統計

言葉の定義

母集団 ： 統計の対象となる集まり

例 ・20歳の日本人全体

・ある工場で作られる製品全体

・ある場所の1年間の天気

個体 ： 母集団を構成する個々のもの

特性標識 ： 着目する性質

⟶ 統計学では着目する特性について母集団全体の特徴を調べたい

特に次のものを考える

母平均 ： 母集団の平均値

例 20歳の日本人の平均身長

母分散 ： 母集団の分散値

母比率 ： 条件Cをみたす確率

例 ・工場の製品の不良品の割合

・身長が160cm以上165cm以下の人　　など

このような母集団の特徴を母分布という

→ これらの値を求めるにはどうすればいいか.

方法1. 全数調査：全部調べる

→ 時間と費用がかかりすぎる.

方法2. 標本(サンプル)調査：一部を取り出して調べる

→ 標本の数が多い. 時間と費用がかかる.

標本の数が少い 信頼性に問題

・標本の数のことを標本の大きさ(サイズ)という

・標本を取り出すときは. 無作為に個体をとり. 特性を調べた後.
その個体をもとに戻す(復元抽出). これを何回もくり返す.
このような方法を独立な任意抽出という.

標本変量

特性 X が数値で表されるとき, 1回目に取り出される標本の特性を X_1 とする

→ X_1 の値は母分布に従うので, X_1 は確率変数のように扱える

例. 製品の不良品の割合が3%とする ← 母分布

1つ取り出したとき. X_1 を不良品なら1, そうでなければ0とするとき.

X_1 のとる値と確率は

X_1	0	1
確率	0.97	0.03

となる

これをくり返すと、母分布に従う独立な確率変数

$X_1, \cdots, X_n$ を得る。これらを**標本変量**といい、

$(X_1, \cdots, X_n)$ を**標本**という。

実際に抽出を行うと、$X_1, \cdots, X_n$ には具体的な数 $x_1, \cdots, x_n$ が得られる。

この $(x_1, \cdots, x_n)$ を $(X_1, \cdots, X_n)$ の**実現値，標本値**という。

標本 $(X_1, \cdots, X_n)$ に対し、次の量を定義する。

標本平均 ： $\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$

標本分散 ： $S^2 = \frac{1}{n-1}\{(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2\}$

$= \frac{1}{n-1}\{X_1^2 + X_2^2 + \cdots + X_n^2 - n\overline{X}^2\}$

標本標準偏差 ： $S = \sqrt{S^2}$

これらも確率変数として扱える。

例. 標本 (3.2, 1.5, 3.5, 2.8, 3.0) の平均 $\overline{X}$ と分散 S^2 の実現値を求めよ

答 $\overline{X} = \frac{1}{5}(3.2 + 1.5 + 3.5 + 2.8 + 3.0) = 2.8$

$S^2 = \frac{1}{5-1}\{3.2^2 + 1.5^2 + 3.5^2 + 2.8^2 + 3.0^2 - 5 \times 2.8^2\} = 0.595$ である

問題 標本 (1.67, 1.72, 1.70, 1.72, 1.69) の平均 $\overline{X}$ と分散 S^2 の実現値を求めよ

答 $\overline{X} = \frac{1}{5}(1.67 + 1.72 + 1.70 + 1.72 + 1.69) = 1.70$

$S^2 = \frac{1}{4}(0.03^2 + 0.02^2 + 0 + 0.02^2 + 0.01^2)$

$= \frac{1}{4}(0.0018) = 0.00045$

定理 2.5.1 母平均をμ，母分散をσ^2とすると．

標本$(X_1, \cdots, X_n)$の標本平均$\bar{X}$は

(i) $E(\bar{X}) = \mu$

(ii) $V(\bar{X}) = \frac{\sigma^2}{n}$ をみたす．

∵ 1つの標本変量X_iは母分布に従うので．

$E(X_i) = \mu$，$V(X_i) = \sigma^2$ である．∴ 定理1.7.2より求まる．

定理 2.5.2．上の条件で．nが十分大きければ．

$\bar{X}$はだいたい $N(\mu, \frac{\sigma^2}{n})$ に従う．

定理2.5.5 上の条件で

$E(S^2) = \sigma^2$ をみたす

$$
\begin{aligned}
∵ E(S^2) &= E\left(\frac{1}{n-1}(X_1^2 + X_2^2 + \cdots + X_n^2 - n\bar{X}^2)\right) \\
&= \frac{1}{n-1}\{E(X_1^2) + E(X_2^2) + \cdots + E(X_n^2) - n \cdot E(\bar{X}^2)) \\
&= \frac{1}{n-1}\{(\sigma^2 - \mu^2) \times n - n \cdot (\frac{\sigma^2}{n} - \mu^2)\} \\
&= \sigma^2
\end{aligned}
$$

標本度数と標本比率

標本として得られる結果を、k個の階級 $C_1, C_2, \cdots, C_k$ に分類するとする。

各階級 C_i に属する結果の個数 N_i をその階級の標本度数という

標本が n 個の場合

$N_1 + N_2 + \cdots + N_k = n$ になる。

また、標本の中で各階級が占める比率

$P_i = \dfrac{N_i}{n}$ を標本比率という。

$P_1 + P_2 + \cdots + P_k = 1$ をみたす。

これらを表に整理したものを度数分布表という。

階級	C_1	C_2	C_3	$\cdots$	C_k
標本度数	N_1	N_2	N_3	$\cdots$	N_k
標本比率	P_1	P_2	P_3	$\cdots$	P_k

例. 100人のクラスの身長を度数分布表にしたとすると、例えば…

階級	150cm以下	150〜160	160〜170	170〜180	180以上
標本度数	3	28	43	21	5
標本比率	0.03	0.28	0.43	0.21	0.05

となる