

オンライン学習履歴データの統計的取り扱いについて

草 薙 邦 広

広島大学外国語教育研究センター

はじめに

1.1. 「オンライン学習履歴データのお蔵入り」問題

外国語教育、そのなかでも高等教育機関が実施する語学学習カリキュラムにおいて、WBT (Web-Based Training) やそれに類するオンライン教材 (または e-learning) が使用されることは、いまやさして新規性の高いことではなくなっている。昨今のオンライン教材の大半には、学習者の学習履歴を自律的に記録・保存する機能が実装されており、教育業務の従事者が、そのような学習履歴データを利活用し、新たな知見を創出したり、教育成果の質保証や、そのほかの業務改善に役立てたりすることが望まれるようになってきている。現在では、ラーニング・アナリティクス、エデュケーショナル・データマイニング、そして教育工学といった分野において、「教育ビッグデータ」なる概念がキーワードとなっていることは思い出すまでもなく、日々、我々の教育業務のなかで記録・保存されるオンライン学習履歴データの利活用は、教育業界一般において共通の話題となってきているといえよう (e.g., 森本, 2015)。

しかしながら、我が国の高等教育機関における外国語教育の現状は、このような社会のニーズや期待に沿っているとはいいがたいものである。学習履歴を管理する大多数の外国語科目担当教員は、得てしてビッグデータとよばれるような量的に莫大なデータの解析や、それによる知識の創出を職能とするわけではない。さらに、データ解析を専門とする関係者との連携もけっして容易なものではない。そのため、日々累積的に記録・保存されるオンライン学習履歴データは、かろうじて保存こそなされているものの、ほとんどの場合が未整理のまま、利活用の機会を得ないままに留まっている。このような状況を「オンライン学習履歴データのお蔵入り」と言ったとしても過言ではない。

1.2. これまでのオンライン学習履歴データ解析に見られる諸問題

「オンライン学習履歴データのお蔵入り」問題の背景には、様々な要因があると考えられる。第一に、オンライン学習履歴データの利用権限や提供範囲がこれまで法的に曖昧であったことがあげられよう。一般的にいて、外国語教育においては、教育機関が民間の教育関連企業と契約を結び、当該の企業が提供するオンライン教材の使用権を個々の学習者に割り当てる場合が典型である。このような場合、当該の企業がオンライン学習履歴データを保存・管理していることも往々にしてあり、実際に指導にあたる外国語科目担当教員が容易にアクセスすることができないときもある。しかし、オンライン学習履歴の利用権限や提供範囲については、主に教育工学分野を筆頭として、現状に即した新しいガイドラインを策定する動きが見られる (森本・はが・高瀬・鮫島, 2015)。例として、『学習履歴の利活用に関するガイドライン』 (森本他, 2015) では、指導者は、教育業務、そして研究・調査目的のいずれの場合でもデータの提供を受けることができるという指針が明確に示されている。今後、関連する法制度やガイドラインをめぐる動向によっては、さらに明瞭な方向性が見られるようになるであろう。

次に、我が国の外国語教育研究において主流であった研究方法論が、莫大な規模のオンライン学習履歴データの分析に対して、ことごとくミスフィットであったことがあげられよう。我が国の外国語教育研究は、その主たるものに視野を限定すると、1990年代頃より思想的に認知主義(cognitivism)の影響を色濃く受けた第二言語習得研究などの関連分野と交流が盛んになり(草薙, 2016)、第二言語習得研究において既に一般化していた実験計画法と統計的仮説検定を主たる方法論として受け入れた。また、2000年代以降では、心理統計(psychometrics)や教育心理学、そして言語テストといった関連分野の影響から、多変量解析(multivariate analysis)が外国語教育研究の方法論として新たに加わった(e.g., 草薙・川口・田村, 2016; 草薙・石井, 2016)。

外国語教育研究が、長期間その主たる研究方法論の座に据えてきた統計的仮説検定は、得てして小標本を対象とし、ある特定の統制された実験計画上で応用され、多変量解析は主に質問紙データを対象として応用された。また、現在とは異なり、解析技術上の問題によって、研究対象は正規分布に従う変数のみに限られていた。

一方、いうまでもなくオンライン学習履歴データは、調査・実験用にあらかじめ準備された変数ではなく、むしろ教育業務に従事している間に、自然発生するデータである。さらに、測定する構成概念が理論的に、そしてデータの発生に先んじて導出されている調査・実験用の変数とは異なり、各変数が教育業務上、または研究上どのような意味をもつのか、またはどのような概念を測定するのか、そして、どのようにその値を解釈すべきなのかがそれほど自明ではない。

