

# 統計データを含むレポート作成指導のためのシステム化への 公共データの活用

吉根勝美\*1

\*1 南山大学

## Utilization of Public Data to System for Teaching Report Writing with Statistical Data

Katsumi Yoshine \*1

\*1 Nanzan University

現学習指導要領では、高校数学で統計に関する内容が必修化され、次期指導要領では、小中学校の統計教育が充実するが、実社会では、実際の統計データに基づいたレポートが求められる。このとき、使用する統計に関わる分野の知識が必要となるので、こうしたレポートの作成スキルの育成は、大学で実施するのが適切である。本稿では、このスキル育成を目的とするシステム化への公共データと白書・年次報告書の活用について検討する。

キーワード: 公共データ, 白書, 実データ, 統計スキル, 大学教育

### 1. はじめに

現行の学習指導要領では、高等学校の数学で統計に関する内容が必修化され、次期指導要領では、小中学校でも統計教育が充実される予定である。現在の高等学校学習指導要領では、数学 I に「データの分析」が新設され、四分位偏差、分散、標準偏差、散布図、相関係数について、すべての高校生が数学で学ぶことになった。また、小学校は 2020 年度、中学校は 2021 年度より全面実施される次期指導要領では、理数教育の充実の一環として、統計教育を充実させることになっており、改訂のポイント<sup>(1)</sup>の中で「必要なデータを収集・分析し、その傾向を踏まえて課題を解決するための統計教育の充実(小:算数,中:数学)」と説明している。

しかし、小中学校、高等学校で受けた統計教育だけで、大学生・社会人が学習や仕事に使えるとはいえない。小中学生・高校生が統計を学ぶための教材に掲載されているグラフや数表は、学習のための架空のデータであったり、実際のデータであっても分かりやすいように編集されていたりする。

筆者は、大学生・社会人が必要とする統計に関するスキルには 2 種類あると考え、本稿では、2 種類の統計スキルを育成することを目的とする学習システムを設計するにあたり、公共データと白書・年次報告書をどのように活用すればよいかについて議論する。

### 2. 大学生・社会人に必要な統計スキル

本章では、大学生・社会人が学習や仕事に必要な統計スキルについて考察する。

#### 2.1 実際の統計データに関するスキル

小中学校、高等学校で受けた統計教育だけで、大学生・社会人が学習、研究、調査、仕事に使えるとはいえない。小中学生・高校生向けの教科書、資料集、ウェブサイトに掲載されているグラフや数表は、学習のための架空のデータであったり、実際のデータであっても分かりやすいように編集されていたりする。一方で、大学生・社会人が学習、研究、調査、仕事のために使うのは実際の統計データである。実際のデータを活用するには、基礎的な統計学や統計ソフトウェアの

操作方法のみならず、そのデータが属する分野に関わる専門的な知識から、統計の作られ方まで理解している必要がある。

例えば、小中学生が貿易について学ぶとき、主な品目の輸出・輸入先については円グラフが提示されることが多いが、国別の輸出額・輸入額を正式な統計データから得ようとするとは貿易統計の仕組みに関する知識が必要である。日本貿易会が小中学生を対象に開設したサイト<sup>(2)</sup>には、日本の主な輸出品目の輸出先として、自動車輸出額の国別構成比を表す円グラフが掲載されている。このグラフの元となる国別自動車輸出額データは、財務省関税局による貿易統計ウェブサイト<sup>(3)</sup>の「概況品別国別表」から入手することができる。概況品とは、いくつかの品目をまとめて一般的な名称を付したものである。ここでは、たまたま自動車という名称の概況品が見つけられるので、国別自動車輸出額データの入手は比較的容易である。しかし、詳細な品目について輸出額・輸入額データを必要とする場合には、品目ごとにつけられているコード番号を調べた上で統計表を検索する必要が生じ、データ入手までの手間がたいへんかかるようになる。

すなわち、大学生・社会人が学習、研究、調査、仕事のために統計データを使用するときは、データが属する分野に関する知識を前提として、統計データの入手の仕方やその仕組みまで、統計に関するスキルとして必要となる。

