

統計学テキストの「分散」の表記に関する調査

中村 永友¹花岡 重利²土屋 高宏³

要 旨

日本の統計学の書籍・テキスト（大学の教養レベル以上で使う教科書や専門書）において、分散の「呼称」や「記号」の使い方にばらつきがある。統計学の基本である分散は、その表記方法の主なものとして標本分散が S^2 、不偏分散が U^2 、母分散が σ^2 などの記号が使われている。その理論を説明するストーリーなどから、様々な記号が著者独自の視点で使われている。しかし、物理学や化学などの基礎科学のテキストでも統一された記号が使われていることと比べると、統計学の分散は特異である。実際のばらつきを調べるため、札幌学院大学図書館の蔵書と著者個人の蔵書を合わせて約650冊に対して調査した。また、初学者への配慮として、どのような道筋で「分散」を導くべきかを提言した。

キーワード：標本分散，不偏分散，母分散

1 はじめに

日本の大学の教養レベル以上で使う統計学のテキストや専門書において、データのバラツキを表す尺度である「分散」の呼称やその記号の使い方いくつかのヴァリエーションがある。その実情を把握するために、札幌学院大学図書館の蔵書（主として図書分類番号417「確率論・数理統計学」）と著者所有の蔵書をあわせて約650冊に対して調査をした。

物理学や化学などの基礎科学を学ぶとき、基本的な定数や数式はどのテキストを見ても統一された記号が使われている。その数式の成り立ちが物理量の単位などにも密接に関わっているためである。例えば、万有引力定数 G 、光速 c 、ボルツマン定数 k 、気体定数 R 等を見れば明らかであろう。しかし、統計学の基本中の基本である分散の表記は、主なものとして標本分散は S^2 や S 、不偏分散は U や V などの記号が使われている。さらに、理論を説明するストーリーや独自の視点などから、2つの分散は入り交じって σ^2 、 S 、 S^2 、 V 、 V^2 、 U 、 U^2 等々の記号が使われている。大文字が確率

変数、小文字がその実現値という視点では、 σ^2 を除いて、少なくとも単純にこれらの倍の種類があると考えても良い。

一方、それらの呼称においては、異なる式であるのに同じ用語で「分散」や「標本分散」ということもある。また、同じ用語なのに異なる式と言うこともある。さらに、「不偏分散」は「標本不偏分散」と「不偏標本分散」という呼び名が用いられたりする。これは“unbiased sample variance”に対する訳語である。統計学関連の海外からの専門用語の和訳が統一されていないのは、教育的な視点でも大きな問題であろう。これはまだ訳語なので表層的な問題であるが、なぜ分散の呼称と記号が多様に使われるのか。すぐに考えられることとしては、統計学の発展が多様な分野、例えば生物学、農学、数学、医学、経済学、……などでなされたため、異分野間の交流がなかったこと、さらに積極的に統一をしようとしなかったことではないだろうか。

以後、第2節では分散の分類と問題の根源を指摘する。第3節では調査の概要を、第4節ではその結果を示す。第5節では調査の過程で統計学者ではない著書の指摘を示し、第6節ではより望ましい教育の過程を考察する。

¹ 札幌学院大学経済学部; nagatomo@sgu.ac.jp.

² 札幌学院大学経済学部.

³ 城西大学理学部; takahiro@josai.ac.jp.

表1：分散の分類

| 記号・定義式 | 意味 | 可能な呼称 |
|---|--|-----------------------|
| N | 母集団の全要素数, 母集団の大きさ | |
| n | 標本の要素数, 標本の大きさ | 標本数, データ数 |
| X_i | 無作為標本=確率変数. ($i=1, \dots, n < \infty$) | 無作為標本, 確率変数 |
| x_i | X の実現値. ($i=1, \dots, n < \infty$) | 実現値, データ |
| $S = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$ | 母集団の分散 (=母分散) の最尤推定量でありモーメント推定量. 推定量として \hat{S} | 標本分散, 分散 |
| $s = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ | S の実現値. 推定量として \hat{s} (\bar{x} が母平均のとき母分散) | (\bar{x} が母平均の母分散) |
| $V = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ | 母集団の分散 (=母分散) の不偏推定量 | 不偏分散, 標本不偏分散, 標本分散 |
| $v = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ | V の実現値 | |
| $U = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ | 母平均既知の母分散の不偏推定量 | 平均既知の不偏分散 |
| $u = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$ | U の実現値 | |
| $\sigma^2 = \frac{1}{n} \sum_{i=1}^N (x_i - \mu)^2$ | 母集団の分散. $\{x_i\}$ ($i=1, \dots, N$) はすべての母集団の要素. $\mu = \bar{x}$ なので $s = \sigma^2$ となる. | 母分散 |

2 分散の定義式

表1は, 一般的な状況における分散の定義式, その式の意味, 可能な呼称をまとめたものである. これらを基として, 本論文で対象となる分散の定義式を示す.