オンライン学習履歴データに見られる数理的特性にも、これまでの外国語教育研究が対象としてきた変数とは大きな差がある。一般的なオンライン学習履歴データは、ある学習コンテンツに対する学生のログイン回数、こなしたコンテンツ数といったカウントデータ、または、問題の正誤反応といった二値データを含む。これら離散分布に従う変数は、近年になって一般化線形モデル(generalized linear model)や種々の発展形モデルが外国語教育研究に導入されるまでは非常に扱いにくいものであって、ときに離散変数を連続変数として扱うといった代替措置が取られるなどしていた。また、あるコンテンツへのログイン時間、コンテンツ消化時間、または回答時間、読解時間、そしてそれらの和の分布などがおしなべて正規分布に従う保証はもちろんない。むしろ、このような時間データに関する解析技術の発展が進んでいる認知心理学、数理心理学、そしてeテストの分野では、対数正規分布、ガンマ分布、ワイブル分布、指数正規合成分布(ex-Gaussian distribution)などがすでに解析に用いられている(e.g., Baayan & Milin, 2010; Heathcote, Popiel, & Mewhort, 1991; 植野・永岡, 2005)。残念ながら、このような種々の分布を使用した分析方法は、外国語教育研究ではあまりなされていない。分布といった変数をもつ数理的特性を十分に活かした分析こそがもとめられる。

このように、オンライン学習履歴データの統計的取り扱いについては、技術的および根本的な問題が複数見られる状況であり、ここに議論の余地が大いにある。

1.3. 本稿の目的と構成

そこで本稿では、高等教育機関における外国語教育プログラムの管理運営を念頭においた上で、オンライン学習履歴データに関する統計的取り扱いについての基本的な指針を検討する。特に、上記の問題点を受け、本稿の視野は、その中でももっとも基礎的な部分である(a)オンライン学習履歴データの測定論、(b)オンライン学習履歴データの数理的性質に限定する。適宜、技術的な説明を含むが、本稿の内容は統計学の基礎を逸脱するものではない。

2. オンライン学習履歴データの測定論

2.1. 観測変数としてのオンライン学習履歴データ

最初に、オンライン学習履歴データの測定に関わる基本的な観点に触れる。まず、第一にオンライン学習履歴データは、観測が可能な観測変数 (manifest variable, observational variable) であり、観測という行為自体は、自律的に、そして自動的に、コンピュータが行い、それを記録するものである。

一方、外国語教育研究がこれまで主たる関心を寄せていたのは、観測変数それ自体よりも、観測変数に影響を及ぼすと考えられ、そして直接的な観測ができない潜在変数 (latent variable) であった。潜在変数は、心理測定やテスト理論の伝統の上では、構成概念 (construct) とも呼ばれ、外国語教育では、「英語の熟達度」、「英語の技能」、「動機づけの種類」あるいは「動機づけの強度」といったものが典型例である。英語の熟達度は、直接的な観測が不可能であるか、または極めて困難であり、あるテスト項目の回答といった観測変数をもって代替されたり、または複数の観測変数、たとえば、あるテストの全項目の回答などといった多数の情報により推定されたりする。これは、研究対象である潜在変数と任意の観測変数群に、一種の連関が見られることを利用しているものである。熟達度を例とした場合は、「真に熟達度が高い被験者は、あるテスト項目 x において正答する確率が高い」といった関係性、または、「もしも仮に真の熟達度が高ければ、このテスト項目には正答する」という因果推論に由来する。

潜在変数と観測変数の関係には大きく分けて2種類あり (e.g., Borsboom, Mellenbergh, & Van Heerden, 2003; Edwards & Bagozzi, 2000)、ひとつは反映モデル (reflective model) と呼ばれるものである (図1)。これは、「潜在変数が観測変数の原因である」という関係に基づく。もう一方の形成モデル (formative model) は、その反対であり、「観測変数が潜在変数の原因である」という関係に基づく。一般的な心理学の伝統では、反映モデルとして測定モデルを構築することが多い。

さて、オンライン学習履歴データは観測変数であるが、これは一般に、「固定的な測定モデルを伴わない変数」であるといえる。典型的には、ある質問紙の質問項目は、その質問項目の書き下ろしに先んじて導出された構成概念と測定モデルに基づいて作成されるものである。しかしながら、観測変数としてのオンライン学習履歴データは、それが測定する構成概念や、測定モデルに先んじて発生している。この点において、オンライン学習履歴データは、これまで外国語教育研究が対象としてきたほとんどの観測変数とその性質を異にしている。

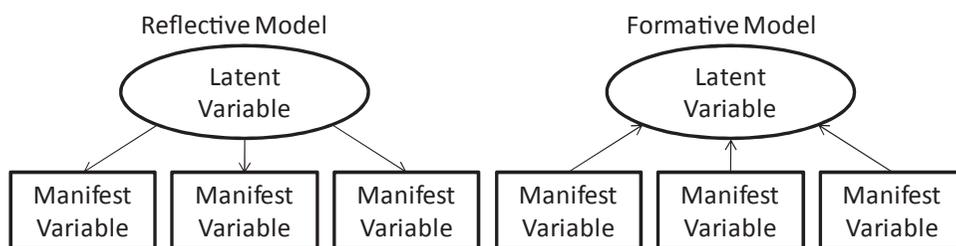


図1 反映モデルと形成モデルの違いを表すパス図の例

2.2. オンライン学習履歴データを含む潜在変数モデル

オンライン学習履歴データの発生は、測定を目指す構成概念の導出に先んじるため、その変数群がどのような構成概念を測定しているのかが研究者の解釈によるところになり、一様に定まる