## 2.2 統計をことばで表現するスキル

「必要なデータを収集・分析し、その傾向を踏まえて課題を解決」した結果は、何らかの手段で表現されなければならないが、特に言語による表現の指導が十分に行われているとはいえない。参考文献(1)に、「学習の基盤としての各教科等における言語活動（実験レポートの作成、立場や根拠を明確にして議論することなど）の充実（小中：総則、各教科等）」という記述があること自体、小中学校での指導が十分ではなかったことにほかならない。

本節では、統計データを解釈して適切にことばで表現するスキルについて、実際にはどのようなスキルが必要か、実例に基づいて検討する

### 2.2.1 英語能力評価試験での事例

英語能力評価試験のひとつである IELTS（アイエルツ）のうち、英語圏の大学、大学院への留学を目指す人が受検する IELTS Academic で出題される Academic Writing Task 1 のサンプル問題を図 1 に示す<sup>(4)</sup>。ライティングでは Task 1 と Task 2 の 2 問が出題される。Task 2 はいわゆるエッセーであるが、IELTS Academic で出題される Task 1 は、与えられたグラフ、表、図解に含まれる情報を記述、要約、説明する問題である。特に、グラフが出題された場合は、グラフから読み取れるデータを記述、要約、説明しなければならない。

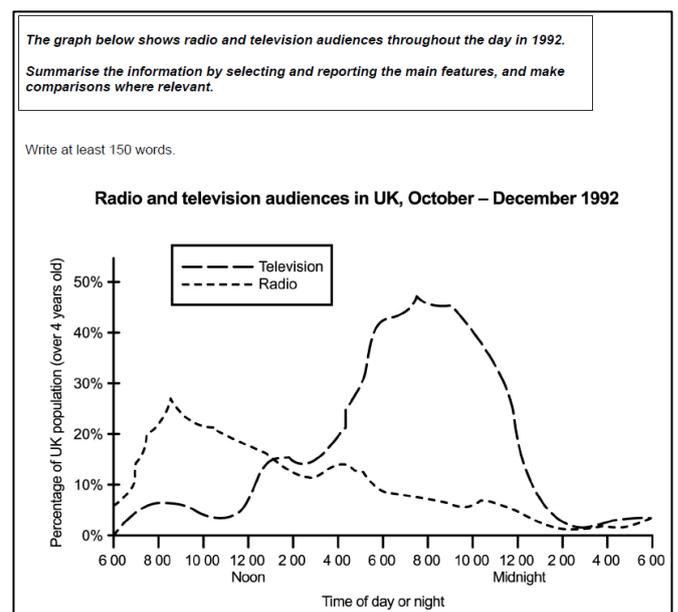


図 1 IELTS Academic Writing Task 1 のサンプル問題

図 1 のサンプル問題に対する解答例が同じサイトで公開されている。以下にその一部を引用する。このテストでは、データの動向の原因を追究する能力ではなく、データの動向を適切なことばで説明する能力を要求していることが分かる。

The bold graph shows the television audiences throughout the day. It shows that the percentage of audiences is zero percent in early morning but it gradually rises upto ten percent at 8:00 am and maintains the same for the next two hours. There is a slight fall in percentage in next two hours however after that it rises sharp upto twenty percent within the next two hours. After this the graph rises very fast and attains its peak at 10 pm which is about forty-five percent. The graph gradually falls down and at 2:00 am it is at five percent. (以下略)

この問題と解答例から、与えられた統計データを短時間で読み取って表現することは、英語圏の大学、大学院に入学するのに必要な最低限のスキルのひとつであることが推察される。なお、データに関する知識自体は要求されていないことに留意する必要がある。

### 2.2.2 リーディングスキルテスト

新井と尾崎<sup>(5)</sup>は、人工知能研究の結果として次のような懸念を表明し、読解力を科学的に測定するテストを開発している<sup>(6)</sup>。「AI と最も差別化できるはずの『よく見、よく読み、よく聞き、よく書き、よく話す能力』が教育できておらず、現代の AI に簡単に代替されるような表層的なスキルしか身に着いていないのではないかとの懸念が生じたのである。(略)そこで、われわれは中高校生が実際にどれだけ教科書に書かれている基本的な文を正しく読むことができるかを直接測定する方法を考案した。それが『リーディングスキルテスト』である」。

