通りである。カッコ内が「調査時点」であり、古い順に並べている。

[a] 情報通信白書(2000 年末)4,708.0 万人

[b] Access Media International(2001 年 2 月)3,263.6 万人

[c] ビデオリサーチネットコム(2001 年 4 月)4,062.0 万人

[d] Nielsen//NetRatings(2001 年 7 月)3,023.0 万人

[e] ジュピターメディアマトリックス(2001 年 7 月)2,507.2 万人

比較に当たって、このデータの大きな問題点は、調査時期にズレがあること、標本数(回答者数)が明記されていないことである。何らかの方法で、これらを補う必要がある。

個々の調査機関に問い合わせる等すれば、回答者数等の情報は得られるかも知れないが、あえてそれは行わなかった。この節の目的は、回答者数等の情報が欠落している状況での比較の方法を示すことにある。また、仮に情報があつたとしても、得られる結果は本節での結果と大差ないものになるであろうと考えられる。

まず、[1]には平成 12 年(2000 年)末の「インターネット人口の推計値」「人口普及率の推計値」として、4,708 万人、37.1% という数字が記載されている。4,708 万人 \div 0.371 = (約)1 億 2,690 万人であるから、これを「調査時点でのわが国の全人口」として採用する。これにより、各調査機関の「インターネット人口の推計値」から算出した「人口普及率の推計値」は、次のようになる。

[a] 37.1%

[b] 25.7%

- [c] 32.0%
- [d] 23.8%
- [e] 19.8%

次に、「標本数(回答者数)」としては、[1]の平成 16 年(2004 年)1 月調査での 11,653 人に近い「キリのいい数」として、各調査機関とも「1 万人」とであると想定する。

以上の設定を行うことにより、各調査機関の「人口普及率の推計値」に基づいて、第 5 節と同様の方法を用いて、「人口普及率の信頼区間」を構成することが出来る。ここでは、信頼度(約)99%の信頼区間を求める。その計算結果は次の通り。

- [a] 0.3585 ~ 0.3835
- [b] 0.2458 ~ 0.2682
- [c] 0.3080 ~ 0.3320
- [d] 0.2270 ~ 0.2490
- [e] 0.1877 ~ 0.2083

これらの数値に各々、当時の「全人口」である「1 億 2,690 万」を乗じることにより、「インターネット人口」に関する信頼度(約)99%の信頼区間として、次の結果を得る。

- [a] 4,549.4 万人 ~ 4,866.6 万人
- [b] 3,119.2 万人 ~ 3,403.5 万人
- [c] 3,908.5 万人 ~ 4,213.1 万人
- [d] 2,880.6 万人 ~ 3,159.8 万人
- [e] 2,381.9 万人 ~ 2,643.3 万人

まず、これらを、「ほぼ同じ時期の調査に基づくもの」と見なし比較してみる。[b]と[d]には「重なり」があるので、「矛盾した結果」とは言えず、「単なる偶然による誤差の範囲内」と見なすこともできよう。しかし、この二つ以外には、全く「重なり」がないので、明らかに「互いに矛盾した結果」とであると判定される。

次に、「調査時点に若干のズレがあつて古い順に並べていること」を考慮すると、[b]と[c]に関しては、「2ヶ月間の間に爆発的に(500 万人以上)インターネット利用者が増えた」と「極めて好意的に」解釈すれば、一概に「矛盾している」とは断言できない。しかし、これら以外は、やはり、明らかに「互いに矛盾した結果」となっている。(なお、[d]と[e]は「全く同時期」と考えている。)

ここでは、信頼度 99%の信頼区間を考察した。第 5 節の(1)

で信頼区間の公式を与えた際に述べたように、信頼度を 95%にすると、各々の区間は狭くなるため、「重なり」は減ることになり、より「矛盾」が生じやすくなる。信頼度 99%というのは、できれば「矛盾」を生じないように、「好意的に」設定した数値と言える。(なお、「好意的に」設定する理由は、第 4 節の最後で述べた。)