定義式A:

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{または} \quad \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

定義式B:

$$\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 \quad \text{または} \quad \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

定義式C:

$$\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 \quad \text{または} \quad \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

定義式Aは, 母分散の最尤推定量であり, モーメント推定量でもある. 定義式Aの后者は $n=N$ とするとき, (有限) 母集団の分散 (=母分散) を計算する式である. 定義式Bは, 推測統計の観点から不偏性という性質を持つ推定量で, 一般には「母分散の不偏推定量」=「不偏分散」と呼ばれることがほとんどである. 定義式Cは, 平均 μ が既知のときの分散の不偏推定量である. また, 定義式Cの后者は μ が母平均の時の分散である. この定義式は本調査では扱わない.

手元のデータの扱いとして母集団 (これがすべての

データである) か, 母集団からの標本であるかという違いがある. 言い換えると, 手元のデータに対して記述統計的視点と推測統計的視点があるということである. 平均 $\bar{x} = (1/n) \sum_{i=1}^n x_i$ は, 不偏性や一致性があり, さらに, この式の中に他の推定量が入っていないため, 分散のような混乱は生じない. 定義式Aと定義式Bは推測統計に関連した性質の有無に違いがあるために, 混乱が生じていると考えられる.

基本として定義式Aと定義式Bの2種類の分散を調査対象とするが, 厳密にはそれぞれ確率変数と実現値によるものがある. 本調査では, 全体の傾向を調べるためこれらは同一のものと見なした.

3 調査概要

分散の記号と呼称の実情を調査するため, 社会科学や自然科学での「計量～」を含む統計学に関して, 図書館と個人蔵書に対して調べた. 図書分類番号として,

- 417 確率論・数理統計学
- 331 経済学・経済思想 (計量経済関係の図書)

を対象とした. 以下の図書については今回の調査から外した.

- 350 統計
- 500 技術・工学 {505 逐次刊行物・年鑑・統計書}

表 2 : 調査書籍の分類

| ジャンル | 入門 | 専門 | ソフトウェア | 読み物 | その他 | 合計 |
|------|-------|-------|--------|------|------|-----|
| 冊数 | 294 | 205 | 63 | 43 | 34 | 643 |
| 割合 | 45.7% | 31.9% | 9.8% | 6.7% | 5.3% | |

表 3 : 分散の呼称

| 定義式 | A : $\frac{1}{n} \sum (x_i - \bar{x})^2$ | | | B : $\frac{1}{n-1} \sum (x_i - \bar{x})^2$ | | | |
|-----|--|-------|------|--|-------|-------|-----------------|
| 呼称 | 分散 | 標本分散 | 母分散 | 標本分散 | 分散 | 不偏分散 | 不偏標本分散 + 標本不偏分散 |
| 比率 | 56.9% | 34.7% | 8.3% | 52.8% | 22.2% | 21.5% | 3.2% |
| 度数 | 164 | 100 | 24 | 167 | 70 | 68 | 10 |

表 4 : 検索エンジンによる不偏分散のヒット数 (2014.1.14時点)

| 用語 | Google 検索 | Google Scholar 検索 |
|----------------------------|-----------|-------------------|
| 不偏標本分散 | 約2,750件 | 4件 |
| 標本不偏分散 | 約27,200件 | 14件 |
| "unbiased sample variance" | 約55,600件 | 532件 |
| "sample unbiased variance" | 約2,500件 | 17件 |

- 600 産業 {605 逐次刊行物・年鑑・統計書}
- 300 社会科学 {305 逐次刊行物・年鑑・統計書}

調査対象図書の基本的データは次の通りである。

- 調査冊数：643冊
- 出版年代：1964～2012年
- 内容の分類：入門書，専門書，ソフトウェア (R, S, SPSS, SAS, ……) の解説，読み物系，その他 (確率論等)
- のべ調査時間：約100時間

4 調査結果

調査結果の集計表をいくつか示すが，それらの合計は調査した書籍の数とは必ずしも一致しない。それは，1つの分散を複数の呼称や記号を使っている場合など，適宜複数の組み合わせをしているためである。