ものではない。ある研究事例において、学習者のある学習コンテンツにおける消化率、消化時間、ログイン回数、ログイン時間、このようなオンライン学習履歴データから「学習従事度」といった潜在変数の値をある測定モデル下において推定するか、または、類似のある潜在変数を測るものと一意的にみなしたとする。もちろん、このような測定モデルについては、十分な妥当性の検証 (validation) が必要となることはいうまでもない。しかしながら、これまでの外国語教育研究においては、オンライン学習履歴データが、一体どのような潜在変数を測定しているかを明らかにしてこなかった上、そのモデルの妥当性を検証する研究も見られない。

まず、第一に、反映モデルと形成モデルのどちらが適切かを研究者が十分に議論するべきだと考えられる。反映モデルでは、学習従事度といった特性を学習者がもち、観測変数であるところのオンライン学習履歴データは、この一元的な特性から因果を受けているという仮定に立っている。つまり、学習従事度が高い学習者は、より頻繁にログインを繰り返し、長い時間をかけてコンテンツを消化し、より多くコンテンツに正答する、といった関係に注目するのである。

一方の形成モデルでは、より頻繁にログインを繰り返し、長い時間をかけてコンテンツを消化し、より多くのコンテンツに正答する、といった観測が得られる学習者であれば、学習従事度といった変数の値も高い、という関係に注目する。つまり、潜在変数は観測変数の値の結果になる。このような形成モデルの代表例は、「社会経済的地位」(socio-economic status) である。潜在変数であるところの社会経済的地位は、一般的に、「最終学歴が高い」、「高収入の職についている」、「文化的な活動に従事する」といった観測変数によって測定する。しかし、この社会経済的地位の値は、ある観測変数の値が変わるとそれに伴って変化すると考えられるが (e.g., 「最終学歴がより高くなれば社会経済的地位も高くなる」)、その逆は非常に考えにくい (e.g., 「社会経済的地位がより高くなれば、最終学歴が上がる」)。

オンライン学習履歴データに対しては、一見、どちらのモデルであってもそれらが適用可能なものに見える。しかしながら、反映モデルの場合、その妥当性の検証が非常に困難であることが容易に予想される。その理由のひとつは、観測変数であるところのオンライン学習履歴データに影響を及ぼすであろう潜在変数が、これまでの外国語教育研究において広く受け入れているものに限っても、無数にその例があげられることである。たとえば、コンテンツ内の正答率は、一般に熟達度の影響を強く受けるものと考えられる (e.g., 「熟達度が高い学習者は、あるコンテンツに正答しやすい」)。このことを踏まえると、学習従事度といった、ある一元的な構成概念がオンライン学習履歴データのすべてに強い影響を及ぼしており、当該の構成概念の測定において妥当であるなどとは容易に考えることができない。

このように、オンライン学習履歴データをもちいて、ある構成概念を測定することは、特に反映モデルの仮定の上では困難であり、この点に関してより慎重な議論がもとめられるであろう。また、少なくとも著者には形成モデルの方が相対的に自然なモデリングであるように見受けられる。

2.3. 特徴量としてのオンライン学習履歴データとデータマイニング

オンライン学習履歴データを利用して、ある構成概念を測定する心理測定的モデル (psychometric model) を構築することが難しいと述べたところで、比較的多数の変数群であるオンライン学習履歴データを、数理的に集約することの重要性は依然として揺るがないものである。このような状況では、ある特定の構成概念を測定する心理測定的モデルの構築を狙うよりも、データがもつ情報の実用性を優先し、データマイニングの技術を援用することが望まれるかもしれない (e.g.,

植野, 2007a)。データマイニングは、オンライン学習履歴データのような、有益な利活用の可能性が見込まれ、さらにその意味が自明ではない情報から、新たに有益な情報を抽出する技術の総称である。データマイニングの根本思想は、これまで外国語教育研究において主流であった認知主義の見方とは異なり、強い実用主義 (pragmatism) の影響を受けている。人間の認知や心理特性の構造などといった直接的観測が不可能な機構の解明に注力するというよりは、観測できるものを活用し、ある意味において帰結主義的ないし功利主義的 (utilitarianism) な観点から、社会の効用を高めることを優先するのである。ここに、これまでの主な外国語教育研究の研究思想とは隔たりがあることには注意されたい (草薙, 2016)。

データマイニングでは、利活用が見込まれるがその情報が自明ではない変数を、一般に特徴量 (feature value) として扱う。このような特徴量をもちいて、ある種のパターンやクラスを探索的に推定する。クラス分類を例にとると、ある別のカテゴリカル変数が所与のとき (e.g., 試験の可否, 熟達度), このカテゴリカル変数の値を、特徴量から統計的に予測する分類器 (教師あり学習) を作成する手続き、などが典型である。もちろん、分類器に組み込む特徴量の選択 (feature selection) は非常に重要な手続きではあるものの、仮にその分類による帰結が十分に効用をもつならば、特徴量が私たちが取り扱う科学的概念の何を表しているか、そしてそれはなぜか、といったことは比較的重要ではない。

