

日本における統計教育～課題と解決のアイデア～

(研究会実施日:2009年6月17日)

講師: ^{いわさき}岩崎 ^{まなぶ}学 [成蹊大学理工学部教授]

研究領域

統計的データ解析, 特にコンピュータ利用の多変量解析の方法論の研究, ならびに教育分野, 医薬分野などでのデータ解析の実践

経歴 所属学会

茨城大学工学部助手, 防衛大学校数学物理学教室講師・助教授, 成蹊大学工学部経営・情報工学科助教授, 教授を経て, 現在, 成蹊大学理工学部情報科学科教授。
日本統計学会(理事長, 評議員), 日本行動計量学会(理事, 編集委員), 日本計量生物学会(評議員), 日本計算機統計学会(評議員, 編集委員), 応用統計学会(評議員), 日本品質管理学会(代議員), American Statistical Association など。

著書

- ・岩崎 学 (著) (2007). 確率・統計の基礎 東京図書
- ・岩崎 学 (著) (2006). 統計的データ解析入門 ノンパラメトリック法 東京図書
- ・岩崎 学 (著) (2006). 統計的データ解析入門 実験計画法 東京図書
- ・岩崎 学 (著) (2006). 統計的データ解析入門 単回帰分析 東京図書
- ・岩崎 学・吉田清隆 (著) (2006). 統計的データ解析入門 線形代数 東京図書
- ・岩崎 学 (著) (2004). 統計的データ解析のための数値計算法入門 朝倉書店
- ・岩崎 学・中西寛子・時岡規夫 (著) (2004). 実用統計用語事典 オーム社
- ・岩崎 学 (著) (2002). 不完全データの統計解析 エコノミスト社
- ・柳井晴夫・岡太彬訓・繁榊算男・高木廣文・岩崎 学 (編著) (2007). 多変量解析実例ハンドブック 朝倉書店
- ・岩崎 学 (著) (2000). 統計的データ解析のレシピ 日本評論社
- ・岩崎 学 (著) (1994). 混合実験の計画と解析 サイエンティスト社

要 約

小中学校ならびに高等学校の学習指導要領が改訂され, 統計の内容が算数・数学の中にこれまで以上に盛り込まれるようになった。本論ではそれに関連して, 米国の高大接続プログラムの AP Statistics を取り上げ, 実際に筆者が AP Statistics に準拠した授業を実施した経験を基に, 高等学校および大学初年級でのあるべき統計教育の姿を論じる。また, 実際にその種の内容を教える際に遭遇するであろう困難さについても言及する。

I. はじめに

小中学校ならびに高等学校の学習指導要領が改訂され、統計の内容が算数・数学の中にこれまで以上に盛り込まれるようになった。統計学を生業とするものにとっては誠に嬉しい限りであるが、その反面、教育上の成果を達成するための責任も負うことになった。筆者は長らく大学における研究・教育に従事し小中学校における教育に疎いこともあり、ここでは主として高等学校および大学初年級の初等統計の教育的なことを絞って議論したい。

統計教育の議論は日本でもそして諸外国でもこれまで長く行われてきているし、それは終わることはない永遠のテーマである。しかし近年、特に英米において、コンピュータならびにネットワーク環境の飛躍的な発展に触発された大規模なデータの蓄積とその分析への需要の広がりが高まり、そして学生気質の変化も相俟って、統計教育の中身を替えなければならないという議論が、統計の専門誌の中でも多く見られるようになってきている¹。また、統計の方法論でなく教育方法に関する書籍の刊行も見られる²。

そのような中、筆者は AP Statistics なるものの存在を知った。AP Statistics は米国 College Board が提供する高大接続プログラムで、その中身を調べるうちに、それが日本の統計教育を見直す意味で絶好の材料であると考えるに至った。そしてその教育内容ならびにその効果の理解のためには自らの実践が第一であることから（統計学では実際にデータを取る作業が重要）、2009 年度の前期のお茶の水女子大学での講義「応用統計学」で半年間実際に AP Statistics に準拠した授業を実施した。その授業経験により、統計学の内容を高等学校の数学の科目の中で教えたり、あるいは大学初年級での初等統計学として講義したりする上での難点がいくつか見えてきた。その一つが統計手法のあいまいさである。

