岐阜数学教育研究
2015, Vol. 14, 1-6

不偏分散から統計学の理解を促す教材についての考察

赤堀　克己[1]

統計学はおおまかにいって，記述統計学と推測統計学に分かれる。共に，統計学ではあるがその概念は大きく違い，その違いを理解することが統計学の学習において必須である。さらに，その違いが如実に現れるのが分散と不偏分散の関係性である。本論文では，大学１年生を対象とした不偏分散の理解を深めるための教材とそれを用いた授業実践の結果を報告すると共に，高校生への適用の可能性についても述べる。

＜キーワード＞　分散，不偏分散，記述統計学，推測統計学，標本調査，無作為抽出

１．はじめに

現代社会において，データを適切に処理し必要な情報のみを抽出する際に統計の手法が必要になる場面が多々見受けられる。それを受けて平成２０年度の中学校学習指導要領数学科の改訂において，「資料の活用」が領域として追加された。また，大学においても文系理系問わず多くの大学生が「統計学」を履修する。

しかしながら，従来より高等学校において，様々な理由から統計的な内容（現行の学習指導要領では「数学Ｂ」の「確率分布と統計的な推測」）はあまり履修されていないのが現状である。そのような現状を踏まえて作成した教材を大学１年生に対して実践した。以下，実践内容とその結果を報告するとともに，「統計学」の特性と生徒の数学の学習過程を考慮した統計学の効果的な学習方法についての考察を述べる。

２．研究のねらい

２．１．高等学校での履修状況

毎年，統計学の内容の講義の前に高等学校での履修状況についてのアンケートを実施しているが，統計的内容の履修率は極めて低い。今回の受講者４１名では，履修者はわずか３名（７%）であった。中学校，高等学校，大学とスムーズな内容の深化が理想であるが，現状は理想とはかけ離れている。このような現状を鑑みて，大学１年生向け，あるいは大学での「統計学」を視野にいれた高等学校での発展的な内容としての教材を検討した。

２．２．「統計学」の特性

「統計学」を学習する際，「記述統計学」→「推測統計学」の順で進めることは適切であろう。実際，中学校学習指導要領([1])において，第１学年で「資料のちらばりと代表値」，第３学年で「標本調査」となっている。さらに，高等学校学習指導要領([2])においても，「数学Ｉ」で「データの分析」，数学Ｂで「確率分布と統計的な推測」となっている。また，大学生や一般社会人向けの「統計学」の入門書もほとんど

[1]岐阜薬科大学

が「記述統計学」→「推測統計学」の順で記載されている。しかしながら，「記述統計学」から「推測統計学」に移行する際，授業者は生徒にその違いを強く強調する必要があると，筆者は考える。一般に,「推測統計学」では始めに点推定として，標本平均と不偏分散を学習する。標本平均については中学校第３学年の「標本調査」でも扱っている。「母集団」と「標本」の違いがよく理解できていなくても, 平均と標本平均の式は同じなので理解の欠如が露見しない。しかし，分散と不偏分散は式が異なるので，理解不足の生徒は「統計学」がわからなくなる端緒となる。標本平均も不偏分散もあくまで「推定量」であることを生徒に理解させることが肝要であると, 筆者は感じている。

では，「母集団」から「標本」を抽出して「母平均」や「母分散」を推測する過程を生徒に体験させるにはどうような教材がよいか？高等学校学習指導要領において，「コンピュータの積極的な活用」を掲げている。本実践においては，Excel の関数を利用して標本抽出のシミュレーションを通じ, 不偏分散の母分散の推定量としての妥当性の体験的理解を促すことを試みた。

２．３．学習者の意識

では，統計学の学習において，実験（シミュレーション）の結果による体験的理解で十分だろうか？山路([3],図１)においては,「標本の平均値が，全数調査の平均値とほぼ等しくなる理由をもっと考えたいか？」という問に対して, 約 67%の生徒が肯定的な回答を示している。数学において，論理的思考の学習に慣れ親しんでいる学習者がそう思うのは至極当然である。そこで，本実践においては，シミュレーションの後に引き続いて不偏分散が推定量として「不偏性」を満たしていることを数式によって示す教材も用意した。

