

突撃インタビュー

渡辺美智子先生に聞く

恒例の突撃インタビューも13回目となりました。今回は、統計教育の第一人者で、新しい教科書の編集にも携わっていらっしゃる渡辺美智子先生です。渡辺先生は、現在、東洋大学経済学部の教授として学生のご指導にあたられる一方で、統計に関する各種学会にも所属したり、放送大学で講師を務められたりと統計教育の普及にもご尽力されています。今回のインタビューでは、先生の統計教育に対する熱い思いや、新しく高校数学で統計分野が必修となったことへの期待など、いろいろとお話を伺いました。

1 プロフィール

先生の生い立ち

九州の博多が出身です。九州大学理学部数学科に在籍していました。学際大学院構想で設置された、大学院総合理工学研究科の1期生です。その情報システム学専攻というところに入りました。そこで、統計と情報学を学んだり、統計の応用では医薬関係の研究会にも所属しました。

統計を専門とされたきっかけは？

実は、私は学部時代は勉強もあまりせずに過ごしていました。だからといって、4年生になって就職活動をするわけでもなく……。ある意味引きこもり状態ですね(笑)。卒業しても就職せずに、塾でアルバイトをしながら数学の勉強をして……。なんていう考えを持っていました。

大半の同級生の就職が決まった頃、たまたま大学構内で数学科の先生に声をかけられて、「今、総合理工学研究科が第1期生を急遽求めているから、どこも決まっていらないのならそこを訪ねてみるといい」と言われました。そこで、現在の研究者としての道を拓いてくださった指導教授に出会い、統計を専門とするきっかけを得たのです。

どのようなことを？

指導教授は、塩野義製薬会社で統計解析センターの所長をされていた先生です。九大の理学部の中にも数理統計学の講座がありましたが、総合理工学研究科の情報システム学専攻の中の私たちの講座では、応用を意識した統計学の研究を行っていました。

そこでは、医療情報研究会というのが立ち上げられました。制癌剤が認可される初期の頃だったので、その臨床評価の統計分析のため、医学部のドクターの方や製薬関係・情報関係の企業の方との交流も多かったです。元来人見知りをするほうなのですが、おかげで鍛えられましたね(笑)。

2 統計の歴史

なぜ高校生や中学生に統計教育？

情報社会が謳われるようになって新聞やニュースなど身近に統計情報があふれてきたこと、またコンピュータネットワークや計測機器の進化で、カードでの購買履歴データからゲノムデータなど、データそのものがあふれる時代になって、その統計分析スキルに対して社会で大きな需要が起こっていることなどを背景に、世界中が20年ほど前から学校教育の中で統計教育の比重をあげています。逆に

日本ではご存じのように、ゆとり教育が言われた平成10年告示の指導要領で統計の内容が軽視され、実質的に中学や高校で教えられてきませんでした。



そこで、日本統計学会を始め日本アクチュアリー会やマーケティングリサーチ協会などの産業協会も含め17の学協会で、国際的なカリキュラムの比較も踏まえて統計教育推進の要望書を文部科学省に提出いたしました。ちょうどOECDのPISA調査で日本の数学の順位が下がったことなども話題になり、数学の問題のなかの4分の1は「不確実性の数理」というデータと分布を扱う統計の内容で、日本の生徒がOECD平均に比べてその部分が弱かったことなども後押しになっていると思いますが、教育行政関係者も海外では非常に早期から学校教育のなかの算数・数学の教科で、系統的な統計教育がなされていることに気が付かれたのではないのでしょうか。高等教育における数学研究の世界でも、文部科学省の科学技術政策研究所が「忘れられた科学・数学」を公表し、世界の数学研究と日本の数学研究とのギャップが大きいことを指摘し、産業界の現代的ニーズに応える応用数学研究の推進に力を入れ始めました。応用数学の中での統計の比重は、データ爆発と言われる現在、非常に大きくなっています。世界の学校教育で統計教育改革が進められる背景は、社会の中で数学・統計を活用しイノベーションを起こす人材育成を長期的に計画しているためなの