このように、オンライン学習履歴データは、心理測定的モデル下では有益でない変数だとしても、実用主義的な観点に立って特徴量として考えるならば、十分に価値のあるものだといえる。本稿は、高等教育機関における外国語学習プログラムの運営やその業務改善をその背景としている。そのため、実用主義の見地より、観測変数であるところのオンライン学習履歴データを、少なくとも特徴量として扱うことを推奨する。しかしながら、その構築に関する議論、そして妥当性の検証が十分になされるのであれば、心理測定的モデルの援用を妨げるつもりはない。また、外国語教育の研究者や外国語科目担当教員は、一貫して実用主義、帰結主義や功利主義の類を積極的に受け入れるべきだ、といった種の論にも触れない。

より現実的な視点に戻ると、データマイニングの技法、特にクラス分類やパターン認識の技術は、教育実践上、学習者の学習のあり方を把握するだけに限らず、成績予測や、ドロップアウト学生の事前検知など、さまざまな面に応用できる可能性がある。よって、オンライン学習履歴データの特徴量として扱うことの利点はこのような応用面にこそあるといえよう。

3. オンライン学習履歴データの数理的性質

3.1. オンライン学習履歴データの種類と分布

さて、オンライン学習履歴データは、少なくとも特徴量として扱うべき変数であるとして、これらの変数の数理的性質とはどのようなものであるだろうか。仕様などに大きな差異があることは明白だが、WBT などにおいて記録される代表的なオンライン学習履歴データには、以下のようなものがある (表1¹⁾)。もちろん、表1は網羅的なものではなく、授業外学習を念頭にした基本的なWBTなどの例に限る。また、本稿では、「コンテンツ」を、 k 個の学習項目によって構成される単位と定義する。コンテンツの中に、ユニットや、セクションなどといった小単位がある場合もあるが、これらの構成単位をすべて包含してコンテンツとして取り扱う。WBTの仕様によっては、「正答による進行強制」(学習項目に正答しないと、次の項目やコンテンツに進行することが禁じられている) があることもある。この場合、回答数 (率) と正答数 (率) などと同じ値を取りうる。

さて、これらの変数をもつ数理的特性に目を向けると、これらの変数のうち、回答数、人数、ステップ数などは、一見してカウントデータであるから、離散変数として扱うことが望ましいことが明白である。一般的な離散確率分布には、(a) 二項分布、(b) 幾何分布、(c) ポアソン分布、(d) 負の二項分布などがある。また、これらの変数のうちの一部は比率データとして扱うこともでき、正規分布などの連続確率分布を代用しても、実用には大きな問題がない場合もある。

離散変数の代表値をもとめるとき、平均値や標準偏差の使用は不適切となることがしばしばあることには注意するべきである。平均値の代わりに最頻値 (mode) を報告するとよい。項目消化人数など、二値データを扱う場合は、比率を報告してもよいだろう。ログイン回数に類するものについては、ポアソン分布または負の二項分布、消化所要ステップ数などは、場合によっては、幾何分布や負の二項分布に従うものとして扱うことが推奨される。

一方、ログイン時間などの時間に関する変数は、連続確率分布に従うものである。ただし、時間に関する変数の分布は大きな歪度をもつことが知られているため、正規分布にはなく、(a) ガンマ分布、(b) 指数正規混合分布、(c) ワイブル分布、(d) 対数正規分布などに従うとみなすことが多い。

時間に関する変数の代表値を報告するとき、平均値を代表値として用いることは、しばしば統計的に望ましくないこととされている。その理由は時間に関する変数の分布が大きな歪度をもつことにほかならない。中央値や分位数を代用することが比較的推奨されている。

このように、これまで外国語教育研究が対象としてきた変数とは、それらが従う確率分布において大きな差があることに注目すべきである。オンライン学習履歴データを特徴量として扱うのであれば、このような分布の情報を活用する必要性がある。