4.1 書籍のタイプ

まず，調査した書籍をその内容にしたがって分類すると，表2のようになった。文系総合大学の図書館なので，これらの数字はあくまでも参考である。

4.2 呼称について

表3に調査した分散の呼称に関する集計を示す。単

純に見て，定義式Aは「分散」，定義式Bは「標本分散」という呼称で使われる割合が多いようだ。定義式Aは母分散でもあり，標本分散（最尤推定量やモーメント推定量）でもあることから，単に「分散」と呼んだ方がよいのだろうか？ 定義式Bは標本からの母分散の不偏推定量なので，「標本分散」と呼ばれるのは至極真っ当である。これは単純集計なので，どのような文脈で使われているのか不明であり，詳細に見るためにはクロス集計などで見るか，この点に注目しての再調査が必要があらう。

参考として，「不偏分散」に関して Google 検索と Google Scholar による検索結果を表4に示す。

Google 検索で注意することは，例えば，“標本不偏分散”と入力して検索しても，「標本（不偏）分散」などの括弧が入った文字列も検索対象となることである。ここではあくまでも参考として，単純にヒットした件数を示す。むしろ Google Scholar の件数がより実情を反映しているとみて良いだろう。標本不偏分散と unbiased sample variance が多いことがわかる。

4.3 記号について

定義式Aの呼称と記号を表5に示す。記号の使い方の方の傾向を見るために，大文字と小文字，さらに推定量（値）を表す[^]（ハット）はその文字の大文字でま

表5：定義式A： $\frac{1}{n}\sum(x_i-\bar{x})^2$ をどう表記しているか

| 呼称 | {分散, 標本分散, 母分散} | | | | | | | 標本分散 | 分散 | 母分散 |
|-----|-----------------|-------|----------------|----------------|------|----------------|--------|------|-------|------|
| | 記号 | S | S ² | U ² | V | V ² | Var(x) | | | |
| 比率1 | 1.7% | 52.4% | 0.3% | 7.6% | 1.0% | 0.7% | 18.4% | 4.9% | 11.5% | 1.4% |
| 比率2 | 2.1% | 63.7% | 0.4% | 9.3% | 1.3% | 0.8% | 22.4% | — | — | — |
| 度数 | 5 | 151 | 1 | 22 | 3 | 2 | 53 | 14 | 33 | 4 |

※表記定義式の分散をどの記号を使っているか、もし記号がない場合はどう呼んでいるか、を集計。
 ※比率1は全集計に対する比率、比率2は記号のみでの比率。
 ※定義式AでU²というのは明らかに間違いで、この書籍の第2版では訂正されている。

表6：定義式B： $\frac{1}{n-1}\sum(x_i-\bar{x})^2$ をどう表記しているか

| 呼称 | {分散, 不偏分散, 標本分散, 母分散} | | | | | | | 標本分散 | 不偏標本分散 | 不偏分散 | 分散 |
|-----|-----------------------|-------|----------------|-------|----------------|------|----------------|------|--------|-------|------|
| | 記号 | S | S ² | U | U ² | V | V ² | | | | |
| 比率1 | 0.6% | 46.1% | 0.3% | 13.9% | 11.6% | 1.6% | 7.7% | 1.6% | 0.6% | 12.3% | 3.5% |
| 比率2 | 0.8% | 56.3% | 0.4% | 16.9% | 14.2% | 2.0% | 9.4% | — | — | — | — |
| 度数 | 2 | 143 | 1 | 43 | 36 | 5 | 24 | 5 | 2 | 38 | 11 |

※表記定義式の分散をどの記号を使っているか、もし記号がない場合はどう呼んでいるか、を集計。
 ※比率1は全集計に対する比率、比率2は記号のみでの比率。

表7：定義式Aの呼称と記号のクロス表

| | 呼 称 | 呼 称 | | | 合計 |
|----|------------------|-------------|-------------|-----------|--------------|
| | | 分 散 | 標本分散 | 母分散 | |
| 記号 | S系 | 86 (29.9%) | 83 (28.8%) | 2 (0.7%) | 171 (59.4%) |
| | V系 | 51 (17.7%) | 8 (2.8%) | 1 (0.3%) | 60 (20.8%) |
| | σ ² 系 | 27 (9.4%) | 9 (3.1%) | 21 (7.3%) | 57 (19.8%) |
| 合計 | | 164 (59.9%) | 100 (34.7%) | 24 (8.3%) | 288 (100 %) |

S系：S+S²+標本分散, V系：V+V²+Var(x)+分散, σ²系：σ²+σ²

めた。呼称の中括弧内は、各記号の呼称である。

この表の右3列は、当該テキストの中で定義式のみを示し、それを何らかの記号を割り当てていない場合の集計を示している。比率1は、これらを含めた場合の度数に対する比率である。一方、比率2は、右3列を除いた場合の記号のみでの度数に対する比率である。これらから、定義式Aに対する記号としては抜きん出てS²が最も比率が高い。また、母分散の計算のための式でもあるので、σ²の比率も高い。

