# 統計学 6月30日 補足資料

全学共通科目「統計学」
担当: 社会情報科学部 山本 岳洋
t.yamamoto@sis-u-hyogo.ac.jp

## 1. 標本分散の期待値

母平均 $\mu$ ，母分散 $\sigma^2$の母集団からサンプルサイズ $n$ の標本 $X_1, \cdots, X_n$ を無作為復元抽出することを考えます． このとき，標本平均と標本分散は以下の式で定義されます．

標本平均:

$$\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

標本分散:

$$S^2 = \frac{1}{n}((X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2) \tag{1}$$

本補足資料は，標本分散の期待値 $E(S^2)$ が

$$E(S^2) = \frac{n-1}{n}\sigma^2$$

となること，すなわち母分散 $\sigma^2$ より小さな値となることを示すことが目的です． 本資料を理解することができれば，母集団と標本を考える上でややこしい， 確率変数・期待値・分散などの諸概念を整理できたと言えるでしょう． また，なぜ Excelで `VAR()` を使うと標本分散ではなく標本不偏分散が計算されるのかについても理解が深まります．

## 2. 準備

### 標本 $X_1, \cdots, X_n$ の持つ性質

母集団から要素を1つづつ無作為復元抽出しているので，$X_1, \cdots, X_n$はすべて同じ母集団分布に従います． また，復元抽出のため $X_1, \cdots, X_n$ は互いに独立です． このように，確率変数 $X_1, \cdots, X_n$ が同じ分布に従い，かつ互いに独立であるようなとき， $X_1, \cdots, X_n$は **独立同分布である**，もしくは **i.i.dである** と表現したりします．

さて，$X_1, \cdots, X_n$はすべて同じ母集団分布に従い，かつ互いに独立なので，$X_i$ $(1 \leq i \leq n)$ について以下の性質が成り立ちます．

$$
\begin{aligned}
E(X_i) &= \mu \\
V(X_i) &= \sigma^2
\end{aligned}
$$

つまり，母平均 $\mu$ と母分散 $\sigma^2$ がそのまま，標本として抽出した各要素 $X_i$ の期待値と分散になります．

## 標本平均 $\bar{X}$ の性質

いま求めたいものは標本分散の期待値ですが，後ほど使うため標本平均の性質についてもまとめておきます．（導出は講義資料参照．この導出および下記に示す結果はとても大事です）

$$
\begin{aligned}
\text{標本平均の期待値}: E(\bar{X}) &= \mu \\
\text{標本平均の分散}: V(\bar{X}) &= \frac{\sigma^2}{n}
\end{aligned}
$$

**標本平均の期待値** や **標本平均の分散** を求めていることに注意してください．たとえば，「標本平均の分散」と「標本分散」は全く異なる概念です．

また，本資料では必要ありませんが，標本平均 $\bar{X}$ の標準偏差

$$
\sqrt{V(\bar{X})} = \frac{\sigma}{\sqrt{n}}
$$

は **標準誤差（Standard Error）** と呼ばれ，区間推定や仮説検定でも出てくる非常に重要な統計量であることを講義中に説明しました．

# 3. $E(S^2) = \frac{n-1}{n}\sigma^2$ の導出

標本分散 $S^2$ の期待値 $E(S^2)$ にまずは式(1)を代入すると，

$$
\begin{aligned}
E(S^2) &= E(\frac{1}{n}\{(X_1-\bar{X})^2+\cdots+(X_n-\bar{X})^2\}) \\
&= \frac{1}{n}E(\{(X_1-\bar{X})^2+\cdots+(X_n-\bar{X})^2\}\,(\because E(aX)=aE(X)) \\
&= \frac{1}{n}E(X_1^2+\cdots+X_n^2-(2\bar{X}X_1+\cdots+2\bar{X}X_n)+\bar{X}^2+\cdots+\bar{X}^2) \\
&= \frac{1}{n}E(\sum_i X_i^2-\sum_i 2\bar{X}X_i+\sum_i \bar{X}^2) \qquad (2)
\end{aligned}
$$

ここで，$\sum_i X_i = n\bar{X}$ より 期待値 $E(\cdot)$ の中身の第二項は

$$
\begin{aligned}
\sum_i 2\bar{X}X_i &= 2\bar{X}\sum_i X_i \\
&= 2\bar{X}n\bar{X} \\
&= 2n\bar{X}^2
\end{aligned}
$$