また、ここでは、「標本数(回答者数)」は各調査機関とも「1 万人」とであると想定した。が、前節のように、「43,027 人」さらには「430,270 人」というような場合も考察するべきかも知れない。しかし、実は、これも第 5 節の(1)で信頼区間の公式を与えた際に述べた標本数 n と信頼区間との関係により、標本数を増やすと、各々の信頼区間は狭くなるため、「重なり」は減ることになり、より「矛盾」が生じやすくなる。つまり、「1 万人」という数も、やはり、「矛盾」を生じにくいように、「好意的に」設定した数値と言える。

以上の考察をまとめると、[b]と[c]の間および[b]と[d]の間には「明確な食い違い」があるとは言えないが、これら以外の間には互いに、「調査時点の違い」によっても「偶然変動に伴う誤差」によっても説明できない「本質的な食い違い」があると判定せざるを得ない。

なお、多数の群の間での同時比較のための統計的手法としては「多重比較検定法」が知られている。(例えば、[10]等を参照のこと。)が、この節で扱っているデータは統一的に管理された実験に基づくデータではないこと、調査時期にズレがあること、「推定」に関心があること、等の理由により、信頼区間による比較を行った。ただし、このような状況を無視して形式的に「多重比較検定」を行ったとしても、同様の結論が得られるであろう。

8. 推計値の「本質的な食い違い」の原因の考察

推計値同士の差異が、「単なる偶然変動による誤差の範囲内」ではなく、「本質的な食い違い」だとすれば、その原因は、どこにあるのだろうか？ 調査方法に関する情報が与えられている[1]と[2]の推計値を中心にして論じる。この節の内容は、[3]と重複する部分も多いが、その内容にさらなる考察を付け加え、整理して改めて述べる。また、[5][6]等では広く一般の統計調査・社会調査等に関する様々な問題点が詳細に論じられており、それらの内容とも重複する点があるが、「インターネット利用者に関する調査」ということによる特殊な要素も多いので、これに特化して改めて論じる必要がある。

(1) 「インターネット利用者」の「定義」の問題

第 3 節でも触れたように、「インターネット利用者人口」を調べるには、「インターネット利用者」の「定義」を明確にして

おく必要がある。まず、「使用機器」としては、[1]と[2]の二つの白書ともに、「パソコン、携帯電話、PHS、TV ゲーム機等、インターネットに接続できる全ての機器を対象とする」という趣旨のことが記述されているので、これに関しては差はないと考えられる。しかし、「インターネット利用者」と言っても、「毎日、常時利用している者」から「週に 1 回程度」さらには「契約して線を接続しているだけの者」まで、様々であろう。しかしながら、二つの白書には、単に「インターネット利用者」としか記述されていない。恐らく調査の際には、「インターネットを利用していますか？」というような質問がされたのであろうが、調査員の説明の微妙なニュアンスや調査対象者の受け取り方の違いによって「利用者」の解釈が異なる可能性がある。「定義」が統一されていなければ、本質的に相異なる数値を推計していることになる。まず、ここに第一の問題がある。

(2) 「標本の偏り」の問題

第 3 節で述べたように、各調査機関の推計値は、標本調査のデータに基づいて母集団に関して推測して得られたものである。その際、統計的推定の方法を用いているのであるが、この手法は、標本が母集団から「偏りなく」「無作為に」抽出されている、ということを大前提としている。もし、標本の抽出に「偏り」「作為」があれば、適切な推定値は得られない。この問題に関して、以下、(2-1)から(2-4)で詳しく論じる。

(2-1) 「調査対象の選び方」による「偏り」

[2]では、RDD 方式の電話調査を採用しているもので、その点では一応「無作為抽出」と言えるだろう。しかし、固定電話だけでなく携帯電話やPHSにも電話したのか等の情報は記述されていない。最近の若者の単身世帯では、固定電話を設置せずに携帯電話やPHSのみ所持の場合が多く、また、そういう若者にインターネット利用者が多いことを考え合わせると、もし固定電話のみにしか電話していないとすれば、これによって「偏り」が生じている可能性がある。また、[2]では、「対象地域」を「主要 26 都市およびその近郊」としている。都市部の方が他の地域よりもインターネット普及率が高いのではないかと（このこと自体、調査しなければわからないが）と考えられるので、これによって「偏り」が生じている可能性もある。一方、[1]には調査対象の選び方が明記されておらず、議論の材料すらない。