次に、定義式Bの呼称と、使われている記号を表6に示す。表5と同様に記号の使われ方の傾向を見るために、大文字と小文字の表記は大文字でまとめた。

この表の右4列は、表5と同様に、当該テキストの中で定義式のみを示し、それを何らかの記号を割り当てていない場合の集計である。比率1と比率2も同様である。定義式Bに対する記号もS²が最も比率が高

く、次いでU², Vである。また、母分散の不偏推定量でもあるので、σ²もほどほどに使われている。

4.4 呼称と記号について

表7は、定義式Aに対しての呼称と記号の組み合わせを見るためのクロス表である。S系={S, S², 標本分散}, V系={V, V², 分散}, σ²系={σ², 母分散}という同類の文字をまとめてクロス集計をした。SはSample, VはVarianceの頭文字を由来としているであろうことを基準としたラベルである。

この結果を見ると、S系と「分散」、あるいはS系と「標本分散」という組み合わせの使われ方が相対的に多いことがわかる。次いでV系と「分散」である。

表8は、定義式Bに対しての呼称と記号の組み合わせを見るためのクロス表である。先ほどの分類と、U系={U, U², 不偏分散, 標本不偏分散, 不偏標本分散}

表 8 : 定義式 B の呼称と記号のクロス表

| | | 呼 称 | | | | 合計 | |
|----|--------------|-------------|------------|------------|----------|-----------|-------------------|
| | | 不偏分散 | 分 散 | 標本分散 | 母分散 | | 不偏標本分散+ 標本不偏分散 |
| 記号 | S系 | 49 (15.9%) | 38 (12.3%) | 56 (18.1%) | 1 (0.3%) | 6 (1.9%) | 150 (48.5%) |
| | U系 | 78 (25.2%) | 0 (0.0%) | 3 (0.0%) | 0 (0.0%) | 2 (0.6%) | 83 (26.9%) |
| | V系 | 25 (8.1%) | 21 (6.8%) | 5 (1.6%) | 0 (0.0%) | 1 (0.3%) | 52 (16.8%) |
| | σ^2 系 | 14 (4.5%) | 6 (1.9%) | 3 (1.0%) | 0 (0.0%) | 1 (0.3%) | 24 (7.8%) |
| | 合計 | 166 (53.7%) | 65 (21.0%) | 67 (21.7%) | 1 (0.3%) | 10 (3.2%) | 309 (100 %) |

S系： $S+S^2$ +標本分散，U系： $U+U^2$ ，V系： $V+V^2+\text{Var}(x)$ +分散， σ^2 系： $\sigma^2+\hat{\sigma}^2$

表 9 : 定義式 A と B の呼称のクロス表

| | | 定義 B | | | | 合計 | |
|------|------|-------------|-----------|------------|----------|----------|-------------------|
| | | 不偏分散 | 分 散 | 標本分散 | 母分散 | | 不偏標本分散+ 標本不偏分散 |
| 定義 A | 分散 | 60 (30.6%) | 7 (3.6%) | 25 (12.8%) | 0 (0.0%) | 1 (0.5%) | 93 (47.4%) |
| | 標本分散 | 66 (33.7%) | 7 (3.6%) | 2 (1.0%) | 0 (0.0%) | 5 (2.6%) | 80 (40.8%) |
| | 母分散 | 9 (4.6%) | 2 (1.0%) | 11 (5.6%) | 1 (0.5%) | 0 (0.0%) | 23 (11.7%) |
| | 合計 | 135 (68.9%) | 16 (8.2%) | 38 (19.4%) | 1 (0.5%) | 6 (3.1%) | 196 (100.0%) |

表10 : 定義式 A と B の記号のクロス表

| | | 定義式 B | | | | 合計 |
|-------|--------------|------------|------------|------------|--------------|--------------|
| | | S系 | U系 | V系 | σ^2 系 | |
| 定義式 A | S系 | 48 (24.7%) | 49 (25.3%) | 13 (6.7%) | 10 (5.2%) | 120 (61.9%) |
| | U系 | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) | 0 (0.0%) |
| | V系 | 7 (3.6%) | 17 (8.8%) | 6 (3.1%) | 2 (1.0%) | 32 (16.5%) |
| | σ^2 系 | 23 (11.9%) | 8 (4.1%) | 4 (2.1%) | 7 (3.6%) | 42 (21.6%) |
| | 合計 | 78 (40.2%) | 74 (38.1%) | 23 (11.9%) | 19 (9.8%) | 194 (100.0%) |