表1 WBTなどに代表的なオンライン学習履歴データなどの例

データ	説明
個人ログイン回数	ある学習者 x が、ある時点 t 、または期間 $I_{(t_2-t_1)}$ において、コンテンツにログインした累計数
個人ログイン時間	ある学習者 x が、ある時点 t 、または期間 $I_{(t_2-t_1)}$ において、コンテンツにログインした時間の合計
個人消化数 (率)	ある学習者 x が、コンテンツ内における有限の学習項目 k 個について回答した数、またはその比率
個人正答数 (率)	ある学習者 x が、コンテンツ内における有限の学習項目 k 個について正答した数、またはその比率
個人消化時間	ある学習者 x が、コンテンツ内における有限の学習項目 k 個について回答するまでにかかった時間
個人消化ステップ数	ある学習者 x が、コンテンツ内における有限の学習項目 k 個について正答するまでの誤答数、ヒント等の閲覧回数
コンテンツログイン人数 (率)	ある時点 t 、または期間 $I_{(t_2-t_1)}$ において、あるコンテンツにログインした学習者の累計数、またはその比率
コンテンツログイン時間	ある時点 t 、または期間 $I_{(t_2-t_1)}$ において、あるコンテンツに学習者がログインしている合計時間
項目消化人数 (率)	ある時点 t 、または期間 $I_{(t_2-t_1)}$ において、コンテンツ内におけるある学習項目 y について回答した学習者の数、またはその比率
項目正答人数 (率)	ある時点 t 、または期間 $I_{(t_2-t_1)}$ において、コンテンツ内におけるある学習項目 y について正答した学習者の数、またはその比率
項目消化時間	コンテンツ内におけるある学習項目 y について、回答者が回答するまでにかかった時間の合計、平均など
項目消化ステップ数	コンテンツ内におけるある学習項目 y について、回答者が回答するまでにかかった誤答数、ヒントなどの閲覧回数の合計や平均

3.2. オンライン学習履歴データの可視化と最尤推定によるモデル化

これらオンライン学習履歴データは、その後どのような分析を行うにせよ、変数の特徴を捉えるべきであり、まずはその経験的分布を可視化し、その分布関数を吟味することが不可欠である。たとえば、 n 人の学習者について、ある時間 t におけるコンテンツの個人正答数の分布はヒストグラムによって、以下のように可視化できる（図2）²⁾。このデータは、広島大学2016年度前期 Semester 科目「コミュニケーションIB」において、著者が担当した6クラス240名の学生が、ある一ヶ月間、北辰映電株式会社提供のWBTプログラム「ぎゅっとe」（グラマーセクション）を用いて学習した履歴による実例である。

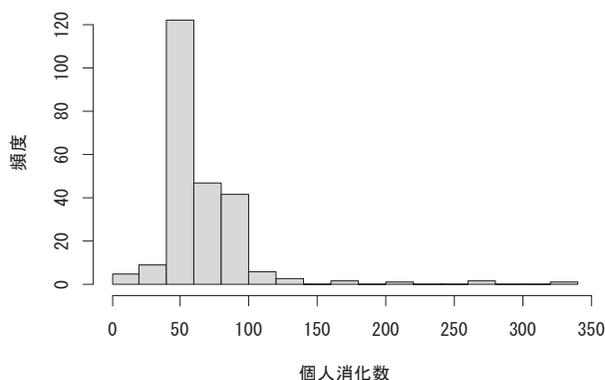


図2 ヒストグラムによる個人消化数の可視化の例 ($n = 240$)

また、個人消化時間の分布も同様に図3のように可視化できる。このデータは、同学生が同 Semester 内全体で学習した時間である。

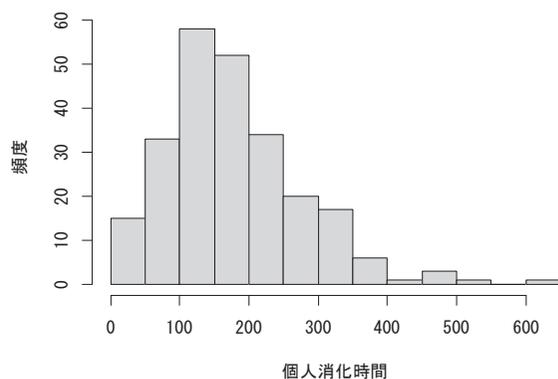


図3 ヒストグラムによる個人消化時間（単位：分）の可視化の例 ($n = 240$)

図3からも明白なように、時間に関する変数は大きな歪度をもつ歪んだ分布を示すため、カーネル密度推定 (kernel density estimation) によって可視化することもあり得る。同じデータをカーネル密度推定によって可視化すると、図4のようになる。

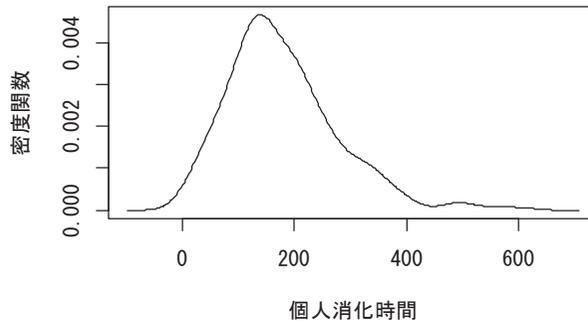


図4 カーネル密度推定による個人消化時間（単位：分）の可視化の例（ $n = 240$ ）

このような強い歪みをもつデータに対して、正規分布を仮定して分析を進めることは情報の捨象にはかならない。たとえば、図4のカーネル密度推定の図に、データから計算した平均と標準偏差を元に推定した正規分布関数（破線）を重ね合わせると、図5のようになる。図5からも正規分布を仮定した分析が明らかに適していないことがわかる。