２．４．学習効果の評価

中学校，高等学校において「統計学」は「数学」の１分野として扱われている。しかし，「推測統計学」においては推測の要素があるため「数学」とは性質を異にする。ゆえに, 高等学校学習指導要領においても「コンピュータの積極的な活用」による実験的な理解を推奨している。ところが，それだけでは学習者の理解度が深まらないのではないかという不安が残る。そこで，我々は授業前と Excel の実習後と数式での証明後で受講者の理解度についてのアンケートを実施し, 双方の学習効果を数値化して評価を試みた。

３．実践内容

講義名：「情報処理科学」

実践日：平成 27 年 1 月 21 日(月)

場所:岐阜薬科大学三田洞学舎村山情報処理センター

対象：大学１年生（４１名）

３．１．受講者について

今回の受講者のほとんどが高等学校の「情報」の授業で Excel を経験しており，さらに前期の「情報基礎実習」でも Excel の操作方法を復習済みである。また，前回までに，Excel を用いて平均，分散等を計算する実習も終えている。

３．２．Excel を用いた標本抽出の実習と数式による不偏性の証明

まず, 以下の操作手順に従って受講者に Excel の実習をしてもらった。実習後，以下の数式による不偏分散の「不偏性」を示す課題も提示した。

[EXCEL による実習]

（操作手順）

① 標準正規分布に従うデータ（=NORMSINV(RAND( )) ４個を１００回分用意する。

② １回目について，偏差平方和を =DEVSQ(データ範囲)で求める。

③ 偏差平方和を個数４で割る。

④ 偏差平方和を（個数－１）の３で割る。

⑤ 同様の計算を２回目から１００回目までのデータについて行う。

⑥ １００回分のデータの平均値をそれぞれ計算する。

⑦ 同様のことを８個のデータを１００個分用意して行ってみる。

⑧ 下の表をうめる。

⑨ 理論値１とどちらが近いか比較する。

n	偏差平方和／n	偏差平方和／（n-1）
4		
8		

[数式による不偏性の証明]

＜問＞　母平均μと母分散σ^2をもつ母集団から無作為に抽出した n 個の標本を $X_1, X_2, \cdots,$ X_n とし，$\overline{X} = (X_1+X_2+\cdots+X_n)/n$ とするとき，

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Xi - \overline{X}\right)^2$$

は母分散の不偏推定量であることを，空欄$\boxed{}$を埋めて示せ。

以下，確率変数 X について E[X]，V[X]はそれぞれ平均と分散を表す。

証明)) n 個の標本 X_1，X_2，・・・，X_n は，同一の母平均μと母分散σ^2をもつ母集団から無作為に抽出されているので，それぞれ平均μと分散σ^2をもつ同一の確率分布に従う互いに独立な確率変数とみなせる。

よって　E[Xi]=($\boxed{\mu}$)，V[Xi]=($\boxed{\sigma^2}$)　（i=１，２，・・・，n ）

$$E[S^2] = E\left[\frac{1}{n-1}\sum_{i=1}^{n}\left(Xi - \overline{X}\right)^2\right]$$

$$=\left(\boxed{\tfrac{1}{n-1}}\right) E\left[\sum_{i=1}^{n}\left(Xi-\overline{X}\right)^2\right]$$

[　Xi－$\overline{X}$ = (Xi－μ)－($\overline{X}$－μ)　として展開すると]

$$=\left(\boxed{\tfrac{1}{n-1}}\right)E\left[\sum_{i=1}^{n}\left\{\left(\boxed{Xi-\mu}\right)^2-2\left(\boxed{\overline{X}-\mu}\right)\left(\boxed{Xi-\mu}\right)+\left(\boxed{\overline{X}-\mu}\right)^2\right\}\right]$$

$$=\left(\boxed{\tfrac{1}{n-1}}\right)E\left[\sum_{i=1}^{n}\left(\boxed{Xi-\mu}\right)^2-2\left(\boxed{\overline{X}-\mu}\right)\sum_{i=1}^{n}\left(\boxed{Xi-\mu}\right)+\left(\boxed{\overline{X}-\mu}\right)^2\sum_{i=1}^{n}1\right]$$