です。

海外で統計教育が推進されるきっかけは？

1990年代初めにアメリカで統計教育改革が始まっています。それは1980年代後半に顕在化したアメリカの経済危機と、逆に戦後の経済復興を果たし、一躍経済大国に押しあがった日本が背景にあります。アメリカは、当時、”Japan as No.1”と呼ばれる高品質な製品を生み出す製造業の企業内教育に着目し、統計的問題解決の研修、いわゆる、トヨタの「KAIZEN」とよばれる統計的品質管理教育をMITなどでもかなり研究し、積極的に、企業教育や大学教育、そして学校教育内で体系化したわけです。

日本の統計的品質管理は、アメリカの統計学者であるデミング博士が戦後GHQにより使節団として来日した際に、企業関係者に指導したことで始まっています。日本の製造業界ではデミング博士は神様のように言われている学者で、1980年代後半、アメリカでは「日本に品質を指導した学者」として非常に有名になりました。デミング博士が1980年代からアメリカで統計的品質管理を指導した際には、統計的問題解決の考え方をモノづくりだけではなく、金融・経営・行政そして教育の「KAIZEN」の手法として、国民すべてが身に付けるべき「生活を向上するスキル」として位置付けています。当時のアメリカ政府レポート、1992年のスキャンズレポートに、「デミング博士が導いた日本の成功にならい、来るべき21世紀のアメリカのハイパフォーマンスを成し遂げる」と書いてあり、「まずそのための産業界と教育界の連携による教育改革を行う」ことが明記されていて、統計教育が国益と直結することがひろく認識されています。これはすぐに、イギリスを始めとするヨーロッパや韓国・中国などのアジア諸国にも広がっています。

1990年代の日本には、統計教育推進校というものが最盛期で145校ほどありました。日本

の研究をしていた MIT のレポートには「日本の小中学校では、企業教育と同様に学校教育でも問題解決型の統計教育を盛んに行っている」と報告され、そのことも世界の統計教育改革のきっかけになっています。しかし、世界が統計教育を推進している間に、本家の日本がどんどん統計教育を学校教育から縮小していったことは、時代のニーズとも逆行し非常に皮肉な結果を生んでいるわけです。

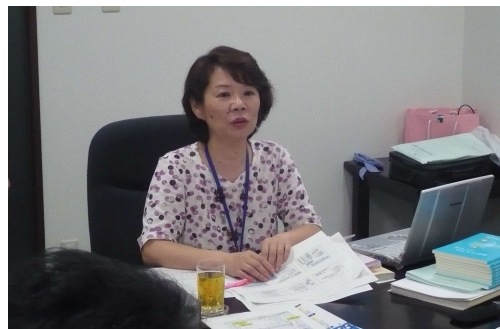
3 統計の重要性

何のための統計か

社会での統計的問題解決の枠組みってというのは、難しいことではなく、身近などんな問題にでも、データで客観的に判断し、解決の要因を探る、というようなものです。問題の前に、課題、issue があります。例えば、簡単にいえば、男の子だと女の子にモテないとかモテないなあというのが課題、企業だと売り上げが伸びないなあとか利益があがらないなあというのが課題、国だと景気が悪いなあとかが課題になります。ここで、この課題に対して科学的に対処するとは、それをデータで解ける問題 problem に変えなければいけません。そのためには、課題の成果（アウトカム）や目的（ターゲット）となる評価指標 Y を作らないとダメなんです。「女の子にモテない」というんだったら、「もてるかもてないか」について、何を Y で測るかっていうことを考える。そこで、ひとまず、女の子から一日に来るメールの回数を Y としてみよう…ということを考えてみると、Y が人によって分布するわけです。そして、マクロな視点を持って、全体の中で自分の現状を客観的に評価する。そのために、できるだけデータを取って全体の分布を踏まえて自分の相対位置を確認し、自分が標準的な状態なのか、どちらかの方向に偏っている

のかを判断する必要があります。そのときに、平均値しか知らなかったら平均より上か下かの判断しかできません。そこで、標準偏差が必要になってくる。標準偏差というのは概念がとても大事で、公式や計算の仕方よりも分布を取り扱う数理的な指標として、その意味の理解と活用を知っておくことが大切だと思います。