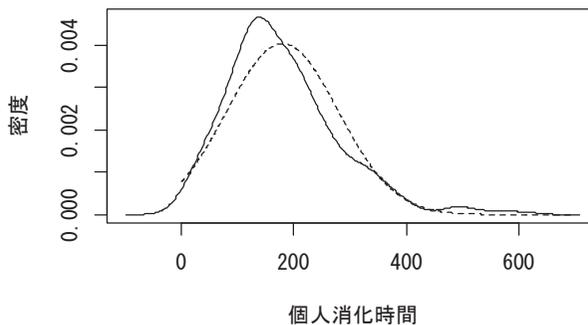


図5 カーネル密度推定による個人消化時間と正規分布による可視化の比較（ $n = 240$ ）

変数の特徴を正確に捉えるために、ヒストグラムやカーネル密度推定によって可視化を行うことは非常に有益であるが、任意の確率密度関数ないし確率質量関数をデータへ当てはめることによって、より柔軟なデータの要約およびモデル化をすることができる。このようなデータの要約およびモデル化は、「ある現象の数理的な近似を得る」ことを目的とするものである。ここでは、最尤推定（maximum likelihood estimation）によって、データへ確率密度関数ないし確率質量関数をフィットさせ、その母数（parameter）を推定する方法について触れる。

最尤推定は、母数 θ の一番もっともらしい値を探す方法であるが、これは観察データにおける母数のもっともらしさ、つまり尤度（likelihood）を最大化する手続きである。現在では、統計解析環境の普及によって、手軽に最尤推定を行うことができるようになった。データのモデルへの当てはまり具合を対数尤度、AIC（赤池情報量基準）、そしてBIC（バイズ情報量基準）などによって評価する。

たとえば、図1における例であれば、これは離散変数であるから、ここではそれぞれ幾何分布、ポアソン分布、負の二項分布などへデータを当てはめ、その当てはまりを比較してみる。それぞ

れの関数に最適な推定母数下における当てはまりの結果は表2のようになる。

表2 それぞれの離散確率分布への当てはまり

分布	対数尤度	AIC	BIC
幾何分布	-1250.10	2502.18	2505.67
ポアソン分布	-2367.20	4736.40	4739.88
負の二項分布	-1131.93	2267.87	2274.83

このデータでは、負の二項分布が最も優れた当てはまりを示した。負の二項分布は、母数として成功回数ないしサイズ母数 r と成功確率 p をもち、その確率質量関数には数種の記述の仕方があるが、一般に (i) 式で与えられる。このデータに対して最ももっともらしい母数の値はそれぞれ、 $r = 5.88$ 、 $p = .08$ であった。この関数を可視化すると図6のようになる。

$$f(k|r, p) = \binom{k+r-1}{k} (1-p)^r p^k \quad (i)$$

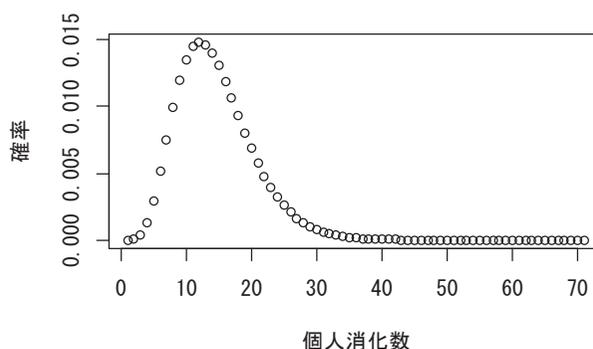


図6 負の二項分布によるモデル化の例

また、図2の消化時間のデータは連続変数であるため、正規分布と、それぞれ時間の解析に使用されるガンマ分布、対数正規分布、指数正規合成分布、ワイブル分布を使って同様にモデル化することができる。この最尤推定の結果を表3に示す。

表3 それぞれの連続確率分布への当てはまり

分布	対数尤度	AIC	BIC
正規分布	-1449.15	2902.30	2909.26
ガンマ分布	-1432.03	2868.07	2875.04
対数正規分布	-1452.09	2908.17	2915.14
指数正規合成分布	-1429.55	2865.09	2875.55
ワイブル分布	-1430.62	2865.25	2872.22

ここでは、表3のように、指数正規合成分布とワイブル分布がともにほかよりも優れた当てはまりを示したが、ここではまずワイブル分布に着目する。ワイブル分布は、母数として尺度母数

λ と形状母数 k をもち、その確率密度関数は、 $x > 0$ のとき、(ii) 式で与えられる。このデータに対して最ももっともらしい母数の値はそれぞれ、 $\lambda = 201.95$ 、 $k = 1.89$ であった。

$$f(x|\lambda, k) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k} \quad (\text{ii})$$