[$\sum_{i=1}^{n}Xi=n\overline{X}$, $\sum_{i=1}^{n}1=n$　だから]

$$=\left(\boxed{\tfrac{1}{n-1}}\right) E\left[\sum_{i=1}^{n}\left(\boxed{Xi-\mu}\right)^2-2\left(\boxed{n}\right)(\overline{X}-\mu)^2+\left(\boxed{n}\right)(\overline{X}-\mu)^2\right]$$

$$=\left(\boxed{\tfrac{1}{n-1}}\right) E\left[\sum_{i=1}^{n}\left(\boxed{Xi-\mu}\right)^2-\left(\boxed{n}\right)(\overline{X}-\mu)^2\right]$$

$$=\left(\boxed{\tfrac{1}{n-1}}\right) \left\{E\left[\sum_{i=1}^{n}\left(\boxed{Xi-\mu}\right)^2\right]-\left(\boxed{n}\right)E\left[(\overline{X}-\mu)^2\right]\right\}$$

ここで

$$E\left[\sum_{i=1}^{n}(Xi-\mu)^2\right]=\sum_{i=1}^{n}E\left[\left(\boxed{Xi-\mu}\right)^2\right]=\sum_{i=1}^{n}\left(\boxed{\sigma^2}\right)=\left(\boxed{n}\right)\left(\boxed{\sigma^2}\right)$$

$$\begin{aligned}E\left[(\overline{X}-\mu)^2\right]&=V\left[\left(\boxed{\overline{X}}\right)\right]=V[(X_1+X_2+\cdots+X_n)/n]\\&=1/n^2\left\{\left(\boxed{V[X_1]}\right)+\left(\boxed{V[X_2]}\right)+\cdots+\left(\boxed{V[X_n]}\right)\right\}\\&=1/n^2\left\{\left(\boxed{\sigma^2}\right)+\left(\boxed{\sigma^2}\right)+\cdots+\left(\boxed{\sigma^2}\right)\right\}\\&=1/n^2\cdot\left(\boxed{n\sigma^2}\right)\\&=\left(\boxed{\tfrac{\sigma^2}{n}}\right)\text{だから}\end{aligned}$$

$$E[S^2]=\left(\boxed{\tfrac{1}{n-1}}\right)\left\{\left(\boxed{n\sigma^2}\right)-n\left(\boxed{\tfrac{\sigma^2}{n}}\right)\right\}=\sigma^2$$

(証明終)

４．実践における受講者の活動の様子

４．１．Excel による実習

Excel の操作自体に戸惑っている受講者はおらず, 操作方法についての質問はでなかった。標本数が小さい n＝４の場合においては,（偏差平方和）／ｎの方が（偏差平方和）／（ｎ－１）より理論値１に近い場合も出現するが, 再計算を繰り返すとほぼ (偏差平方和) ／（ｎ－１）の方が (偏差平方和)／ｎより理論値１に近くなることを確認できている様子だった。

４．２．数式による証明

本実践の前の分散に関する実習において，平均Ｅ[X]と分散Ｖ[X]に関する公式

ｎ個の独立な確率変数 X_1，X_2，・・・，X_nについて，次の公式が成り立つ。(a_1,

a_2,・・・, a_nは定数)

(1) $E[a_1X_1+a_2X_2+\cdots+a_nX_n]$

$=a_1E[X_1]+a_2E\{X_2\}+\cdots+a_nE[X_n]$

(2) $V[a_1X_1+a_2X_2+\cdots+a_nX_n]$

$=a_1^2V[X_1]+a_2^2V[X_2]+\cdots+a_n^2V[X_n]$

を復習しておいたが,使いこなせない受講者が多くみられ,証明を最後まで完結している受講者は少数であった。

５．アンケート結果と考察

授業の前,Excel の実習後,数式による証明後と各時点での理解度の深まりを確認するために以下のアンケートを実施した。

Ｑ１．今現在（受講前),分散と不偏分散の違いがよく理解出来ている。

1. 全然理解できていない（理解度１０％未満)。
2. 少ししか理解出来ていない（理解度１０％以上４０％未満)。
3. 半分ぐらいは理解している（理解度４０％以上６０％未満)。
4. 大体理解している（理解度６０％以上９０％未満)。
5. 完璧に理解している（理解度９０％以上)。