問題解決では Y の分布を確認した後に、では Y の分布を変化させる要因 X は何なのかを探索することで、女の子からのメールの回数を増やす戦略のヒントが生まれるわけです。そこで、新課程の数学 I の「データの分析」の内容の箱ひげ図のように要因 X が質的属性の場合に、並べて Y の分布を比較するツールや要因 X が量的属性の場合に X と Y の関連をみる散布図・相関係数の手法が必要になってきます。



社会でいう「KAIZEN」のヒントを掴む…というのは、非常に簡単にいえばこのようなことで、国際的な統計教育の方法として、まず、問題解決のフレームを教え、個々のグラフや統計量がどういう役割をもつのかを教えるというのが現在の主流です。計算やグラフ作成はコンピュータソフトがしてくれる現在、活用の枠組みと結果の解釈こそ求められる力量になっているわけです。

日本国内では

先日、女子サッカーのなでしこジャパンの

世界一の話題がありました。サッカーの世界でも、11人+11人+ボールの計23個のオブジェクトがどう動いたかのデータを取り、選手やボールがどこをどう動いたかを集計・分析して、戦略を立てる。コンビニのレジやレストランのオーダーをとる端末は、そのまま顧客の消費行動データとして分析され、つねに最適な品揃え、メニュー構成が考えられている。社会には、データを集計して分析結果を解釈し戦略を立てられる人、データアナリスト、データマイナー、モデラーなどいろいろな名称がありますが、そういう人たちが圧倒的に必要とされる時代が来ています。

医学の分野で

統計の応用領域は非常に多いのですが、医学部の大学入試や医師の国家試験に統計の問題が出るくらい、医学の分野での統計は重要です。最初に話した大学院で医療情報研究会に入る前は、薬は飲めば治るもの、みんなに効くものと思っていたのですが、その研究会で、効く人もいれば、効かない人もいる、薬の効果のデータは人によるばらつきが非常に大きいことを知りました。その中である薬は「効く」という判断を下すためには、非常に厳密な統計の知識が要求されます。日本では医薬統計家も不足しているのが現状です。

4 科学的探究のために

本来のねらいは

90年代以降、統計教育のカリキュラムの改革を世界中が行いました。そこでは、問題解決のサイクルの枠組みやその中で使う手法を教えて、概念教育を行う。解決の枠組みをX、Yで考えて、コントロールできないYをコントロールできるXでどう改善していくのか。Yの現状分析だけの分布の教育は現状分析、X

とYで考えて、初めてその効果や関連の強さとかがいろんな候補で見つけ出せるから、分析になる。

ここで何が科学的探究かという点、実は関係があるとか相関があるということと、因果があるということは別なんです。因果のエビデンス (evidence) として、データを分析するためには、ここで数学が必要になってきますが、これはものすごく難しい。いまだに諸種の研究が続いているくらい難しく、それで学校教育では数学の中でその数理をどう教えるのかということが大事になってきます。海外の大学教育では、統計学科は数学科とは独立に存在していますが、高校までは、きちんと数学の教師が統計教育を担う責任があるわけです。単純に比較して分布に差が出ることや散布図から相関関係がみえることのみで、因果関係を立証したことにはならないことを知る必要があります。また、それだけではなく、それが因果のエビデンスとなるためにはどのような条件が必要なのか、どうすれば因果のエビデンスに近づけるのか… そこまでを海外の教育では、しっかりとその学年に応じた身近な例で、考えさせています。



PISA 調査の科学リテラシーの問題でも温暖化とCO₂排出量の問題が出ましたよね。太郎さんは「CO₂が温暖化の原因だ」と因果関係と結びつけているのに対し、花子さんは「CO₂以外にも関係している要因 (confounder: 交絡要因) があるんじゃないか」と

言う。交絡要因というような難しい言葉を使わずに、「それは公平な比較ですか」とか「他に一緒に連動している要因、隠れているものがあるのでは？」というような問いかけを、海外では小学校からデータ分析の結果を議論させるときにしています。