また、指数正規合成分布は、母数として μ 、 σ 、 τ の 3 母数をもち、その確率密度関数は (iii) 式のように与えられる。このデータでは、それぞれ $\mu = 93.13$ 、 $\sigma = 51.90$ 、 $\tau = 85.97$ という値を示した。

$$f(x|\mu, \sigma, \tau) = \frac{1}{\tau} \exp\left(\frac{\mu}{\tau} + \frac{\sigma^2}{2\tau^2} - \frac{x}{\tau}\right) \Phi\left(\frac{x - \mu - \frac{\sigma^2}{\tau}}{\sigma}\right) \quad (\text{iii})$$

図 7 に、このデータの分布を示すヒストグラムに、ワイブル分布によるモデル (実線)、指数正規合成分布によるモデル (破線) を描き足したものを示す。両者がともに、データに対する良い数理的近似になっていることがわかる。

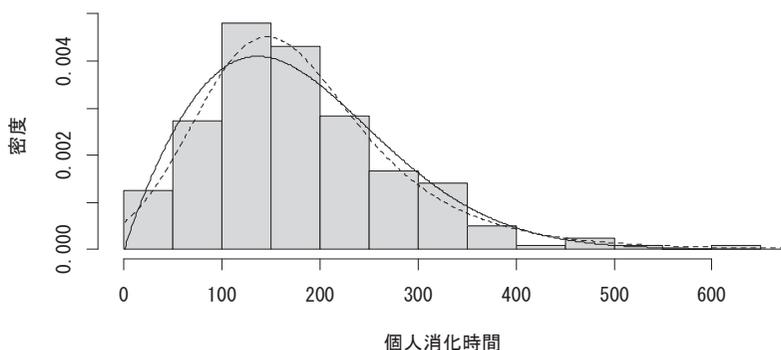


図 7 ワイブル分布と指数正規合成分布によるモデル化の例

このように、最尤推定によってオンライン学習履歴データをモデル化することで、より少ない情報で、変数が元来もつ情報を失うことなく適切に集約することができる。従来の外国語教育研究で行われてきた、正規分布を仮定した上での平均値と標準偏差のみの報告といった統計的処遇と比べると、その利点は明らかである。

しかしながら、分布が元来もつ情報こそ失われてしまうものの、変数の変換によってオンライン学習履歴データを正規分布に近づける方法が悪いわけではない。たとえば、自然対数 (e) を底とする対数変換 (logarithmic transformation) は (iv) 式の要領で行う。対数変換によって正規分布にデータを近づけることができれば、従来使用されていた分析を応用することができるかもしれない。しかしながら、これはあくまでも解析上の工夫であり、データを記述する、ないし現象自体の数理的な近似をもとめる、という目的には向かないということに留意するべきであろう。

$$f(x) = \log_e x \quad (\text{iv})$$

3.3. オンライン学習履歴データのより高度なモデリング

その詳細な記述は本稿の範囲外ではあるものの、近年は一般化線形モデルの普及によって、正規分布以外の変数を応答変数とするモデリングができるようになってきている。オンライン学習履歴データを一般化線形モデルで分析することは、有効な方法であるといえよう。また、近年、外国語教育研究においても流行の兆しを見せているマルチレベル分析やベイズ統計の援用も同様であろう。特に、個人のオンライン学習履歴データは間違いなく、その個人が所属するクラスや、そのクラスを受けもつ担当教員や、学年、大学に複雑にネストした構造をもっているため、マルチレベル分析は有効であろう。また、ある学習項目は、より高次の学習項目のカテゴリーにネストしているとみなすこともできる。このようなデータの階層性を適切にモデリングすることが望まれるだろう。データの時系列性も、時点間において、そのデータが個人にネストしているという点で階層性のあるデータと同様である。オンライン学習履歴データの時系列解析も今後の重要な課題になってくると考えられる。

e テスティングの分野では、オンライン学習履歴データに関連して、永岡・植野（1991）、永岡・植野（1992）、植野（2007b）など、非常に先見的で技術的に高度な分析方法の開発研究が見られる。e テスティングと WBT などとはけっして同じものではないが、項目の困難度や個人の能力を、項目反応理論などを援用し、回答時間なども含めて e テスティングで得られるさまざまなデータから推定しようとするこのような試みについても、外国語教育研究との関わりが強いものである。外国語教育研究者による今後の応用研究が必要になるだろう。

4. 総括

本稿では、オンライン学習履歴データの統計的取り扱いに関する基礎を概観した。本稿の主張は、概ね以下の3つの点にまとめられる。

第一の点は、オンライン学習履歴データはこれまでの外国語教育研究において主流であった研究方法論とミスフィットがあるという点である。この点は、認知主義や実用主義、帰結主義、功利主義といった研究方法論の背景にある思想の対立からも明確である。本稿では、オンライン学習履歴データの利活用は、必然的に実用主義的な観点に依るべき面が多く、思想面において、より詳細な観点の整理がまとめられると主張した。

第二の点は、上記の点に関するものであるが、実用主義的な観点なもと、オンライン学習履歴データを、少なくとも特徴量として扱うことが妥当であるということである。オンライン学習履歴データは、その測定概念の導出に先んじて自然発生するものである。よって、心理測定論的モデルの構成は困難であり、そのようなモデルの運用には水準の高い妥当性の検証手続きが必要であると主張した。

