統計教育の諸問題

宮 本 一 郎*

On the Problems about Statistics Education

Ichiro MIYAMOTO

The aim of statistics education is the understanding of statistical ways of thinking and their techniques, and developing the attitudes for applying them in various fields. And the basic idea is the relationship between "population" and its "sample". But the mathematical theory of the relation is difficult for the students in the elementaly stage, and it is desirable that statistics should be taught with as little mathematics as possible, so as to catch the general view of statistics. For this purpose, how should the contents and their treatments be? In this paper, the concerning problems and their counterplans are proposed.

ま え が き

統計教育の目的は，統計的な考え方とその手段を理解させ，これをいろいろの方面に活用する態度を養うことにあると考えられる。その基礎は「母集団」と「標本」との関係にある。しかしこれを数理統計学的に論じると，統計学が難しいものとなるので，初歩的段階において数理的な部分をできるだけ割愛してむしろその考え方を強調し，統計学そのものの概観を把握させることが望ましい。（その後，必要があればさらにそれを数理的に見直せばよい。）そのためには，その教授内容・範囲とその指導・取扱いをどのようにすべきかについての諸問題とそれに対する方策に関して具体的に詳述する。

1．内容・範囲

基礎教養としての統計学と専門統計学との境界をどこにおくかが問題である。

(1) 推定論

この部分は標本論として専門的には重要であるが，入門レベルとしてはそれほど深入りする必要がないと考えられる。不偏性を重点とし，一致性は常識的に認め，有効性はたとえば$\bar{x}$は$\tilde{x}$より有効であるなどの例により概念的に理解させる程度でよい。最尤性については数学的に難しくな

＊教養部

るので割愛する方がよい。

(2) 確率分布

二項分布，超幾何分布，一様分布以外は密度関係の誘導をやらないでヒストグラムなどから直観的に導き，その応用例を中心として展開する。従って積率母関数にはふれない。ただし分布の相互関係については十分理解させておくようにする。

例　超幾何分布と二項分布，二項分布の正規近似・ポアソン近似，ポアソン分布と指数分布，標準正規分布とｔ分布，ｔ分布とＦ分布など。

取扱う分布としては，二項分布，超幾何分布，ポアソン分布，一様分布，指数分布，正規分布，χ^2分布，ｔ分布，Ｆ分布とし，ガンマ分布，パスカル分布は余力教材（補節，付録など）とするとよい。

(3) 分散分析

これは従来，専門的な分野で，入門レベルではふれないものが多かったが，現在手法としてよく用いられるようになったので，むしろ平均値の比較検定の延長として簡単に導入した方がよい。レベルは二元配置（繰り返しのない場合）までで十分であろう。

(4) ノンパラメトリック検定

現在用いられている推定検定の手法は正規分布を前提とするものが大部分であるが，正規性が疑わしい場合，あるいは分布そのものがよくわからない場合に，依りどころとなるものとして，簡単にふれておくとよい。その内容としては，符号検定，順位和検定，順位相関などが有用であると考えられる。

(5) 数値表

正規分布表，ｔ表，χ^2表，Ｆ表，（一様）乱数表を基礎とする。さらにできれば順位和検定の表[5]，ポアソン分布表があると便利である。しかし順位和検定は大標本では正規近似できるし，またポアソン分布の確率は漸化式 $P(x+1)=\lambda P(x)/(x+1)$ で電卓で簡単に求められるからなくても差支えない。正規分布表は通常の $I(x)=\int_0^x \phi(x)\,dx$ の表のほか，片側パーセント点 $Z_{\alpha/2}$ の表（日科技連：新編数値表 P.5）を入れるとよい。Ｆ分布表は５％，１％，2.5%，0.5%の４種のものがほしい。（５％，１％しかないものが多いが，これでは両側検定の場合に困る。）

２．指導・取扱い

(1) 数式の取扱い

- まずどの程度の数学学力を前提とするかが問題であるが，最少限，$n!$，${}_nC_r$，Σ，$\int$の基本計算が必要である。特に統計学ではΣが中心となるので，定義とその基本性質（線形性）を最初に

十分準備確認することが必要である。ただし$\sum x_i$，あるいは$\sum x$のような略記法でよい。

・定理の証明は簡単に代数計算でできるもの，たとえば$\bar{x}$，sの基本性質や，これを用いた簡便計算法の原理などはやった方がよい。しかし分布に関する難しい定理の証明は省略して，できればこれを実験や数値例[6]により直観的に把握させることが望ましい。たとえば中心極限定理はチップ実験が有効である。また定理の前提条件，近似条件（頑健性も含めて）を明らかに示しておくことが必要である。たとえば二項分布の正規近似については，

「$B(n, p)$においてpまたは$q=1-p$があまり小さくなく，nが十分大きいとき(実用的には$np>5$，$n>30$程度のとき正規分布$N(np, npq)$で近似してよい。」