本稿では、次の第2節で AP Statistics の概要とその特徴について述べ、第3節では統計

教育で支障となると思われる上述の統計手法のあいまいさについて議論する。最後の第4節は簡単なまとめとする。

II. AP Statistics とは

AP は Advanced Placement の略であり、米国 College Board が 1955 年に創始した高等学校と大学を結ぶ高大接続プログラムである。高等学校において、正規の授業以外に AP コースを設け、そこでは大学初年級レベルの内容の授業を行う。AP コースは主として高等学校の最後の学年に配置され、毎年5月に全国一斉試験 (AP Exam) が実施される。その試験でよい成績を収めると（概ね5段階の4以上）、大学入学後の当該科目の単位として認められる³。

AP Statistics は 1977 に開始された比較的新しい AP コースである。しかしその人気は高く、近年では 11 万人以上の高校生がそのコースを取っているとのことである。その理由は、大学入学後にどのような学問分野を専攻するにせよ、統計的方法論がその基礎にあるとの認識が広く持たれ、実際ほとんどすべての大学ではその専攻に関わらず、統計学入門の科目がほぼ必修の形で展開されていることから、それを高等学校段階で先取りすることのメリットがあることによるようである。

AP Statistics コースの先取りにより大学での統計学入門の単位に振り替えられるということは、AP Statistics が米国の大学で展開されている入門的な統計学の授業内容を反映していることに他ならず、その意味で AP Statistics の中身の精査は米国における統計学入門を知ることにつながる。AP Statistics に関しては、日本での受験参考書と同様いくつかの参考書・問題集が米国で出版されている（筆者は Hinders (2008), Mulekar (2008), Simmons, Bland and Wojciechowski (2008) および Sternstein (2007) を入手し、主として Sternstein (2007) を参考にして授業を行った）。Sternstein (2007) は出版社の名をとって Barron's と呼ばれることが多いので、以降本

¹ Brown and Kass (2009), Meng (2009), Smith and Staetsky (2007) などを参照

² Gelman and Nolan (2002), Hulsizer and Woolf (2009) など

³ Roberts, Scheaffer and Watkins (1999) および最新の情報については College Board のホームページを参照

稿でも当該書を **Barron's** と呼ぶことにする。

AP Statistics は、高校生対象であること、ならびに大学入学後の専攻を問わないことから、数学的なものとはなっていない (**non calculus** といわれる)。そしてその内容は、大きく分けて

- (i) データの集計とグラフ化
- (ii) 標本調査と実験の方法
- (iii) 確率と確率分布
- (iv) 統計的推測

の4分野からなっている。日本の大学の通常のカリキュラムでは (iii) の確率と (iv) の統計的推測を半年ないしは1年かけて授業するのが普通であろう。あるいは文科系の学部であれば、コンピュータの操作を含めた情報教育の一環として (i) が講義されることもある。 **AP Statistics** で特徴的であり、筆者も驚いたのは (ii) のデータ取得の方法論が授業内容に含まれている点である。

統計的データ解析でも最も重要なのはよいデータを取得するための方法論であり、その意味で実験計画法と標本調査法は欠かせないのであるが、日本の統計学入門の範囲に含まれることは極めて稀であろうし、実際筆者も学部段階で講義したことはなかった。また、実験の方法といった場合、日本では品質管理分野での実験計画を思い浮かべるが、 **AP Statistics** での実験は主として新薬開発の臨床試験のような実験研究を指している。そして、実験研究と疫学調査のような観察研究との相違を生徒に理解させようとしている。健康科学が大学における重要な専門分野であることが色濃く反映されていると感じられた。

AP Statistics コースの特徴は、座学だけではなく、生徒が自らデータ取得の計画を立て、実際にデータを取った上でその分析を行い、そして結果を解釈するという実践的なプロジェクトが組み込まれ、それが重要視されているという点である。高等学校では大学初年級のような大規模授業はありえず、きめ細かな指導ができることからむしろ大学に入ってからよりも密度の濃い教育が可能な点が **AP** プログラムが人気となる所以ではないとも考えられる。残念ながら筆者のお茶の水女子大学での授業ではそのようなプロジェクトを実