最後の点は、オンライン学習履歴データの数理的特性についてであった。特徴量としてオンライン学習履歴データを利活用するためには、変数としてオンライン学習履歴データがもつ分布といった数理的特性の情報を捨象せずにモデリングすることがまとめられる。この点に関し、所与のデータを、まずは適切な方法によって可視化し、それらがもつ分布の特徴を吟味することが重要であると論じた。また、最尤推定などによって、任意の分布関数にデータを当てはめ、単変量のモデル化を試みる事例を、筆者の教育実践による実例とともに紹介した。

これらの点は、オンライン学習履歴データの利活用に資するであろう統計解析技術のうち、表面的なものの一部にすぎない。今後、より発展的な解析技術の検討が必要となるであろうし、そ

のような優れた解析技術は、今後の教育実践をより豊かなものにしていくだろう。

注

- 1) より詳細な整理については、植野（2007a）などを参照されたい。植野は、LMSにおいて蓄積されるデータの例をあげている。
- 2) 本稿における分析はすべて、総合的な統計解析環境であるR(R Core Team, 2016)を使用した。また、Rのパッケージである、MASS (Venables & Ripley, 2002) を使用している。

参考文献

- 植野真臣（2007a）. eラーニングにおけるデータマイニング, 『日本教育工学会論文誌』 31 (3), 271-283.
- 植野真臣（2007b）. eラーニングにおける所要時間データの異常値オンライン検出, 『電子情報通信学会論文誌』 J90D (1), 40-51.
- 植野真臣・永岡慶三（2005）. ガンマ分布に依るeラーニング所要時間データのオンライン解析, 『日本教育工学会誌』 29 (2), 107-117.
- 草薙邦広（2016）. 認知科学化した外国語教育研究とその後の方向性, 第4回外国語教育メディア学会中部支部外国語教育基礎研究部会年次例会ワークショップ. 名城大学.
- 草薙邦広・川口勇作・田村祐（2016）. 英語教育研究における教育的処遇の効果とリスク, 『中部地区英語教育学会紀要』 45, 61-68.
- 草薙邦広・石井雄隆（2016）. 量的研究の最前線—ベイズ統計とデータマイニング—, 第42回全国英語教育学会埼玉研究大会特別ワークショップ. 獨協大学.
- 永岡慶三・植野真臣（1991）. 回答所要時間における項目応答理論, 『行動計量学』 18 (2), 1-8.
- 永岡慶三・植野真臣（1992）. 自信—正誤反応における項目応答理論, 『電子情報通信学会論文誌』 J75A (2), 407-413.
- 森本康彦（2015）. eポートフォリオとしての教育ビッグデータとラーニングアナリティクス, 『コンピュータ&エデュケーション』 38, 18-27.
- 森本康彦・はが弘明・高瀬亜富・鮫島正洋（2015）. 教育ビッグデータにおける学習履歴の活用に関するガイドラインの作成, 『第40回教育システム情報学会全国講演論文集』 235-236.
- Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, 3(2), 12-28.
- Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review*, 110, 203-219.
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, 5, 155-174.
- Heathcote, A., Popiel, S. J., & Mewhort, D. J. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, 109, 340-347.
- R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from <https://www.R-project.org/>.
- Venables, W. N. & Ripley, B. D. (2002). *Modern applied statistics with S*. Springer, New York. ISBN 0-387-95457-0

ABSTRACT

On the Statistical Treatment of Online Learning Log Data

Kunihiro KUSANAGI

Institute for Foreign Language Research and Education

Hiroshima University

The aim of this paper is to discuss the appropriate statistical treatment of online learning log data that are presently viewed as informative and useful resources to improve foreign language teaching practices in higher education. However, current research methodologies in foreign language teaching research that include experimental designs, null hypothesis significance testing, and psychological scaling are not attuned to utilizing disordered online learning log data. This paper indicates that online learning log data should be treated as “feature values” like in data mining rather than manifest variables under a certain psychometric latent variable model, since the data substantially precedes the theoretical derivation of the construct in measurement. The rationale of this treatment is based on relatively new pragmatic, utilitarian, and consequential perspectives on foreign language teaching research, unlike orthodox research techniques that are strongly supported by cognitivism. Furthermore, this paper underscores the importance of examining mathematical properties of online learning log data. Typically, online learning log data follow non-normal distributions, such as (a) binomial distribution, (b) Poisson distribution, (c) geometric distribution, (d) negative binomial distribution, (e) log-normal distribution, (f) Gamma distribution, (g) Weibull distribution, and (h) ex-Gaussian distribution. Due to these distributional properties, data analysts are concerned about visualizing the empirical distributions of the given online learning log data by histograms or kernel density estimation, and then fitting specific probability density functions to the given data using the maximum likelihood estimation method or other estimation methods. This paper demonstrates the suggested statistical treatment of online learning log data with numerical examples from the author’s teaching experiences.