2016 年から 2017 年にかけて実施された同テストのうち、ある文章を読んで、その内容を正しく表す円グラフを 4 つの選択肢から選ぶ問題の正答率は非常に低調であった<sup>(7)</sup>。対象は、全国の小中高校生や大学生、社会人合わせて約 24,000 人であった。「メジャーリーグの選手のうち 28% はアメリカ合衆国以外の出身の選手であるが、その出身国を見ると、ドミニカ共和国が最も多くおよそ 35% である」という文章を読んで、その内容を表す正しい円グラフを 4 つの選択肢から選ぶ問題がある。この文章は中学社会科の教科書からの引用である。正しいグラフを選択するための情報は、与えられた文章にすべて含まれており、この文章以外の知識を知らなくとも解ける問題である。選択肢となる円グラフは、どれも 3～4 の項目からなる単純なものであり、正解と不正解は容易に区別できるはずであるが、中学生の全国平均が 12.3%、高校生の全国平均が 27.8% と低調であった。新井らの分析によると、中学生の 6 割、高校生の 5 割は、文中の「28%」を「米国出身の選手」の割合だと読み違えていた。

リーディングスキルテストの結果は、統計グラフと文章を結びつけるスキルを身につけるための教育が実施されてこなかったため、グラフを正しく解釈する能力が育てられていないことを示している。

### 2.3 大学生・社会人のための 2 種類の統計スキル

以上をまとめると、大学卒業時まで身に付けさせたい統計に関するスキルは 2 種類ある。ひとつは実際の統計データに関するスキルであり、もうひとつはデータ分析の結果を適切なことばで表現するスキルである。前者のスキルが大学生や社会人に必要な理由は、学習や仕事で使う統計データは実データであり、それを適切に扱うには、データの背景や知識をある程度必要とするからである。後者のスキルが大学生や社会人に必要な理由は、統計グラフと文章を結びつけるスキルの教育が不十分だったからである。

この 2 種類のスキルを学ぶには、公共データと白書・年次報告書を利用することが有効であると筆者は主張する。

大学生や社会人が使用する実際の統計データはその人ごとにまちまちであるから、そのための学習システムを用意するには、様々な分野の統計をほぼ無料で入手することができる公共データの活用が適している。民間企業で作成しているデータを入手するには、手続きが面倒であったり、有料であったりすることが多い点で、学習システムには使いにくい。

データ分析の結果を適切なことばで表現するスキルを獲得するには、文章を書く以前に、例となるような文章を数多く読む経験が必要であり、白書・年次報告書の活用が適している。1 冊の白書・年次報告書の本文には、公共データを引用して作成されているグラフや数表が多数掲載されており、データ分析の結果を適切なことばで表現している例文を数多く読むことができる。

そこで、以下では、公共データと白書・年次報告書を利用した学習システムの構築を目標として、実際に試みていることを述べる。

## 3. 公共データと白書・年次報告書を活用した学習システム

本章では、大学生、社会人が統計データについて学ぶことを前提として、学習システムを設計するにあたって、公共データや白書・年次報告書がどのように活用できるかについて検討する。

### 3.1 公共データの活用

大学生、社会人が統計データについて学ぶための学習システム的设计においては、公共データをうまく利用する工夫が必要である。大学生や社会人が、学習、研究、調査、仕事に必要な統計データを入手するとき、最初に候補として考えるべきは、政府や自治体が作成した公共データである。なぜなら、公共データは信頼性が高く、その多くはウェブを通じて無料で公開されているからである。

ただし、2.1 節で貿易統計の例でも述べたとおり、統計データの作られ方は、それぞれに異なっており、かつ複雑であることが多く、システム設計の段階で仕様を定めることは難しい。

さまざまな公共データを一元的に提供する試みはある。総務省統計局は、2017年5月12日に統計ダッシュボード<sup>(8)</sup>の公開を始めた。統計ダッシュボードでは、国や民間企業等が提供している主要な統計データをグラフで表示したり、統計データを検索したりするシステムである。さらに、API (Application Programming Interface) 機能も公開されており、これを利用して新たなアプリケーションを作成することも可能である。