(2-2) 「回答した者」と「回答しなかった者」の間での「偏り」

そもそも、「インターネットに関するアンケート調査」を行った場合、「インターネットへの関心が強い人」「インターネットを利用している人」の方が、「関心が弱い人・ない人」「利用していない人」よりも積極的に調査に応じるであろうと考えられる。

よって、必然的に「偏り」が生じることは避けられない。これは、各調査機関による推計値を実際の値（真の値）よりも高くする方向に作用すると考えられる。

(2-3) 「調査方法」による「偏り」の違い

一般に、アンケート調査の方法としては、「電話調査」「訪問調査」「郵送調査」「街頭調査」等が行われている。(2-2)で述べたように、どのような方法を用いても、ある程度、「偏り」が起こるが、さらに、これらの調査方法による違いも生じる。第 5 節に記したように、[1]は郵送調査、[2]は電話調査を行っている。回答する者の立場になって考えれば明らかのように、郵送調査の方が、電話調査に比べて、「調査に協力しよう」という積極的な意思が求められる。従って、仮に他の条件が同じならば、郵送調査による推計値は、電話調査による推計値よりも、さらに高い値となりやすいと考えられる。（なお、最近の種々のアンケート調査の主流となりつつある「インターネットによる調査」を用いて「インターネットを利用していますか？」等の質問を行うことは当然ながら無意味であることを付け加えておく。）

(2-4) 「回答率」の問題

第 5 節で紹介したように、[2]では、有効回答世帯率は 42.98%で、半分に満たない。さらに、[1]に至っては、白書からは回答率を知ることはできず。なお、[3]には、2000 年の「通信白書」に記載の推計値の算出の基礎となる調査について、「回収率 31.0%」と書かれている。もし、その後身の「情報通信白書」に関しても同程度の数字であるとしたら、回答率は 3 分の 1 程度しかかないことになる。調査に回答しなかった者の状況がわかれば、全く異なる推計結果が出てくる可能性は大きい。

(3) 「算出方法」の不明確さ

推計値の算出方法を検証するには、調査方法、データ、計算方法に関する全ての情報が必要だが、[1]では、これらがほとんど記載されていないので、検証のしようがない。[2]には、かなり詳しい説明が記述されている。インターネットの利用形態を「自宅の機器で」「勤務先・学校の機器で」「携帯電話・PHSで」に分類して、「世帯単位」でかなり詳細な集計を行って、「インターネット人口の推計値」を算出している。しかし、それでも、算出に必要な全ての情報が記載されているとは言えず、また、例えば、「同一世帯の中に異なる利用形態の者が混在している場合、どのように計算しているのか？」等の点で不明な部分が残る。

(4) 政治的な思惑による情報操作(?)

何らかの政治的・政策的思惑によって、故意に数字が「大きめ」あるいは「小さめ」になるように操作する、ということも考

えられる。実際、第 1 節でも述べたように、[1]と[2]とは、わが国のインターネット普及状況に関する評価や見通しにかなりの違いが見てとれるのは、単なる偶然だろうか？ その真偽は定かでないが、一般には、政治的・政策的思惑も、いわゆる「統計のウソ」が生じる大きな原因の一つである。このような問題に関しては、[5][6]等でも詳しく述べられているし、本稿の目的ではないので、深入りは避ける。

第 1 節で述べたように、本稿は、「インターネット人口」の「正しい推計値」を求めることが目的ではない。が、以上の考察から総合的に見て、「利用者の定義」「調査対象の選び方」「算出方法」「政治的思惑」等の影響に関しては不明であるとしても、各調査時点での実際の「インターネット人口」「人口普及率」は、[1]の推計値よりはずっと少なく、さらには[2]の推計値よりも少ないのではないだろうか、ということが推測される。

9. おわりに

「インターネット人口の推計値」に限らず、新聞・テレビ等のマスコミや様々な機関・団体等による「調査結果の発表」に関して、問題と思われることをまとめて述べる。これらは、本稿を執筆する動機となったとも言える。個々の具体例は割愛するが、例えば、新聞やテレビでの種々の「世論調査」の結果発表、テレビのいわゆる「健康番組」等での種々の「実験」の結果発表、等の多くが該当する。

(1) アンケート調査や実験等で得られた限られた数のデータの数値に基づいて、誤差評価等をせずに点推定値のみを挙げたり、有意差の検定等を行わずに「大小」「優劣」「増減」等を断定するような主張が、至る所で見られる。

