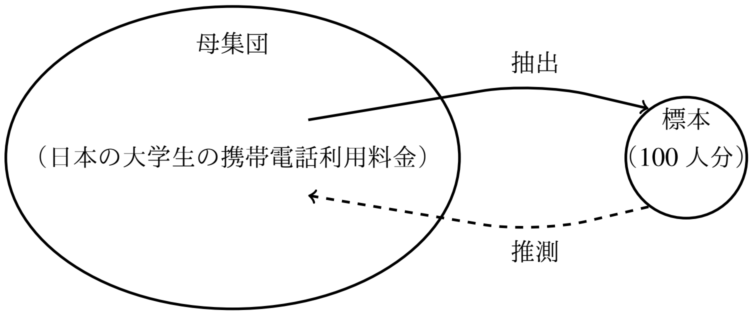
応用計量経済分析 TA セッション
第 1 回：平均・分散・標準偏差

TA：北村友宏 [*]

2015 年 10 月 6 日

1　統計的推測

- 調査対象の集団全体を**母集団**（population）とよぶ。
  - ⋆ e.g., 日本全国の大学生の 1 か月間の携帯電話利用料金（全員を調べるのは大変）
- 母集団の中から個体を抽出して集めたものを**標本**（sample）とよぶ。
  - ⋆ e.g., 日本全国の大学生のうち、ある 100 人の 1 か月間の携帯電話利用料金
- 標本から母集団の特徴を推測することを**統計的推測**（statistical inference）とよぶ。
  - ⋆ e.g., 日本の大学生 100 人の 1 か月間の携帯電話利用料金のデータを用い、「日本全国の大学生は 1 か月間に携帯電話利用料金が何円程度かかっているのか、また、金額には人によってどの程度のバラつきがあるのか」を推測する。



2　統計量

- 標本の関数を**統計量**（statistic）とよぶ。
  - ⋆ e.g., 標本平均、標本分散など
- 標本に含まれる個体を**観察値**（observation）とよぶ。

[*] Email: kitamu.tom@gmail.com　URL: http://tomkitamura.html.xdomain.jp

2.1 平均

- n 個の観察値 $(x_1, x_2, \ldots, x_n)$ を観測したとき、**算術平均**（arithmetic average）は、

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}(x_1 + x_2 + \cdots + x_n).$$

★「標本平均」（統計量の 1 つ）として、母集団の平均の推定に使われる（統計的推測の 1 例）。

例題 1. $\displaystyle\sum_{i=1}^{n}(x_i - \overline{x}) = 0$ を証明しなさい。

（証明）

$$\sum_{i=1}^{n}(x_i - \overline{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n}\overline{x} = \sum_{i=1}^{n} x_i - n\overline{x}.$$

ここで、$\overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i$ なので、$n\overline{x} = \displaystyle\sum_{i=1}^{n} x_i$ である。よって、

$$\sum_{i=1}^{n} x_i - n\overline{x} = n\overline{x} - n\overline{x} = 0. \quad \text{（証明終）}$$

※証明するときは、始まりと終わりを必ず明記すること。「（証明）〜（証明終）」や"*Proof.*〜□"など。

$\boxed{\text{この式から分かること}}$ 算術平均からの差（**偏差**、deviation）を、標本内の全ての個体について合計すると 0 になる。

例題 2. $\displaystyle\frac{1}{n}\sum_{i=1}^{n}(ax_i + b) = a\overline{x} + b$ を証明しなさい。

（証明）

$$\frac{1}{n}\sum_{i=1}^{n}(ax_i + b) = \frac{1}{n}\sum_{i=1}^{n} ax_i + \frac{1}{n}\sum_{i=1}^{n} b = \frac{1}{n}\cdot a\sum_{i=1}^{n} x_i + \frac{1}{n}\cdot nb = a\cdot\frac{1}{n}\sum_{i=1}^{n} x_i + b = a\overline{x} + b. \quad \text{（証明終）}$$

$\boxed{\text{この式から分かること}}$ 標本の全ての観察値を a 倍して b を加えたものの算術平均は、もとの算術平均を a 倍して b を加えたものに等しい。

2.2 分散

- n 個の観察値 $(x_1, x_2, \ldots, x_n)$ を観測したとき、**分散**（variance）は、

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 = \frac{1}{n-1}\left\{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2\right\}.$$

★ 観察値のバラつきの程度を表す。

★ 単位は観察値の 2 乗。

★ $s^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2$ と定義する場合もあるが、$n-1$ で割ったバージョンで定義すれば母集団の分散 σ^2 を偏りなく推定できる。つまり、$E\left(\dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i - \overline{x})^2\right) = \sigma^2.$

★ σ^2 の推定も統計的推測の 1 例（s^2 を統計量の 1 つである「標本分散」として推定に用いる）。

2.3　標準偏差

- n 個の観察値 $(x_1, x_2, \ldots, x_n)$ を観測したとき、**標準偏差**（standard deviation）は、

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2} = \sqrt{\frac{1}{n-1}\left\{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2\right\}}.$$

  - ★ 分散の平方根。
  - ★ 分散とは違い、単位はもとの観察値と同じ。
  - ★ これも統計量の 1 つ（「標本標準偏差」として、母集団の標準偏差の推定に使われる）。
  - ★ $s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2}$ と定義する場合もある。

※今回登場した平均や分散は、確率変数の平均 $E(X)$ や分散 $V(X)$ とは別物。混同しないように！