施する時間的（および人的）余裕がなく、残念であった。

AP コース終了後の統一試験 (**AP Exam**) では、5者択一式問題が40問（90分）、論述式問題が6問（90分）出題される。択一式の問題の採点は機械でできるが、論述式問題の採点はそうは行かない。全米各地から大学の統計学者および高等学校の **AP** コース担当者が1箇所集まり、泊り込みで集中的に採点することによって、先日の **Amstat News**（米国統計学会のニュースレター）にも採点ボランティア募集の広告が出ていた。極めて大変な仕事であるが、参加者からは合宿気分で大米の仲間と情報交換するよい機会であって楽しんでいるとの感想もあった。日本でもこの種の集まりが開けるものであろうか、と気になった。大学の教員も高等学校の教員も何かとむやみに忙しいのが日本の現状である。であるが故に、リフレッシュの期間が必要であるかもしれない。

AP Statistics で何を教えるべきか（あるいは何を教えないか）の議論は相当あったようで、その開始時期は当初予定からずれ込んだようである (**Roberts, et al., 1999**)。カリキュラムの統一の陰ではこの種の突っ込んだ議論が不可欠である。

最近、**AP Statistics** を受講して大学に入学した学生に関する興味深い調査報告が公表された (**Patterson, 2009**)。それによると、**AP Statistics** コースの受講により統計学に対する興味が湧いたと答えた学生がいる一方で、**AP Statistics** で単位を振替取得した後にさらに進んだ統計学のコースを取らなかった学生がいたり、**AP Statistics** により統計学に失望したという学生も（幸いなことに）少数いたようである。**Meng (2009)** も大学初年級の統計学入門の授業の重要性を指摘する論文の中で '**Among the Harvard undergraduates I asked, the most frequent reason for not considering a statistical major was a "turn-off" experience from an AP statistics course**' と書いている。しかしながら、総体として **AP Statistics** は成功を収めているとの評価が一般的である。米国では、高等学校間の格差も大きいようで、**AP** コースを設けていることがその学校のステ-

タスの一つとなっている。高等学校の教員も、大学レベルのコースを担っているという自負を持っているようである。日本の高等学校にも、大学受験のための勉強に時間を費やすのではなく(決して無駄というつもりはないが)、AP コースのような大学での勉強を先取りするようなシステムがあってもよいであろう。これまでも高大連携プログラムが実施されているというもののそれは単発的かつ非組織的なものでしかない。

Ⅲ. 統計手法のあいまいさ

統計手法の中には、統計学者間でさえ(であるが故に)合意できていない問題が相当数あいまいなまま残されている。統計が数学の中で教えられる場合、初等中等レベルの数学ではほとんどあいまいさが見られないことから、統計手法のあいまいさは教師も生徒も混乱させる原因となろう。もちろん、現実問題では解が一意に決まるというほうが珍しく、何がしかのあいまいさを内包したままに意思決定がなされることから、その種の訓練を早いうちから積んだほうがいいとの意見もあろうが、やはり諸刃の刃であり、下手をすると統計学に対する不信を招きかねない。ましてや、試験に出題するような問題では微塵のあいまいさも許されないと想像される。ここでは筆者が AP Statistics の内容を実際授業するに当たって遭遇したあいまいさの例をいくつか挙げる。