しかし、ひとつのシステムですべての統計データを網羅することは不可能であるため、本研究で設計する学習システムでは、公共データごとにサブシステムとして追加していく方針とする。前述の統計ダッシュボードAPIでは、統計データ取得APIのほかに、データのメタ情報を取得するAPIとして、系列、地域、用語、社会事象、統計情報の5種類のAPIを提供している。これに対して、本研究では、あらかじめ仕様を定めることはせず、公共データごとに個別に対応していく。

### 3.2 学習システム内のデータベース機能

大学卒業時まで身に付けたい統計に関するスキルとして、統計データそのものを知るためのスキルと、統計データを適切にことばで表現するスキルの2種類を設定するとき、これらを同時に学ぶ教材には、公共データと白書・年次報告書がふさわしい。特に後者のスキルを向上させるには、グラフや数表を説明する文章のサンプルを、数多く学ぶことが効果的である。そこで、白書・年次報告書に掲載されているグラフあるいは数表と、それについて説明している本文中の文章

の組み合わせを、学習システム内にデータベースとして用意しておくといよい。

そこで、図2に示すようなデータベース機能を考える。白書・年次報告書から抽出したグラフと文章を蓄積する。このとき、グラフと文章の対応付けも同時に記録する。また、統計データの表現に使われる用語を抽出し、表現集として蓄積する。

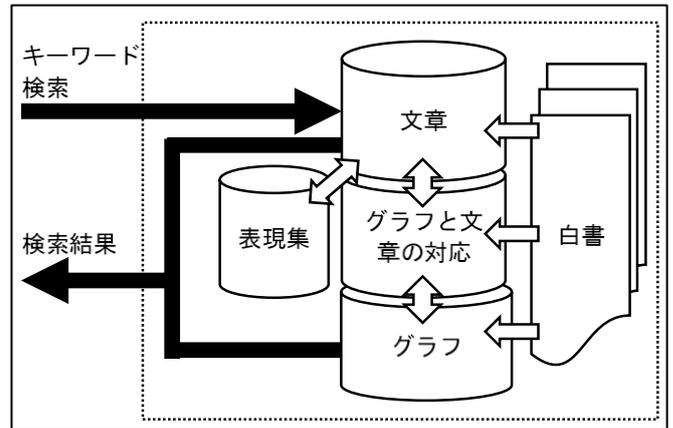


図2 学習システム内のデータベース機能

### 3.3 白書から文章とグラフの自動抽出

以下では、「平成25年度水産白書」のうち、「平成25年度水産の動向」の第II章「平成24年度以降の我が国水産の動向」を具体例として<sup>(9)</sup>、白書から、文章とグラフを自動抽出し、両者を対応付けるための手順について述べる。なお、白書からの文章とグラフの自動抽出は、プログラミング言語Pythonを用いて記述する。

#### 3.3.1 文章の抽出

白書の目次は、図3に示すように1ページのウェブサイトに掲載されており、そこから本文へのリンクが設定されている。



図3 水産白書の目次の一部

以下に示すように、図3における各ページへのこのリンク先のURLが、/j/kikaku/wpaper/h25\_h/trend/1/

t1\_2\_1\_1.html のような形式になっていることを利用し、目次のページの HTML 文書からリンク先の URL を抽出し、順次本文の HTML 文書にアクセスする。

```

<h2>目次</h2>
<p class="fw-b">
<a href="/j/kikaku/wpaper/h25_h/trend/1/t1_2_1.html">第 1
節 我が国における水産資源の管理</a></p>
<div class="section">
<p><a href="/j/kikaku/wpaper/h25_h/trend/1/t1_2_1_1.html">
(1) 我が国周辺の水産資源の状況</a></p>
<p><a href="/j/kikaku/wpaper/h25_h/trend/1/t1_2_1_2.html">
(2) 資源評価の実施体制</a></p>
<p><a href="/j/kikaku/wpaper/h25_h/trend/1/t1_2_1_3.html">
(3) 資源管理の必要性</a></p>
<p><a href="/j/kikaku/wpaper/h25_h/trend/1/t1_2_1_4.html">
(4) 我が国の漁業管理の特徴</a></p>
<p><a href="/j/kikaku/wpaper/h25_h/trend/1/t1_2_1_5.html">
(5) 我が国における資源管理の具体的実施状況</a></p>
<p><a href="/j/kikaku/wpaper/h25_h/trend/1/t1_2_1_6.html">
(6) 自主的な資源管理とそれに参加する漁業者への支援</a></p>
<p><a href="/j/kikaku/wpaper/h25_h/trend/1/t1_2_1_7.html">
(7) 資源管理の国際協力</a></p>
</div>
<br />
<p class="fw-b">
<a href="/j/kikaku/wpaper/h25_h/trend/1/t1_2_2.html">第 2
節 我が国の漁場環境をめぐる動き</a></p>