のように示しておくとよい。

・最小2乗法による回帰式の推定においては，最初は一応微分法を用いて導くが，以後正規方程式を次のように構成していくとよい。(これは偏微分法を適用するのと本質的に同じである。)

$$\hat{y}=a+bx \longrightarrow na+b\sum x=\sum y,\ a\sum x+b\sum x^2=\sum xy$$

$$\hat{y}=a+bx_1+cx_2 \longrightarrow \begin{cases} na+b\sum x_1+c\sum x_2=\sum y \\ a\sum x_1+b\sum x_1^2+c\sum x_1x_2=\sum x_1y \\ a\sum x_2+b\sum x_1x_2+c\sum x_2^2=\sum x_2y \end{cases}$$

(その他の場合も同様)

(2) 推定・検定の指導

・推定は点推定論に深入りしないで区間推定を中心として展開する。また有限母集団の標本調査についても簡単にふれておくことが望ましい。

・検定の考え方は導入の段階において案外わかりにくいので，その手順を次の5項目で箇条書にするとよい。

①帰無仮説H_0(対立仮説H_1も併記)　②標本分布　③棄却域　④実現値　⑤結論(危険率も併記)

例　あるマーケットで1袋10g入りの商品をランダムに25枚とり出してその平均を求めたら9.8gであった。この商品が10g入りであることは信用できるか，ただしこの商品の重量は正規分布に従い，その標準偏差は安定しており約0.5gであったとする。(有意水準5%)

帰無仮説H_0：$\mu=10$g　($\sigma=0.5$g)

　　　　H_1：$\mu \gtrless 10$g　($\sigma=0.5$g)

標本分布：$n=25$のとき　$\sigma_{\bar{x}}=0.5/\sqrt{25}=0.1$であるから，$\bar{X}$は$N(10, 0.1^2)$に従い，$Z=(\bar{X}-10)/0.1$は$N(0, 1)$に従う。

棄 却 域：$\alpha=0.05$，両側検定　　　$R：|Z| \geqq 1.96$

実 現 値：$z=(9.80-10)/0.1=-2 \in R$　　H_0は棄却される。

結　　論：この商品は10g入りであるとはいえない。〔$\alpha=5$%〕

注　第2種の誤まり，検出力については正規検定の例題について具体的に一応説明するが，以後深入りしない。

・母数の推定，検定は同じcaseについて別々にやる方がよいが，母集団間の有意差については検定のところで推定もまとめてやる方が能率的である。(従来のテキストは別々にやっているものが多いが) またその延長として分散分析の基礎を指導するとよい。

(3) 標本分布(χ^2分布，t分布，F分布)の導入

これについては2通りの導入の仕方がある。

①標本分布としてまとめてやる。

②正規分布を中心として推定検定をやり，後必要に応じてt分布，χ^2分布，F分布を別々に導入する。

①が普通であるが，証明を省略するときは無味乾燥になる嫌いがあるので，次のような取扱いが望ましい。

・χ^2分布はN(0, 1)から標本2乗和の分布として導入し，N(μ, σ^2)における$\chi^2=\sum(X_i-\mu)^2/\sigma^2$の分布を導き，次に$\chi^2=\sum(X_i-\bar{X})^2/\sigma^2$が自由度($n-1$)の$\chi^2$分布に従うことを「自由度」の説明と合せて直観的に理解させる。

・t分布，F分布は生成定理を無証明に示し，これにより正規母集団の標本定理を導く。

・分布式は参考に示すだけとし，数値表の引き方を重点として簡単な例題を入れる。

②は羅列の欠点を除き，実際の例題から興味をもたせて導入する方法であるが，分布そのものの説明が十分でないという欠陥がある。指導時間が少い場合に限った方がよい。

(4) 数値計算

電卓を用いて自分でやってみることで面白味がわかる。入門レベルで統計用電卓やパソコンを用いることは好ましくないと思われる。(一通り学習してからその実際の応用例を見るためにはパソコンは有用であるが)。ただし電卓の計算で途中の数字を出すときは，あまり数字を丸め過ぎないようにし(有効数字4けた程度保存し) 最後の結果を適当に丸める（3けた程度）指導が必要である。

3．記　号

記号の使用[3)]がテキストによってまちまちであることが初学者に少なからぬ混乱を起しているので，これについての検討が必要である。記号の統一に関しては，次の要件を配慮すべきである。

（i） 整合性

他の場合と混同して誤解を生じないこと。

横の整合性（関連分野）と縦の整合性（拡張，発展）に留意。

（ii） 実用性

よく用いられる記法はなるべく保存する。(慣習，多用性)

いろいろあるものは主流となっているものにする。(学術用語集，JIS，辞典など)