(1) 標準偏差と標準誤差

母集団が N 個の個体 x_1, \dots, x_N からなり、全体の平均値(母平均)を μ とするとき、(母)標準偏差は

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (1)$$

で定義され、母集団からの独立な n 個の観測値を x_1, \dots, x_n とするとき、その(標本)標準偏差は、 \bar{x} を標本平均としたとき、

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

と計算される。取り敢えずここで混乱が生じ

る。標本標準偏差は第2節の AP Statistics の最初の (i) の段階で導入される。(2) で除数を $n-1$ とする理由はこの時点では(あるいは最後まで)謎のままである。統計学的には、(2) で母平均 μ が既知であれば(標本)標準偏差は

$$s^* = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2} \quad (3)$$

で計算されることになるが、新たに (3) などを導入するとさらに混乱が増すばかりであろう。実際筆者の授業でも (1) と (2) の違いについて質問を受けたが、不偏性は出さないまでも等式

$$\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2$$

を用いて (2) が過小評価である旨を説明しようとしたが理解されたとは思えなかった。除数を n とするか $n-1$ とするかで計算結果は当然異なるので、「 n が大きければどちらでもよい」などというのは統計的には極めて全うな回答であっても、計算間違いはご法度の数学の中では許されないであろう。

次に、母標準偏差を σ としたとき、 n 個の観測値(正確には確率変数)の標本平均 \bar{x} の分散の平方根

$$\frac{\sigma}{\sqrt{n}} \quad (4)$$

は標準誤差 (standard error) と呼ばれる(と筆者は思っている)。それに対し、(2) の標本標準偏差(母標準偏差の推定値)を用いた

$$\frac{s}{\sqrt{n}} \quad (5)$$

は「(4) の標準誤差の推定値」である。しかしいちいちそう呼ぶのは面倒なので (5) を標準誤差と言い習わしている(と筆者は説明している)。ところが少なくとも Barron's では、(4) はあくまでも「母集団標準偏差 (σ) を標本サイズの平方根 (\sqrt{n}) で割ったもの」であり、(5) を標準誤差と呼んでいる。少なくとも言葉の定義は統一させておかねばならない。見かけ上は (4) は (1) の σ を \sqrt{n} で割ったものに過ぎないが、標準偏差と標準誤差の区別は統計的推測を理解するうえでの大きな

キーポイントであることは強調せねばならない。とはいえそこが、理解されることは少ないのが悩みの種である。数式の表現のみを見てその意味を見ようとしないう学生が多い。

(2) 中心極限定理と二項確率の信頼区間

試行回数 n 、確率 p の二項分布 $B(n, p)$ は n が大きいとき同じ期待値と分散を持つ正規分布 $N(np, np(1-p))$ で近似される。これは中心極限定理からの帰結であり、二項確率の統計的推測では最も基本的な性質である。具体的には、 $X \sim B(n, p)$ とし、 $Y \sim N(np, np(1-p))$ としたとき、近似的に

$$\Pr(X \leq a) \approx \Pr(Y \leq a) \quad (6)$$

であり、連続修正を施して

$$\Pr(X \leq a) \approx \Pr(Y \leq a + 0.5) \quad (7)$$

とするとさらに近似はよくなる。

ところがこの「近似される」が数学では曲者である。Barron's では、正規近似が有効であるための条件として

$$np > 10 \text{ および } n(1-p) > 10 \quad (8)$$

としていて、その条件が成り立つかどうかを必ずチェックさせている。しかしこの条件は書物によって一定しているとは言い難い。実際、Barron's には 10 ではなく 5 としている著者もいると書いてあるし、 $np(1-p) > 5$ としている書物もある。問題文の条件が成り立つかどうかのチェックは数学では基本であるが、その条件自体があいまいでは、生徒は混乱するであろう。

また、確率計算に関しては Barron's では連続修正を施した (7) を一貫して用いているが、初等統計の書物では (6) としているのも少なくない。逆向きの不等式では連続修正は $\Pr(a - 0.5 \leq Y)$ となるが、多分生徒は混乱するであろう。現に大学の授業でも混乱している。これも式の意味を理解しないことに起因している。

二項確率 p の信頼区間の構成は、簡単なようであるが実は一筋縄ではいかない。通常の本物では (Barron's でも) p の近似的な $100(1-\alpha)\%$ 信頼区間は、その点推定値を $\hat{p} = x/n$ とし、 $z(\alpha/2)$ を標準正規分布の上側 $100\alpha/2\%$ 点として

$$\hat{p} \pm z(\alpha/2) \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \quad (9)$$