隠れた要因を探すために

Y を動かす要因 X や X 以外の隠れた要因を考えるために、「特性要因図」(図 1)¹⁾ を使います。海外では「cause and effect diagram」「fishbone diagram」「Ishikawa diagram」として知られています。

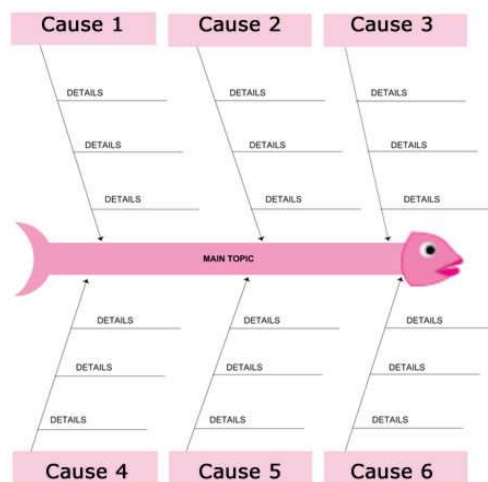


図 1: 特性要因図

こういうものは数学ではないと言われる人もいますが、関数を現実の問題で活用しようと思えば必ず必要になる道具です。自分の考えている問題で、どちらが要因 (independent variable: 原因変数) で、どちらが結果 (dependent variable: 結果変数) であるかというのは、先にきちんと意識させておかないと、関数でモデルを作ることができません。数学的

モデリング、統計的なモデリングの思考の道具になることから、海外では小学校からイラストのようなダイアグラムで教えていますが、これも日本企業が統計的品質管理教育で使用しているのを見習ってのことです。

統計の手法を問題解決の道具として位置付け、実際のデータで活用を教えることが新課程の「データの分析」に意図されていますが、そのためには、ランダムサンプリング（無作為抽出）やランダムアロケーション（無作為割付）のようなデータの取り方が結果の解釈に影響を与えることや、デザインしてデータが取れる実験データとデザインが効かない観察データの違いなども、知っておくといいのではないかと思います。

今海外と日本では

海外の大学では、統計学部や統計学科がたくさんあります。そのため、その卒業生は、Statistician（統計家）という職業区分に位置付けられています。

例えば、イギリスでは王立統計学会による Statistician の資格認定制度や中国では国家資格（統計師）にもなっています。アメリカでは、政府の職業区分の中で、Statistician の仕事内容、就職先、平均給与、人数などがきちんと公表されていて、最近の USA Today 紙に「Statistician の給料が、平均で 26 % あがった」ということや Fortune 紙に、「アメリカの失業率は高くても Data Scientist だけは別で、企業は自社の中のデータを意味のある戦略に繋げるために統計家を雇い続けている」、New York Times 紙には「Google のチーフエコノミストがこの先 10 年の most sexy job は、Statistician だと言っている」などの報道がなされています。

日本では、世界でもめずらしく、大学に統計学部や統計学科がひとつもないので、統計家という職業自身が認知されていませんが、事情は何ら変わることはありません。

¹⁾ 「Creately K-12 Education Graphic Organizer Templates」
<http://creately.com/examples/Graphic-Organizer-Templates>

今回、学習指導要領が改訂され「データの分析」が入ったことを、経団連の方や日本品質管理学会の企業の方々はとっても期待しています。日本の経済が停滞し企業が教育にお金と時間を割けなくなっていて、学校教育に期待しているわけです。

数学の先生方に望むこと

データの読み方、分析の仕方の考え方をもっと小さいときから、ちゃんと教えるということが統計教育なのですが、数学だけではなく、科学教育だったり、情報教育だったり、全部でやっているのが海外です。ただし、数学の中で体系的に毎学年、教えられています。