S系： $S+S^2$ +標本分散，U系： $U+U^2$ ，V系： $V+V^2+\text{Var}(x)$ +分散， σ^2 系： $\sigma^2+\hat{\sigma}^2$

を加えてクロス集計をした。U は Unbiased variance の頭文字を反映したラベルである。

この結果を見ると、最も多いのが U 系と「不偏分散」という組み合わせである。そして、S 系と「標本分散」、「不偏分散」、「分散」という組み合わせが相対的に高い。

表 9 は、定義式 A と定義式 B の呼称の組み合わせを見るためのクロス表である。{定義式 A, 定義式 B} として記述すると、{標本分散, 不偏分散} と {分散, 不偏分散} の組み合わせが最も多く、次いで {分散, 標本分散} であった。

表10は、定義式 A と定義式 B の記号使われ方の組み合わせを見るためのクロス表である。記号の数が多いので、S 系、U 系、V 系、 σ^2 系の視点で集計した。{定義式 A, 定義式 B} として記述すると、比較的多いの

は {S 系, U 系}, {S 系, S 系} で、次いで { σ^2 系, S 系}, {V 系, U 系} であった。

4.5 JIS と ISO の定義

日本工業規格 (JIS) と国際標準規格 (ISO) でも、統計学に関する基本用語が規定されている。その一部を付録 A と付録 B に示しておく。それらをまとめると、 σ^2 は確率変数・確率分布の分散、 s^2 または V を定義式 B で定義し、「不偏分散」または「標本分散」と規定している。ISO でも同様の記述があり、JIS はこれに準じているものと考えられる。

4.6 総括

今回の統計学書籍の調査をまとめると、以下のよう

になる。

- (1) 定義式Aでは S , S^2 , V , V^2 , Var , σ^2 が、定義式Bでは S , S^2 , U , U^2 , V , V^2 , Var , $\hat{\sigma}^2$ が使われていた。
- (2) 定義式Aにおいて、呼称では、分散、標本分散、母分散、記号では、 S^2 , V , σ^2 が多かった。
- (3) 定義式Bにおいて、呼称では（おおむね）不偏分散、標本分散、分散、記号では S^2 , U^2 , V が多かった。
- (4) 呼称の組み合わせは、{標本分散, 不偏分散}, {分散, 不偏分散} が比較的多かった。
- (5) 記号の組み合わせは{S系, U系}, {S系, S系}, { σ^2 系, S系} が比較的多かった。

5 統計学専門外からの指摘

分散に対する指摘が書かれた書籍があるので紹介する。まずは、京極 一樹 著「ちょっとわかればこんなに役に立つ 統計・確率のほんとうの使い道」から、第3章第3節「母集団・標本と平均・分散の関係」（101ページ）に以下のような記述がある。

……統計学は非常に便利なのですが、反面統計学では文字や記号が非常に多く、また「書籍によって使う記号がバラバラ」であることが極めて厄介です。この問題点は統計学会が何かで統一してほしいものです。

たとえば標本分散や標本標準偏差は、書籍によって2種類の定義があり、このあいまいさが解消される動きがまったくありません。
……

この後半の指摘は統計学を表面的に勉強したのちにおきる疑問であろう。この著者は標本や母集団という概念を的確に理解していないか、用語の真意が伝わっていないようだ。一方、教科書・テキスト執筆者（統計学者）への自戒として言えることは、テキストなどで、どうしてそうなのかということが記述されていない、あるいは簡潔な記述しかないということも指摘できよう。

次に、大上 丈彦 著「マンガでわかる統計学 素朴な疑問からゆる〜く解説」から「死んだ専門用語」（188ページ）には、以下のような記述がある。

たとえば「標本分散」。本書の定義では「標本の分散」なのですが、本によっては「標本の分散から推定した母分散」を指すことがあります。

ます。この違いは定義式にしたときに、前者は「 n で割る」、後者は「 $n-1$ で割る」ということに対応します。「では、どちらが正しいんだ?」と思って岩波数学事典をみると「標本分散： $V=1/n$ 〜(ただし、 $1/(n-1)$ 〜のこともある)」と書いてありました。……

つまり、本によっては定義が違うのを、岩波数学事典は許容しているのですね。これはとても困ったことです。人によって意味が違う専門用語はもはや専門用語として、役立たず、「標本分散」という専門用語は死んでいるのです。