としている。このときの正規近似が成り立つための条件もあいまいで、Barron's では $n\hat{p} > 10$ および $n(1-\hat{p}) > 10$ としているがこの条件は (8) と同じではない。Agresti and Franklin (2007) では $n\hat{p} > 15$ および $n(1-\hat{p}) > 15$ としている。(9) の信頼区間はスコア型と呼ばれるもので、真値の被覆確率が名目の $100(1-\alpha)\%$ を下回るため統計学者の間では評判がよくない。Agresti and Franklin (2007) でも (9) を提示しているが、一方で世界的に著名な統計学者である A. Agresti は Agresti and Coull (1998) において別の手順を推奨している。A. Agresti も初等統計の書物を書きながら疑問に思ったに違いない。岩崎 (2007) も (9) は紹介しつつもそれは第二選択であるとしている。また、Barron's では (9) の導出過程が示されていないが、その過程においては (6) の連続修正なしの正規近似を使用していることになり、確率計算での一貫しての (7) の使用との整合性が取れない、という問題もある。

(3) 正規分布の母平均の差の信頼区間

正規母集団 $N(\mu_1, \sigma_1^2)$ および $N(\mu_2, \sigma_2^2)$ からのそれぞれ独立な m 個および n 個の観測値の標本平均をそれぞれ \bar{x} , \bar{y} とするとき、差 $\bar{x} - \bar{y}$ の標準誤差 (SE) は

$$SE[\bar{x} - \bar{y}] = \sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \quad (10)$$

である。母平均の差 $\mu_1 - \mu_2$ の点推定値は当然 $\bar{x} - \bar{y}$ であり、その標準誤差 (の推定値) は σ_1^2 および σ_2^2 の推定値を各標本分散 s_1^2 および s_2^2 としたとき、

$$SE = \sqrt{\frac{s_1^2}{m} + \frac{s_2^2}{n}} \quad (11)$$

とするのが極めて自然である。一方、 $\mu_1 - \mu_2$ の $100(1-\alpha)\%$ 信頼区間は、 m および n が大きいときは $z(\alpha/2)$ を標準正規分布の上側 $100\alpha/2\%$ 点とし、 SE を (11) の値として

$$(\bar{x} - \bar{y}) \pm z(\alpha/2)SE \quad (12)$$

で与えられる。しかしここでも「 m および n

が大きいとき」はどういう場合かのあいまいさが残る。AP Statistics では t 分布も導入していて、(12) の $z(\alpha/2)$ の代わりに t 分布のパーセント点を用いる手順も示されているが、そのときの自由度の決め方が一定していない。

普通の統計学の書物では $\sigma_1^2 = \sigma_2^2 (= \sigma^2)$ と仮定し、未知の共通分散 σ^2 の推定をプールした分散により行い、 t 分布の自由度を $m + n - 2$ とするのが常道であるが、 $\sigma_1^2 = \sigma_2^2$ と仮定できない場合の話はあまりしないことになっている（教師側にこのあたりの話がきちんとできるだけの力量が要求されるのだが）。生徒（学生）はこの話題のあたりに来ると相当疲労していて、ただでさえなかなかきちんとした理解にまで至らない。

IV. おわりに

高等学校での統計教育は大学初年級の統計教育と切り離しては考えられない（はずである）。大学入試では、高等学校までの各教科内容の修得状況の評価に加え、大学入学後の資質を見極めるといった目的を持っている。すなわち生徒の大学入学後のパフォーマンスを予測するという大きな役割を担っているのである。高等学校の学習内容と大学での勉学はシームレスにつながるべきである。その意味で、AP Statistics は高大双方の学習内容を再確認するよい機会であると思われる。

高等学校での統計学は数学の中で展開されることになっている。しかし前節で見たように、これまでの普通の数学とそぐわない内容が（不可避免的に）統計学の中に潜んでいる。統計学は現実問題で生じる不確定性に関する学問であり、過度に精確な議論は本来できないはずである。そこが数学とは相容れない。