日本の数学の先生からは、「そこまで数学で面倒を見なければならぬのか」などいろいろな質問が来ます。いままでの抽象的、演繹的推論の方式の数学とは異質な、文脈（コンテキスト）の中で機能的推論の方式を教えるのが統計ですから、戸惑われるのももっともだと思います。日本の大学で、統計教育、とくにデータの分析教育自身を受けてきたという方の方が少ない現実があります。しかし、生徒が日本国内で就職したとしても日本企業のグローバル化が進んでいますので、国際通用性の観点からも、またデータがどの組織にも身近に蓄積されている社会情勢から考えても、この方向性は拡大することがあっても後退することはないと思います。その重要性に鑑み、是非、関心をもって取り組んでいただきたいと思います。

5 統計指導における教材例

スポーツのデータ

不確実性の数理を教える際の面白い教材で私が使うのが、野球の球種と球速のデータです。スポーツのデータは、背景が分かり易く

生徒にも関心を持ってもらえるのではないのでしょうか？ 私が大学で使うのは、松坂大輔が2005年のシーズン中に投げた3329球の投球内容のデータで、「データで学ぶ！統計活用授業のための教材サイト」²⁾からダウンロードできます。

まず、データの種類として球種のような質的データと球速のような量的データの2種類の区別を教えます。松坂が投げるまでは、バッターはどの球種で投げているのか、何キロで投げているかはわからない。つまり予測がつかない。これを不確実な現象という。その現象に対して「わからないのだから、何もしない」ではなく、「わからないけれど、科学的に不確実性にチャレンジしていこう」というのが、統計的な思考力になる。

そして、不確実な現象を過去のデータのばらつきで捉えます。2005年のシーズン中3329球の様子を見て、それを分布という概念で捉えると、球種の分布構造は、ストレートが47%、スライダーが27%あわせて約70%ということがわかります。6種類の球種(図2)とその可能性がわかり、その程度の差をみて予測の根拠とする、これが分布を活用する統計的思考です。分布の山と裾(メリハリ)を予測にいかす考え方です。

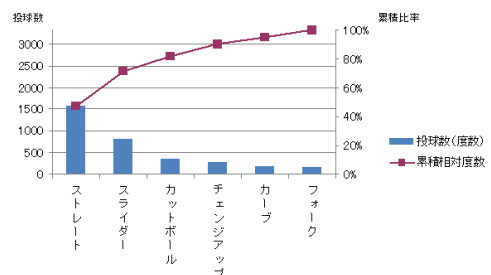


図 2: 球種のパレート図

量的データの球速も一緒に、データの分布をみることで、起こり得る球速の範囲がわかり、

²⁾<http://stat.sci.kagoshima-u.ac.jp/data/>

球種ごとに分布の山と裾をみることで、「ストレートだったら143~147km くらいの範囲に7割は入ってくる」とか「150Km 以上は、全体のわずか数%しかない」などと考えることで、予想に活かします(図3)³⁾。

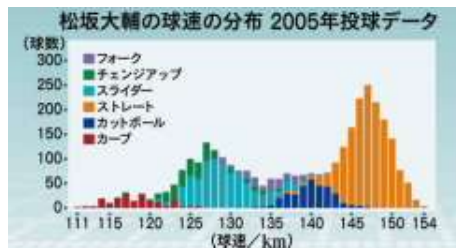


図 3: 球種・球速の層別ヒストグラム

この背景には、多峰性を示す球速(量的データ)の分布が適切な分類(球種別)によって単峰性の分布にできることがあります。

「データの分析」で学習される平均、標準偏差、四分位数などの統計量は、いずれも中心傾向を決めるためのものですが、そのためには分布を単峰性にして、中心傾向を一つにしておくことが大切です。多峰性の分布で計算しても意味のない統計量です。そのこともこの例では示すことができます。球種に分けると、松坂の球速の場合は、単峰で左右対称な分布になり、このような分布の特徴は、平均を中心に±標準偏差で、データの中心付近の2/3を含む球速の区間(中心傾向)がわかるので、平均や標準偏差の有用性を理解させる良い事例だと思います(図4)。



図 4: 球種別の球速データ

³⁾データソース: (株) データスタジアム

また、標準偏差や四分位範囲が小さいほど、中心傾向を狭い区間で捉える事ができ、予測の精度やマネジメントの効率が高くなることも感覚的に理解させておくといいと思います。松坂の球速のデータは、「データで学ぶ統計活用のための教材サイト」の”スポーツ”のところから利用できます。