この指摘は統計学を全く知らない初学者にとっては至極真つ当なご意見であろう。統計学は道具なので、様々な視点でその道具が有効に使えるか否かを判断する必要がある、という視点に立たないといけない。定義式Aは最尤推定量でありモーメント推定量であることから、この推定量で推定した分散も「標本分散」であることは確かである。一方、定義式Bは不偏分散であり、これも「標本分散」である。このように定義式は違えども、双方「標本分散」と言っても統計学的には間違いではない。

辞典としては歯切れの悪いと思われる記述であるが、標本と母集団を理解せずに一刀両断しているのは残念である。

統計学者でない方々からの指摘は、専門外であるからこそのものである。それを専門的に学ぶとき、様々なことを考慮しなければならなくなり、そう簡単には記述や処理ができないということが様々な存在するのである。これはどんな学問分野でも共通なことで、専門家が入門テキストを書くときには、初学者のことを十分に考える必要があるだろう。

6 そもそも標本分散とは

本節では、教養レベルの統計学を学生に教えることを前提として、分散をどのように教えるのがもっとも適切かを考える。

まず、「標本分散」というキーワードに注目するとき、手元のデータセットがどのような状況であるのかを考える必要がある。それは、「背後に母集団があり、そこから観測・抽出されたデータである」と、「これ以上のデータはなく、このデータで母集団である（有限母集団）」という2つの立場である。

念のため「標本 (sample)」の意味を広辞苑 (第6版)

から引くと、「推測統計のための集団から抜き出した個々の要素。これに対して元の集団を母集団という。サンプル、／見本。ひながた。転じて、典型的な例。」とある。これを踏まえると、わかっていることであるが、「標本」と見なしているデータは、あくまでも母集団の一部であり、母集団のミニチュア（典型、縮図）なのである。ということは、「標本分散」と言う場合には「標本」の分散であり、その値は母集団の分散を推定するためのものでなければ意味がない。つまり、標本であることから、そのデータから推測統計の視点を入れて推定した（計算した）値のみ意味があるのである。推測統計の観点からは、定義式Aは母分散の一つの推定量（最尤推定量、モーメント推定量）ではあるが、推定量として望ましい性質である不偏性を持ち合わせていない。初学者が他の性質（不変性、一致性等）を知らないのであれば、不偏分散でない分散にどんな意味があるのか、ということである。結論として、「標本分散」と言えばすなわち不偏性を持った分散＝「不偏分散」を意味するものとして教育すべきではないだろうか。

統計学の入門時（推測統計の考え方を学ぶ前）では n で割る分散はあくまでも手元のデータがそれすべてである、すなわち母集団の時に限り用いるべきで、それは単に「分散」という呼称で良いのではないだろうか。その理由は、統計学の入門段階で、不偏推定量の話をする前に最尤推定量やモーメント推定量を扱う可能性はほとんどないためである。「 n で割る分散」＝「母分散」あるいは単に「分散」と言うべきだろう。

一方、多変量解析では推測統計を意識せずに、そのデータから有意な何らかの構造を見いだすことを目的として、記述統計を基本として作られている手法が多い、この場合は手元のデータがすべてという前提なので、 n で割る分散のほうがわかりやすいのかもしれない。しかし、回帰モデルなどでは係数の仮説検定などを行う関係から、推測統計が自動的に入ってくるため、自由度の必要性がでてくる。しかしながら、推測の要素が入らない限り n であっても $n-1$ であっても本質的な違いが見いだせない。

このように考えると、母集団の設定があるなしにかかわらず、

- ・「標本分散」と言えば「不偏分散」を指す。
- ・「分散」と一般的に言った場合は、これらすべてを含

む、2次のモーメントに関わるバラツキを表す指標の総称として用いる、

と統一するのがすっきりするのではないか。

以上を踏まえると、次のような視点で分散の提示の仕方が良いのではないだろうか。

提示の順番

〈超入門時〉 広がり の尺度としての標準偏差と分散を学ぶ。手元のデータが母集団であることを前提として、定義式Aを提示。

〈他の位置と散布の尺度を学習〉 平均と同じレベルで標準偏差が比較できることが理解できるレベル＋手元のデータの背後に母集団があるということが理解できる、という条件で定義式B（不偏分散）の平方根が標準偏差、その2乗が分散と提示する。定義式Bの根拠は明示しない。

〈母集団の特性の推定〉 推定量のことを理解し、その持つべき性質（不偏性など）のいくつかを提示してから、定義式B（不偏分散）を正式に教える。

「統計学は平均と分散の学問」と言われるが、その片方の柱である分散には、本調査結果のように実に様々な表現がある。ある一定レベル以上の学習をすれば、著者の意図を見抜けるが、やはり統計学初学者にとっては表記や記号の多様性は混乱の原因であることに変わりがないであろう。