と表すことができます．これを式(2)に代入すると，

$$
\begin{aligned}
E(S^2) &= \frac{1}{n}E(\sum_i X_i^2 - \sum_i 2\bar{X}X_i + \sum_i \bar{X}^2) \\
&= \frac{1}{n}E(\sum_i X_i^2 - 2n\bar{X}^2 + n\bar{X}^2) \\
&= \frac{1}{n}E(\sum_i X_i^2 - n\bar{X}^2) \\
&= \frac{1}{n}\sum_i E(X_i^2) - \frac{1}{n}E(n\bar{X}^2) \left(\because \text{期待値の加法性 } E(X+Y) = E(X) + E(Y)\right) \\
&= \frac{1}{n}\sum_i E(X_i^2) - E(\bar{X}^2) \, (\because E(aX) = aE(X)) \qquad (3)
\end{aligned}
$$

さて，式(3) には，$E(X_i^2)$ と $E(\bar{X}^2)$ があらわれています．まずは，$E(X_i^2)$ について考えてみましょう． これは，標本の1つの要素$X_i$ の二乗の期待値ですので，$E(X_i^2)$ が式にでてくるであろう，$X_i$ の 分散 $V(X_i)$ を考えてみると，

$$
\begin{aligned}
V(X_i) &= \sigma^2 \qquad (3) \\
V(X_i) &= E(X_i^2) - E(X_i)^2 \\
&= E(X_i^2) - \mu^2 \, (\because E(X_i) = \mu) \qquad (4)
\end{aligned}
$$

従って，式(3)，(4)より，$E(X_i^2)$ は

$$
E(X_i^2) = \sigma^2 + \mu^2 \qquad (5)
$$

と表すことができることが分かります.

次に，$E(\bar{X}^2)$ について考えてみましょう．こちらも，いまと同様に 標本平均 $\bar{X}$ の分散 $V(\bar{X})$ を考えてみると，

$$
\begin{aligned}
V(\bar{X}) &= \frac{\sigma^2}{n} \qquad (6) \\
V(\bar{X}) &= E(\bar{X}^2) - E(\bar{X})^2 \\
&= E(\bar{X}^2) - \mu^2 \, (\because E(\bar{X}) = \mu) \qquad (7)
\end{aligned}
$$

従って，式(6)，(7)より

$$
E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2 \qquad (8)
$$

とあらわすことができます．式(5)，(8)を式(3)に代入すると，

$$
\begin{aligned}
E(S^2) &= \frac{1}{n}\sum_i E(X_i^2) - E(\bar{X}^2) \\
&= \frac{1}{n}((\sigma^2+\mu^2)+\cdots+(\sigma^2+\mu^2)) - (\frac{\sigma^2}{n}+\mu^2) \\
&= \frac{1}{n}(n\cdot\sigma^2+n\cdot\mu^2) - (\frac{\sigma^2}{n}+\mu^2) \\
&= \sigma^2+\mu^2-(\frac{\sigma^2}{n}+\mu^2) \\
&= \sigma^2-\frac{1}{n}\sigma^2 \\
&= \frac{n-1}{n}\sigma^2
\end{aligned}
$$

となり，

$$
E(S^2) = \frac{n-1}{n}\sigma^2 \tag{9}
$$

を示すことができました．

# 4. 標本不偏分散

いま示したように，標本分散 $S^2$ の期待値 $E(S^2)$ は母分散より少し小さな値となります．サンプルサイズ $n$ が大きければ $\frac{n-1}{n}$ はほとんど1と考えてもよいですが，$n$ が小さいときはそのような考えを使うことはできません．実際， $n=10$ とすれば $\frac{n-1}{n}\sigma^2 = \frac{9}{10}\sigma^2$ となり約1割も母分散から離れた値が期待値となってしまいます． 従って，標本分散の代わりに，以下に示す **標本不偏分散** $S_{\mathrm{u}}^2$ という分散が標本の分散を計算する際に一般的に用いられます．Excelで `VAR()` を用いて分散を計算するときも以下の式が用いられます．

$$
S_{\mathrm{u}}^2 = \frac{1}{n-1}((X_1-\bar{X})^2+\cdots+(X_n-\bar{X})^2)
$$

この標本不偏分散 $S_{\mathrm{u}}^2$ は，その期待値 $E(S_{\mathrm{u}}^2)$ が

$$
E(S_{\mathrm{u}}^2) = \sigma^2
$$

となり母分散 $\sigma^2$ と一致します．このように，標本から計算されるある値の期待値が母集団のある性質と一致する性質を不偏性と言います．不偏分散という名前は，この不偏性が母分散に対して成り立つところからきています．導出は今後課題として出す予定です，式(9)の結果を利用すると容易に示すことができます．