(2) 全数調査と標本調査、母集団と標本、「確率モデルを仮定する推測統計学的分析手法」と「そうではない統計的分析手法」、等の区別がきちんと認識されていないのではないと思われるケースが、しばしば見られる。

(3) 「標本の偏りの可能性」や「標本の偏りの影響」を全く無視していると思われるケースが、非常に多い。

本稿で扱った「インターネット人口」に関しても、例えば[1]に記載された「推計値」は、各種の報告書やホームページ等において引用されて、様々な主張を行う上での根拠とされており、推計値が「正しい数値」として完全に「一人歩き」しているようである。この推計値が何らかの政策決定等に用いられているとしたら、「実害」が生じている恐れがあるかも知れない。他の調査機関から全く異なる推計値が発表されているにも拘らず、また、[3][4]等にもあるように以前から推計値の食い違いが指摘されているにも拘らず、毎年、同様の調査結果が

発表され続けており、問題が放置されているとしか思われな

い。

標本調査のデータには何らかの「偏り」が生ずるものである。また、仮にデータの収集方法自体には「偏り」がなかったとしても、「理想的な無作為抽出」ができたとしても、推測には「偶然変動による誤差」が生じるのは当然である。従って、「常に真の値のみを発表すべきである」等とは主張しない。問題への現実的な対処方法として、次のようなことを提言したい。

(1) 標本調査に基づく推測の結果を発表する際、「100%正しい」等と誤解されるような表現は、厳に慎むべきである。

(2) 「推計値」を発表する際には、点推定値だけでなく、「信頼度」を明記した「信頼区間」も発表すべきではないか？

(3) 「大小」「優劣」「増減」等を主張したい場合には、「有意水準」を明示した「有意差の検定」の結果を付けるべきではないか？

(4) 「調査対象の選び方」「調査対象者数」「回答者数」「回答率」等を明記すべきである。さらに、回答率が低い場合には、その「原因」「意味すること」「調査結果への影響」等を示すべきではないか？

(5) 「標本の偏り」の影響を考慮して、例えば、「真の値は、この推計値よりも大きい(小さい)と考えられる」というような「注意」「断り」も付記すべきではないか？

統計学あるいは統計科学が、多くの研究者の日々の努力によって各方面で目覚ましい発展を続けていることは、間違いない。しかし、一方で、一般人の「統計」「統計学」に対する信頼性を根底から揺るがすような状況が数多く存在することも事実である。「標本の偏りの問題」をはじめとして、既に 40 年近く前に[5]等において指摘されていることが、現在、ほとんど全く考慮されていないように思われるのは、極めて残念なことである。統計学の教育研究に関わる者は、このような問題も注視していく必要があると考える。

引用文献 (本文内での引用順)

[1] 総務省(編集)、「平成 16 年版 情報通信白書」、発行:株式会社ぎょうせい、平成 16 年。

[2] 財団法人インターネット協会(監修)、「インターネット白書2004」、発行:株式会社インプレス、2004 年。

[3] 萩原雅之、「「インターネット人口」をどのように読んだらいいか」、「INTERNET Watch(ホームページ)」の「コラム」、<http://www.watch.impress.co.jp/internet/www/article/2000/0728/popu.htm>、2000 年。

[4] インターネットビジネス研究会、「インターネットビジネス

白書2002」、発行：ソフトバンクパブリッシング株式会社、
2001年12月。

[5] ダレル・ハフ(著)、高木秀玄(訳)、「統計でウソをつく
法」、発行：講談社、1968年。

[6] 松本正生、「「世論調査」のゆくえ」、発行：中央公論新
社、2003年。

[7] 古川淳一、「インターネットの利用状況に関する統計学
的研究」、九州芸術工科大学芸術情報設計学科卒業論文、
2004年2月。

[8] 総務省(編集)、「平成15年版 情報通信白書」、発行：
株式会社ぎょうせい、平成15年。

[9] 坂田年男・高田佳和・百武弘登、「基礎統計学」、発行：
朝倉書店、1992年。

[10] 永田靖・吉田道弘、「統計的多重比較法の基礎」、発
行：サイエンティスト社、1997年。