7 おわりに

統計学の初学者がとくに独学で数種類のテキストを参照したとき、記号がばらばらであることに何らかの戸惑いがあることを聞いたことがある。純粋数学のようにそのテキストの中で閉じている記号の使い方であれば、何ら問題はないであろうが、ある一定の統一性があった方が学ぶ側にとってつまずきの要因、とつつきにくさ、等を減らすことができるのではないか。

「統計学は大人の学問」とも言われる。それは、様々な教養が身につくにはじめて様々な統計学の景色が見えてくる、ということをも含む言葉であろう。したがって、便利な道具として使うためには、やはりそれなりにデータの背景や周辺知識が必要不可欠で、入門段階では2種類の分散があることへの疑問はそのままにしておいて、統計学が本当に必要なときにわかってもら

えばいいのであろう。

今回は和書を中心に調べたが、今後の課題としては、辞典の類と洋書でどのようになっているのかを調べる必要がある。

なお、本論文は花岡（2013）を基礎として、加筆修正したものである。

付録

A JIS における分散の記述（抜粋）

JIS Z 8101-1：1999

第1部：確率及び一般統計用語（Part 1: Probability and general statistical terms）

1.13 分散 ぶんさん variance

確率変数 X からその母平均を引いた変数の二乗の期待値。 $\sigma^2 = V(X) = E[X - E(X)]^2$ である。

備考

1. 確率変数の従う確率分布の分散ということもある。
2. 確率変数 X の分散を X の母分散ということもある。

2.19 標本分散, 不偏分散 ひょうほんぶんさん, ふへんぶんさん variance

各観測値の平均値からの偏差の二乗の和を観測個数から1を引いた数で割ったばらつきの尺度。

備考

1. サンプル (x_1, x_2, \dots, x_n) については, $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$ として求める。ここで, n は観測値の個数である。標本分散は V と表記してもよい。
2. 標本分散は, 母分散の不偏推定量である。
3. 標本分散は, 2次中心モーメントの $n/(n-1)$ 倍である。
4. 混乱が生じなければ, 標本分散を分散と呼んでもよい。

2.20 標本標準偏差 ひょうほんひょうじゅんへんさ standard deviation

標本分散の正の平方根。

備考

1. 標本標準偏差は, 母標準偏差のかたよりのある推定量となる。
2. 混乱が生じなければ, 標本標準偏差を標準偏差と呼んでもよい。

附属書 A（規定） JIS Z 8101-1 で用いられる記号

$E(X)$ 確率変数 X の期待値。（場合によっては, m で期待値を表す。）

.....

s 標本標準偏差。（ s^2 の備考を参照。）

s^2 標本分散。

備考

一般に, 標本が採られた母集団の分散の推定量として,

算術平均からの偏差の2乗和を $n-1$ で割った量を表すために記号 s^2 が, また記号 s はその量の平方根をあらわすために用いられる。

.....

σ 確率変数または確率分布の標準偏差。

$\sigma^2 V(X)$ 確率変数または確率分布の分散。

.....

B ISO における分散の記述（抜粋）

ISO 12491：1997 (en)

3.12 (population) variance, σ^2

For a continuous variable X having the probability density function $f(x)$, the variance, if it exists, is given by $\sigma^2 = \int (x - \mu)^2 f(x) dx$, the integral being extended over the interval(s) of variation of the variable X .

3.13 (population) standard deviation, σ .

Positive square root of the population variance σ^2 .

3.20 sample variance, s^2

Sum of n squared deviations from the sample mean \bar{x} divided by the sample size n minus 1: $s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$

3.21 sample standard deviation, s .

Positive square root of the sample variance s^2 .

Annex C Basic statistical terms and concepts.

C.2.19 arithmetic mean, average

The sum of values divided by the number of values.

NOTE 1 The term “mean” is used generally when referring to a population parameter and the term “average” when referring to the result of a calculation on the data obtained in a sample.

NOTE 2 The average of a simple random sample taken from a population is an unbiased estimator of the mean of this population. However, other estimators, such as the geometric or harmonic mean, or the median or mode, are sometimes used.

[ISO 3534-1:1993, definition 2.26]

C.2.20 variance

a measure of dispersion, which is the sum of the squared deviations of observations from their average divided by one less than the number of observations.

EXAMPLE For n observations x_1, x_2, \dots, x_n with average $\bar{x} = \frac{1}{n} \sum x_i$, the variance is $s^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$.

NOTE 1 The sample variance is an unbiased estimator of the population variance.