でもそれでもよいのであって、Meng (2009) などの文献では、統計的方法 (statistical Methods) よりも統計的な考え方 (statistical thinking) に重きを置くべきであるとの見解が繰り返し述べられていて、それは Brown and Kass (2009) およびその中の何人かの統計学者のコメントにおいても概ね支持されている。しかしそれは相当に困難な道である。公式を覚え、それを当てはめるだけという授業ではそれは望むべくもない。幸いなことに

（特によくできる）生徒は考え方 (thinking) に対する要求が強い。実際筆者がお茶の水女子大学で授業した際も、授業回数を重ねるにつれだんだん時間がなくなってきて、公式を提示しそれを用いて例題を解くだけという状況に陥ってしまったが、それに対し何人かの学生からは（全く正当にも）不満の声を聞いた。

第1節に挙げた統計教育に関する文献を読むと、うらやましいという気がしないでもない。Meng (2009) は、統計学入門のコースを教える教師は

- (I) extensive statistical knowledge
- (II) deep understanding of statistical foundations
- (III) substantial experience in statistical practice
- (IV) great communication skills
- (V) profound pedagogical passion

の5つ (THE FIVE) を持つべしと書いている（もちろん理想にはという意味で）。そして、この中の最初の3つは既に Hotelling (1940) によって指摘されたものであるとも記している。

翻って日本では、統計学の授業が統計学の専門家以外によって教えられることが多く、THE FIVE の最初の3つすら満足されていない状況である、と見ている。それは教える側にとっても教えられる側にとっても不幸なことである。英米の文献で statistical thinking が重要視されているということは、逆に言えば現状はそうでないことも意味している（そうであれば論文には書かれない）。しかし、その方向に向かっていることは事実であり、我々もそれを十分認識する必要がある。勝負はこれからである。

Yes We Can, Too.

〈引用文献〉

- Agresti, A. and Coull, B. A. (1998). Approximate is better than "exact" for interval estimation of binomial proportions, *American Statistician*, **52**, pp. 119-126.
- Agresti, A. and Franklin, C. (2007). *Statistics, The Art and Science of Learning From Data*,

- Pearson Education, New Jersey.
- Brown, E. N. and Kass, R. E. (2009). What is statistics? *American Statistician*, **63**, pp. 105-123 (with comments).
- Gelman, A. and Nolan, D. (2002). *Teaching Statistics, A Bag of Tricks*, Oxford University Press, New York.
- Hedayat, S. and Stufken, J. (2009). Comment on "What is statistics," by E. N. Brown and R. E. Kass, *American Statistician*, **63**, pp. 115-116.
- Hinders, D. C. (2008). *Five Steps to a 5, AP Statistics 2008-2009*, McGraw Hill, New York.
- Hotelling, H. (1940). The teaching of statistics, *Annals of Mathematical Statistics*, **11**, pp. 457-472.
- Hulsizer, M. R. and Woolf, L. M. (2009). *A Guide to Teaching Statistics, Innovation and Best Practices*, Wiley-Blackwell, West Sussex, UK.
- 岩崎 学 (2007). 確率・統計の基礎 東京図書
- Meng, X.-L. (2009). Desired and feared – What do we do now and over the next 50 years? *American Statistician*, **63**, pp. 202-210.
- Mulekar, M. S. (2008). *Cracking the AP Statistics Exam, 2008 Edition*, Princeton Review, New York.
- Patterson, B. F. (2009). AP Statistics: Students' Choices after high school, *Amstat News*, May 2009 Issue (No. 383), pp. 6-12.
- Roberts, R., Scheaffer, R. and Watkins, A. (1999). Advanced Placement Statistics – Past, present, and future, *American Statistician*, **53**, pp. 307-320.
- Simmons, B., Bland, M. J. and Wojciechowski, B. (2008). *AP Statistics 2008 Edition*, Kaplan Publishing, New York.
- Smith, T. M. F. and Staetsky, L. (2007). The teaching of statistics in UK universities, *Journal of the Royal Statistical Society, Series A*, **170**, pp. 581-622 (with discussion).
- Sternstein, M. (2007). *AP Statistics 2008, 4th Edition*, Barron's Educational Series, New York.