

文系学生向け統計学習教材のためのメタデータ作成

Basic Design of Metadata of Learning Materials to help Non-Science Majors Learn Statistics

吉根 勝美

Katsumi YOSHINE

南山大学経済学部

Faculty of Economics, Nanzan University

Email: kyoshine@nanzan-u.ac.jp

あらまし：特定分野の専門家には、専門のための統計理論とデータ分析手法が必要である一方、一般の社会人には、仕事上の課題解決のための統計知識とデータ処理手法が必要であり、文系学生の多くが卒業後使う統計学は後者である。統計を学ぶための題材は多種多様遍在しており、自ら学びたい学習者にも、コースウェア作成者にも不便な状況である。本稿では、様々な題材を統計学習のために活用することを目的としたメタデータの設計を試みる。

キーワード：高等教育、統計学教育、学習教材のメタデータ化

1. はじめに

昨年（2013年）は、統計学やデータの重要性が、一般のビジネスパーソンにあらためて認識された1年であった。年初に出版された一般向けの統計学の本⁽¹⁾が30万部を超えるベストセラーとなり、その後も、類書が次々と発売され、経済紙・経済誌でも、統計学に関する特集記事が掲載された。ビジネスの活性化等を目的に、国や自治体等が保有する公共データをオープンにする動きがみられる⁽²⁾⁽³⁾。このように、統計学は、特定の専門分野の専門家が利用するものだけではなく、一般の社会人がデータを分析するためのツールとなり、データを解釈し、社会背景を考察し、意思決定や問題解決のためのリテラシーとなった。

小中高の学校教育では、統計教育を重視する傾向が始まっている。新しい学習指導要領では、2009年度に先行実施された中学数学に「資料の活用」が追加され、1年生でヒストグラムや代表値等を学ぶ。2012年度に先行実施された高校の数学Iには「データの分析」が追加され、データの散らばりとデータの相関等を学ぶことが必修となった。

一方で、これから社会人となる大学生が、統計学教育を受ける環境が十分に整備されているとは、必ずしもいえない。統計教育の質保証制度を確立しようとしている統計教育大学間連携ネットワークの活動は、2012年に始まったところである⁽⁴⁾。

大学における統計教育は、特に文系の大学生にとって、専門的過ぎたり、具体性に欠けたりする可能性が高い。大学の授業では、専攻分野に適合した統計学を学ぶので、卒業後、特定分野の専門家にならない社会人にとっては、大学で学んだ統計学を、仕事で扱うデータにどのように活用すればいいのか、分かりにくいだろう。

特定分野の専門家になる可能性の低い文系大学生には、できるだけさまざまな統計データを目に触れ

させることが望ましい。しかし、実際の統計データは多種多様遍在しており、学生にとっては、自分がどのようなデータを見るべきかわからないし、見たいデータが分かっているにもかかわらず、どこにあるかわからないことが多い。そこで本稿では、さまざまな統計データや題材を、統計学を学ぶのに役立つ教材として活用できるようにすることを目的とし、どのようにメタデータを付与すればよいかについて検討する。

2. 統計を学ぶための題材と統計データ

統計学を学ぶのに有用な資料は、インターネットを通じて数多く公開されている。例えば、総務省統計局は、学習指導要領改訂により内容の充実が図られた統計教育学習に対応するために、統計学習サイトを開設している⁽⁵⁾。国連欧州経済委員会（UNECE）によるガイドブック“*The Making Data Meaningful*”は、統計に関わる職に携わる人が、テキスト、表、グラフ、地図などを使って、一般の人々に統計を伝えるのに役立つ実用的なツールとなることを意図しており、現在 Part 4 まで公開されている⁽⁶⁾。

文系学生にとっては、統計理論を学ぶばかりではなく、表計算ソフトを利用して、実際の統計データを統計分析し、わかったことを文章に書く実習を繰り返すことが望ましい。筆者が経済学部の学生に対して行っている授業で使用した統計データを抜き出したところ、今年4月以降だけでも、次のように多岐にわたる。内閣府「GDP速報値」／経済産業省「商業動態統計調査」／総務省「家計調査」「小売物価統計調査」「小売物価統計調査（動向編）」／厚生労働省「賃金構造基本統計調査」／国税庁「民間給与実態統計調査」／文部科学省「学校保健統計調査」。

3. 統計データの記述

ウェブ上のさまざまなリソースを、カード目録のように記述する語彙であるダブリン・コアは、リソ

ース記述の基礎として広く使われており⁽⁷⁾、統計データの所在情報や書誌情報の記述にも、ダブリン・コアの基本記述要素(表 1)が適用できる。

例えば、小売物価統計調査年報 平成 25 年⁽⁸⁾の一部の書誌情報は、例えば次のように記述できる。

タイトル	第 1 表 調査品目の月別価格及び年平均価格 - 県庁所在市及び人口 15 万以上の市「1001 うるち米(単一原料米, 「コシヒカリ」)」~ 「1101 まぐろ」
作成者	総務省
記録形式	Excel97-2003 ブック
資源識別子	http://www.e-stat.go.jp/SG1/estat/Xlsdl.do?sinfid=000024595377