```

アクセス先の URL をリスト hrefs に集める Python 言語によるプログラムは、以下の通りである。

```

from bs4 import BeautifulSoup
soup = BeautifulSoup( ¥
    open('index.html', encoding='utf-8'), ¥
    'html.parser')
hrefs = [e.attrs["href"] for e in ¥
    soup.findAll( ¥
        href=re.compile( "^/j.+t1_._.*.html"))]

```

### 3.3.2 グラフの自動抽出

図 4 に示すような図表を含むウェブページに対応する HTML 文書は、画像のファイル名に“f\_077.gif”のように 3 桁の連番が付されていることを利用してグラフを抽出する。

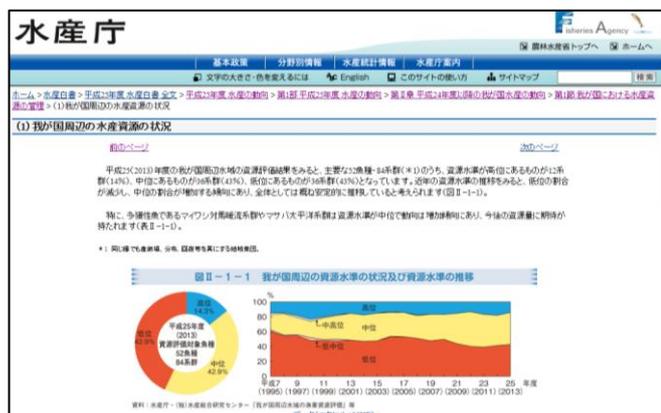


図 4 水産白書のウェブページの一部

```

<div class="ind01">平成 25 (2013) 年度の我が国周辺水域の資源評価結果をみると、主要な 52 魚種・84 系群 (*1) のうち、資源水準が高位にあるものが 12 系群 (14%)、中位にあるものが 36 系群 (43%)、低位にあるものが 36 系群 (43%) となっています。近年の資源水準の推移をみると、低位の割合が減少し、中位の割合が増加する傾向にあり、全体としては概ね安定的に推移していると考えられます (図 II-1-1)。</div>
<br />
<div class="ind01">特に、多獲性魚であるマイワシ対馬暖流系群やマサバ太平洋系群は資源水準が中位で動向は増加傾向にあり、今後の資源量に期待が持たれます (表 II-1-1)。</div>
<br />
<div class="txt_small">*1 同じ種でも産卵場、分布、回遊等を異にする地域集団。</div>
<br />
<div class="center"><br />
<a href="/j/kikaku/wpaper/h25_h/trend/1/other/f_077.xls">データ (エクセル: 29KB) </a></div>

```

白書本文を構成している複数の HTML 文書のファイル名の一覧がリスト flist に入っているとき、画像の alt 属性と src 属性のペアの一覧をリスト ilist に作成する Python 言語によるプログラムは、以下の通りである。

```

ilist = []
for f in flist:
    if re.search("^t1.*.html", f):
        html = open(f, encoding='utf-8')
        soup = BeautifulSoup( ¥
            html, 'html.parser')
        imglist = soup.findAll( ¥
            "img", [{"src":re.compile ¥
                ("/j/kikaku/wpaper/ ¥
                h25_h/trend/1/img/f_[0-9]{3}.gif")})
        ilist.extend([ ¥
            {"alt":img.attrs["alt"], ¥
            "src":img.attrs["src"]} ¥
            for img in imglist])