(iii) 有効性

教育上有効と考えられる記法を積極的に取り入れる。

(iv) 統一性

統一した方がよいものと統一しない方がよいものとあるのでこれを検討する。

次に，統計教育に関するものについて述べる。

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

統計学の入門テキストではほとんど①であるが，QC 関係，JIS では主として②を用いている。またアメリカのテキストでは②がよく用いられている。

①はs^2が偏差の 2 乗平均，すなわち$\bar{x}$のまわりの 2 次モーメントとして自然的であり，また$s^2 = \overline{x^2} - \bar{x}^2$は母集団分布における $D^2(X) = E(X^2) - E^2(X)$にもよく整合する。

②は(平方和)/(自由度) = (不偏分散)の考えから，推定検定でほとんどこれが用いられ，実用性をもつが，整合性を欠き，また導入の段階でなぜ $(n-1)$ で割るのか説明ができない。またこれで導入すると，共分数の定義においても $\sum(x_i - \bar{x})(y_i - \bar{y})/(n-1)$としなければならない。

注 ①で導入するものは②の段階で，不偏分数を

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

と定義するものがあるが，U^2あるいはu^2は記号として不自然である。

また①において標準偏差を σ で表すものもあるが，統計量をギリシヤ文字で表すのは，母数との混同の恐れがあり，整合的でない。

(提案)

①で導入し，分散を V_n，標準偏差をs_nで表す。(nをつける)

推定検定の段階で，不偏分散を V，標準偏差を s で表す。(nをつけない)

また確率分布において分数は V(X)，Var(X)，$\sigma^2(X)$，$D^2(X)$などいろいろの記法があるが V(X)，または $D^2(X)$とし，標準偏差は D(X)が適当である。(期待値は E(X)なので，E，D と親近性があってよい。)

(2) 分布のパーセント点

対称な確率分布の両側確率，上側確率のパーセント点の表し方に混乱が見られるので，これを次のように区別して明確にすることが望ましい。

(提案)

$$\begin{cases} \text{両側確率} \quad P\{|x| \geqq c\} = \alpha \text{ のとき} \quad c = x(\alpha) \\ \text{上側確率} \quad P\{x \geqq c\} = \alpha \text{ のとき} \quad c = x_\alpha \end{cases}$$

例	両側点	上側点	
t分布	$t(\phi, \alpha)$	$t_\alpha(\phi)$	$\therefore t(\alpha) = t_{\alpha/2},\ z(\alpha) = z_{\alpha/2}$
正規分布	$z(\alpha)$	z_α	

ただしχ^2分布，F分布のような非対称分布については混乱しないので，慣例に従って，$\chi^2(\phi_1, \phi_2; \alpha)$でも差支ない。しかし整合性から，$\chi^2_\alpha(\phi)$，$F_\alpha(\phi_1, \phi_2)$の記法が望ましい。

(3) 自由度

自由度を表わすのに，ϕ，f，ν，nなどが用いられている。nは分布式と直接関連してよいが自由度が（$n-1$）になったとき混乱がある。この点でνはよいが，この文字は書きにくいという欠点がある。ϕはQC関係でよく用いられて実用性があるが，分布式中にϕを用いるのは不自然である。fはこれをいくらか折衷したものでfreedomの頭文字でもあるが不自然さは残る。

(提案)

導入段階では分布式と関連してnを用い，数値表では後の実用性を考慮してϕとする。

(4) 確率分布への所属

たとえば「XはN(μ, σ^2)に従う」ことを表す正式な記法はないが，次のようなものがみられる。

①$X \in N(\mu, \sigma^2)$　②$X : N(\mu, \sigma^2)$　③$X \sim N(\mu, \sigma^2)$

いずれも大差はないが，②が最も使い易いように思われる。

また分布の近似を表すには≈が便利である。　例　$B(n, p) \approx N(np, npq)$

おわりに

以上は著者が長年各種の統計教育(高校，高専，大学一般教養，および会社員のQC講習)にたずさわった体験を通してまとめたものの概要で，さらに詳細については拙著(文献1)，2)）を参照願いたく，これに関して諸賢のご批判，ご助言が頂ければ，まことに幸とするところである。

参　考　文　献

1）宮本一郎：よくわかる統計学概要，学術図書出版，1993.
2）　同　：統計入門(講習テキスト)(新訂版)，富山県経営者協会，1992.
3）　同　：数学教育における表記法の検討，福井工業大学研究紀要 Vol. 20，1990.
4）森口繁一編：新編 日科技連 数値表，日科技連出版，1990.
5）P.G. Hoel（浅井晃，村上正康訳）：初等統計学（原書第4版），培風館，1981.
6）森口繁一：初等数理統計学（改訂版），培風館，1957.
7）小和田正，日比野康文，磯村孝志，中川覃夫：統計学の基礎，実教出版，1984.
ほか統計学テキスト多数

(平成5年10月26日受理)