実際の統計表は、図 1 のように提供されており、例えば次のようにして、データの構成要素を記述する必要がある。ただし、統計データごとに構成要素は異なる。

品目番号 / 品目名 / 銘柄 / 単位	1001 / うるち米(単一原料米, 「コシヒカリ」) / 国内産, 精米, 単一原料米(産地, 品種及び産年が同一のもの), 袋入り(5kg 入り), 「コシヒカリ」 / 1 袋, …… , 1101 / まぐろ / めばち又はきはだ, 刺身用, さく, 赤身 / 100g
都市番号 / 都市名	1 / 札幌, 2 / 函館, 3 / 旭川, …… , 81 / 那覇
年月	2013 年平均, 2013 年 1 月, …… , 2013 年 12 月

4. 教材としての統計データ

表計算ソフトによる統計分析の実習を繰り返すには、実習内容に応じた統計データが必要である。例えば、季節変動を取り除く実習ならば、月単位の時系列データが必要であるが、時系列データからトレンドを抽出する実習ならば、長期間(例えば 10 年以上)の年単位の時系列データが必要である。

そこで、統計データに付与するメタデータには、書誌情報や構成要素の他に、実習内容に適した統計データに、分野を問わずにアクセスできるようにするための記述を加えたい。そこで、実習課題そのもののほかに、データの枠組みをメタデータとして付与することを提案する。図 1 のデータを例にすると、「1 年分の月次データ」と「都市別データ」がデータの枠組みである。共通の枠組みをもつデータならば、同じ実習内容に適用できるはずであり、実習内容とデータの枠組みとの関連を別途記述する(図 2)。

5. まとめ

データ公開の方法としてリンクト・オープン・データ(LOD)が普及するにつれて、CSV や Excel データとして公開されていた統計データも LOD として公開するという動きが欧米を中心に始まっており、経済産業省統計データの一部を RDF 化した試案も発表されている⁽⁹⁾。こうした動きも考慮しつつ、多

種多様な統計データに対して、統計学の学習に有用なメタデータの作成を進めたい。

参考文献

- (1) 西内啓: “統計学が最強の学問である データ社会を生き抜くための武器と教養”, ダイヤモンド社 (2013)
- (2) 各府省情報化統括責任者連絡会議: “日本のオープンデータ憲章アクションプラン”, http://www.kantei.go.jp/jp/singi/it2/cio/dai53/plan_jp.pdf (2013)
- (3) オープンデータ流通推進コンソーシアム: “2013 年度のオープンデータに関する優れた取り組みの表彰”, http://www.opendata.gr.jp/news/1403/140313_000615.php (2014)
- (4) 統計教育大学間連携ネットワーク: “平成 24 年度総合報告書”, <http://www.jinse.jp/pdf/24souhou.pdf> (2013)
- (5) 総務省統計局: “統計学習サイト”, <http://www.stat.go.jp/edu/> (2010~)
- (6) UNECE: “Making Data Meaningful”, <http://www.uncece.org/stats/documents/writing/> (2009~2012)
- (7) 神崎正英: “セマンティック HTML/XHTML”, 毎日コミュニケーションズ (2009)
- (8) 総務省: “小売物価統計調査年報 平成 25 年”, <http://www.stat.go.jp/data/kouri/doukou/2013np.htm> (2014)
- (9) リンクト・オープン・データ・イニシアティブ: “統計表から RDF へ”, <http://linkedopendata.jp/works/statslod/> (2013)

表 1 ダブリン・コアの基本記述要素

タイトル Title	寄与者 Contributor	出处 Source
作成者 Creator	日付 Date	言語 Language
キーワード Subject	資源タイプ Type	関係 Relation
内容記述 Description	記録形式 Format	時空間範囲 Coverage
公開者 Publisher	資源識別子 Identifier	権利管理 Rights

図 1 表計算ソフトの形式で提供されるデータの例

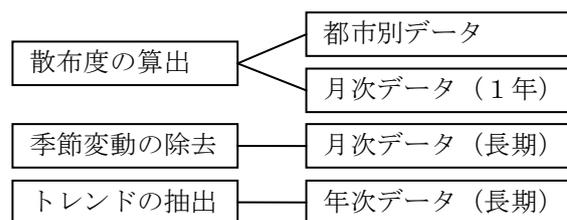


図 2 実習内容とデータの枠組みとの関連性