```

### 3.3.3 グラフと文章の対応

白書の本文は、先に示したように、class="ind01"の div 要素に記述されている。本文でグラフや数表を参照するときは、例えば、「全体としては概ね安定的に推移していると考えられます (図 II-1-1).」のように記述されている。「対象種は約 80 種に及びます(表 II-2-2, 図 II-2-3).」のように複数の図表を参照することがあることを考慮して、図表を参照している文章を集めてリスト dlist に作成する Python 言語によるプログラムは、以下の通りである。

```

dlist = []
for f in flist:
    if re.search("^t1.*.html", f):
        html = open(f, encoding='utf-8')
        soup = BeautifulSoup(html, 'html.parser')
        divlist = soup.findAll("div", {"class": "ind01"}, True, re.compile("[図表] II"))
        dlist.extend([ soup.div.get_text().strip() for div in divlist])

```

#### 4. おわりに

本稿では、大学生、社会人が統計データに関するスキルを身に付けるのに、公共データと白書・年次報告書を利用することが有効であるとし、特に、実際の白書から、文章と図表を自動的に取り出すシステムの実現可能性を議論した。このシステムの最終的な目標は、大学生、社会人のための統計スキルを育成するための学習システムの構築である。ここでいう統計スキルとは、大学卒業時まで身に付けたい統計に関するスキルのことで、本研究では、実際の統計データに関するスキルと、データ分析の結果を適切なことばで表現するスキルで2種類を考えている。これらのスキルの育成には、公共データと白書・年次報告書の活用が効果的であるとし、これらを活用した学習システムの構築を目指している。

昨今では、人工知能が作成した新聞記事が見られる。例えば、日本経済新聞社が日経電子版等で2017年1月25日に配信開始した『決算サマリー』は、人工知能が上場企業の開示資料や決算データから文章を作成し、配信するまで完全自動化されている<sup>(10)</sup>。

しかし、人工知能の進展に関わらず、大学生、社会人にとって、統計データをよく観察して特徴を見つけ出し、レポートにまとめるスキルが不要になることはないだろう。今後は、このシステム構築を進めながら、統計スキルの学習に役立つこととして、例えば、ドリル問題の自動生成や、学生が書いたレポートの評価への応用について検討したい。

#### 謝辞

本研究は JSPS 科研費 JP17H01994 の助成を受けたものです。

#### 参考文献

- (1) 文部科学省：“幼稚園教育要領，小・中学校学習指導要領等の改訂のポイント，” [http://www.mext.go.jp/a\\_menu/shotou/new-cs/\\_icsFiles/afieldfile/2017/06/16/1384662\\_2.pdf](http://www.mext.go.jp/a_menu/shotou/new-cs/_icsFiles/afieldfile/2017/06/16/1384662_2.pdf) (2017)
- (2) 日本貿易会：“日本の主な輸出入品，” JFTC キッズサイト，[http://www.jftc.or.jp/kids/kids\\_news/japan/item.html](http://www.jftc.or.jp/kids/kids_news/japan/item.html) (2018年2月6日確認)
- (3) 財務省 関税局：“財務省貿易統計，” <http://www.customs.go.jp/toukei/> (2018年2月6日確認)
- (4) IELTS 公式サイト：<https://www.ielts.org/> (2018年2月6日確認)
- (5) 新井紀子，尾崎幸謙：“デジタルイゼーション時代に求められる人材育成，” NIRA オピニオンペーパーNo.31，<http://www.nira.or.jp/pdf/opinion31.pdf> (2017)
- (6) 国立情報学研究所ニュースリリース：“文章を正確に読む力を科学的に測るテストを開発産学連携で『読解力』向上を目指す研究を加速，” [http://www.nii.ac.jp/userimg/press\\_20160726.pdf](http://www.nii.ac.jp/userimg/press_20160726.pdf) (2016)
- (7) 読売新聞 2017年9月23日朝刊 31面：“文章構成変わると誤解も 読解力テスト グラフ理解する力不足” (2017)
- (8) 総務省統計局：“統計ダッシュボード，” <https://dashboard.e-stat.go.jp/> (2018年2月6日確認)
- (9) 水産庁：“平成25年度水産白書，” [http://www.jfa.maff.go.jp/j/kikaku/wpaper/h25\\_h/](http://www.jfa.maff.go.jp/j/kikaku/wpaper/h25_h/) (2014)
- (10) 日本経済新聞社ほか：“完全自動「決算サマリー」ベータ版(試用版)，” <http://pr.nikkei.com/qreports-ai/> (2018年2月6日確認)