利用規定は、授業のみの利用です。無断で本や教科書の材料にすることはできません。

また、相関関係の教材には、ワールドカップのオフィシャルサイト⁴⁾のデータを使ったりもします。どのプレーが勝ち点に結びついているのかという問題意識を持たせ、散布図や相関係数が役立つことを実感させます。

紙ヘリコプター

これは、MIT が日本研究をやって、企業研修用に作ったのがこの問題解決のための紙ヘリコプターです(図5)⁵⁾。現在、学校教育でも企業教育でも、難易度をいくらでも変えられるので、使用されています。どのレベルでも出来る教材で、ものすごく盛り上がるらしい。

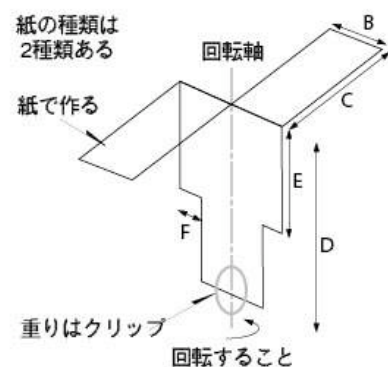


図 5: 紙ヘリコプター

⁴⁾<http://www.fifa.com/worldcup/archive/southafrica2010/statistics/teams/goals.html>

⁵⁾「Tech-On! 日経ものづくり 2006 年 1 月号」
<http://techon.nikkeibp.co.jp/article/HONSHI/20051227/111976/>

紙を裂いて、ある高さから落とすだけでもいいのです。何度も繰り返すと、ストップウォッチの 0.1 秒レベルで間違いなくばらつきが出ます。そこで、単純に最も滞空時間を長くするデザインを探したりとか、予め決めた時間で落ちるものを作らせたり、落ちる場所を特定してそこに落とすとか、目的を変えることで難しくなったり易しくなったりします。

アメリカの NCTM の教材で、ワンクリップとかツークリップとか、紙飛行機にクリップをつけて飛ばす教材例もありますし。

ウェブサイトでは

「科学の道具箱」(図 6)⁶⁾ というサイトには、実際に専門家の方や実務家の方がそれぞれの領域でデータが活用される分析のストーリーを話してくれていて、それに関連するデータやナレーション入りの統計の解説などがあります。新課程の指導要領の中の統計教育のために、制作されました。(並列) 箱ひげ図やラベル付き散布図などが簡単に描ける Excel のマクロを利用したソフトもダウンロードして使えます。



図 6: 科学の道具箱

⁶⁾<http://rikanet2.jst.go.jp/contents/cp0530/start.html>

6 最後に

統計は大学のどの学部でもそこでの専門科目を理解する上で必要になってくる内容です。統計関連学会連合では、「統計学分野の教育課程編成上の参照基準」⁷⁾として、大学で統計がどのレベルまで必要になるのかのガイドラインをまとめました。どの学部に進学するにせよ、統計的データ分析の下地は必須であることがわかっていただけるかと思います。

共通必修科目となった数学Ⅰの中に「データの分析」として統計の内容が入ってきたことで、文系・理系を問わず、この内容が大学入試と関わってくることになります。大学入試センターからは、試作問題も出されるでしょう。

統計検定センター⁸⁾では、日本統計学会公式認定の資格として「統計検定」を始めました。推薦入試の参考情報や大学入試の傾向を知る上で参考になるかと思います。統計教育は、生徒の将来の就業力や研究力・身近なリスク管理能力、総合的問題解決力の育成に繋がる大切な内容です。統計関係の学会や統計数理研究所、総務省統計研修所では、学校現場での統計教育の実践や定着、先生方の研修などをできるだけ支援することにしておりま

すので、どうぞよろしくお願いいたします。



⁷⁾<http://park.itc.u-tokyo.ac.jp/atstat/statedu/shitsu/shitsu.html>

⁸⁾<http://www.toukei-kentei.jp/>