NOTE 2 The variance is $n/(n-1)$ times the central moment of order 2 (see note to ISO 3534-1:1993,

definition 2.39).
[ISO 3534-1:1993, definition 2.33]

Guide Comment: The variance defined here is more appropriately designated the “sample estimate of the population variance”. The variance of a sample is usually defined to be the central moment of order 2 of the sample (see C.2.13 and C.2.22).

C.2.21 standard deviation

The positive square root of the variance.

NOTE The sample standard deviation is a biased estimator of the population standard deviation.

[ISO 3534-1:1993, definition 2.34]

C.2.22 central moment of order q

In a distribution of a single characteristic, the arithmetic mean of the q th power of the difference between

the observed values and their average \bar{x} : $\frac{1}{n} \sum_i (x_i - \bar{x})^q$

where n is the number of observations.

NOTE The central moment of order 1 is equal to zero.

[ISO 3534-1:1993, definition 2.37]

C 標準正規分布の確率

標準正規分布の確率は入門的な教科書では付表として掲載されている。この種類も付加的に調査している。結果は次の通りである。

表11：標準正規分布の確率

| 領域 | $[z, \infty)$ | $[0, z]$ | $(-\infty, z]$ | $(-\infty, z] \text{ \& } [z, \infty)$ |
|----|---------------|----------|----------------|--|
| 度数 | 61 | 68 | 38 | 3 |
| 割合 | 35.9% | 40.0% | 22.4% | 1.8% |

積分区間が $(-\infty, z]$, $[0, z]$, $[z, 1)$ の比率がおおよそ 7 : 8 : 4 となっていて、標準正規分布の確率を求

める表が掲載されているテキストの中では、この3つでほとんどを占めていた。これは、「著者がどのような教科書で統計学を勉強したか」ということに大きく関係すると考えられる。4つめの両裾の確率の表は、これは区間推定や両側検定のみを目的とした数表である。

このような確率の表が必要となるのは区間推定と仮説検定なので、調査した中では「統計学の一通りの入門」の書籍が約170冊あり、約170冊/650冊 \approx 26% \approx 約1/4を占めていた。

参考文献

- 大上丈彦 (2012). マンガでわかる統計学 素朴な疑問からゆる〜く解説, サイエンス・アイ新書, ソフトバンククリエイティブ.
- 京極一樹 (2012). ちょっとわかればこんなに役に立つ 統計・確率のほんとうの使い道, じっぴコンパクト新書, 実業之日本社.
- 国際規格 ISO 12491: 1997 (en), Statistical methods for quality control of building materials and components.
- 日本工業規格 JIS Z 8101-1 : 1999 (統計-用語と記号-第1部: 確率及び一般統計用語).
- 日本工業規格 JIS Z 8101-2 : 1999 (統計-用語と記号-第2部: 統計的品質管理用語).
- 日本工業規格 JIS Z 8101-3 : 2006 (統計-用語と記号-第3部: 実験計画法).
- 日本統計学会 編 (2013). 統計学, 日本統計学会公式認定 統計検定1級対応, 東京図書.
- 日本統計学会 編 (2012). 統計学基礎, 日本統計学会公式認定 統計検定2級対応, 東京図書.
- 日本統計学会 編 (2012). データの分析, 日本統計学会公式認定 統計検定3級対応, 東京図書.
- 日本統計学会 編 (2013). 資料の活用, 日本統計学会公式認定 統計検定4級対応, 東京図書.
- 花岡重利 (2013). 統計学テキストにおける分散のさまざまな表記について, 札幌学院大学 経済学部 平成25年度 卒業論文.

The Study on the Representation of the “Variance” in the Statistics Textbook

Nagatomo NAKAMURA¹

Shigetoshi HANAOKA²

Takahiro TSUCHIYA³

Abstract

Several designations and notations (symbols) for “variance” exist in the statistics books and textbooks which are used in liberal arts education at the universities in Japan. The variance is a basic term in statistics. For example, the symbol S^2 is used for the sample variance, V is for unbiased variance, and σ^2 is for population variance. The authors of the textbook use the symbols for sample variance and unbiased variance according to their own point of view. However, in the textbooks of basic science, such as physics and chemistry, the symbols are used in a unified way. So comparing with it, the use of symbols in the field of statistics is really peculiar. To examine the variation in the actual, we investigated about 650 books of a collection of privately owned and Sapporo Gakuin University library. In addition, and as a consideration for the beginners, we present a method which should lead them to obtain the “variance” in an easy and right way.

Keywords: Sample Variance, Unbiased Variance, Population Variance.

¹Department of Economics, Sapporo Gakuin University; nagatomo@sgu.ac.jp.

²Department of Economics, Sapporo Gakuin University.

³Department of Mathematics, Josai University; takahiro@josai.ac.jp.