Ｑ２．Excel の実習後，不偏分散は偏差平方和を n ではなく n-1 で割ることについて

１．全然わからない（理解度１０％未満)。

２．少ししかわからない（理解度１０％以上４０％未満)。

３．なんとなくわかった（理解度４０％以上６０％未満)。

４．大体わかった（理解度６０％以上９０％未満)。

５．完璧にわかった（理解度９０％以上)。

Ｑ３．不偏分散の不偏性の証明後，不偏分散は偏差平方和を n ではなく n-1 で割ることについて

１．全然わからない（理解度１０％未満)。

２．少ししかわからない（理解度１０％以上４０％未満)。

３．なんとなくわかった（理解度４０％以上６０％未満)。

４．大体わかった（理解度６０％以上９０％未満)。

５．完璧にわかった（理解度９０％以上)。

各問に対する回答の集計結果が[表１]である。

回答番号	Ｑ１	Ｑ２	Ｑ３
1	10 人	3 人	3 人
2	26 人	11 人	7 人
3	3 人	19 人	17 人
4	2 人	7 人	12 人
5	1 人	1 人	2 人
合計	41 人	41 人	41 人

[表１]

さらに,どちらの教材で理解が深まったかを見る指標として，回答番号の差（Ｑ２

－Ｑ１）と（Ｑ３－Ｑ２）を調べた。[表２]がその集計結果である。

回答番号の差	Q2－Q1	Q3－Q2
0 point	12 人	３３人
1 point	23 人	5 人
2 point	5 人	3 人
3 point	1 人	0 人
合計	４１人	４１人

[表２]

[表２]より，Excel の実習で理解度が深まったと回答した受講者は全体の 71%にのぼり実験的理解の学習効果を確認することが出来た。一方，数式による不偏性の証明の学習において理解度がさらに深まったと回答した受講者は20%(8 人)に止まった。しかし，Ｑ３の回答番号が４以上の受講生 14 人に限定してみると 50%(7 人)が数式による証明により理解が深まったと回答した。逆に，Ｑ３の回答番号が３以下の受講生 27 人に限定してみると 3%(１人)しか数式による証明の学習効果が見られなかった。この結果より，理解がある程度進んだ学習者に対してのみ数式による証明は効果が期待できるといえる。

６．今後の課題

６．１．推測する作業の追加

今回の実践では，不偏分散が「不偏性」を満たしていることの理解を深めることのみに絞って行った。母分散の推定量を学習する際，次のような過程が考えられる。

① 標本平均が「不偏性」を満たしていることを Excel の実習で確認する。

② （偏差平方和)/(標本数）が「不偏性」を満たしていないことを Excel の実習で体験する。

③ 母分散の推定量として何が適切か学習者に考えるよう促す。

④ 母分散の推定量の候補を発表してもらい，クラス全体で共有する。

⑤ 各自 Excel の実習を行い，どれが「不偏性」を満たすが確認する。

このような学習過程を経たならば，学習者は③の活動において推測していることを体感でき，「記述統計学」と「推測統計学」の違いをより実感できるのではないかと筆者は考える。

６．２．高校生への適用

母分散の推定は高等学校の「数学Ｂ」の学習指導要領の範囲外ではあるが，分散は範囲内であり標本調査の理解を深めるための発展的内容としては取り上げることは可能である。また，Excel の操作についても「情報」を履修している高校生にとっては全く問題ないと思われる。

引用・参考文献

[1] 文部科学省，2008，中学校学習指導要領解説　数学編

[2] 文部科学省，2009，高等学校学習指導要領解説　数学編

[3] 山路健祐，2013，確率の考えを活用して標本調査のしくみの理解を促す指導，岐阜数学教育研究，Vol. 12，42-51