統計

まず、言葉の定義をいくつかする.

母集団：統計の対象となる集まり

個体　：母集団の中の1つずつ

特性標識：着目する性質

⟶ 統計学では着目する性質について.母集団全体の特徴を調べたい.

特に次が重要.

母平均：母集団の平均値

母分散：母集団の分散値

母比率：条件Cをみたす確率

例 製品が不良である割合など.

母分布：上記のような母集団の特徴

⟶ これらを調べるにはどうしたらよいか？

方法1. 全数調査：全部調べる ⟶ 時間と金がかかる.ただ正確.

方法2. 標本(サンプル)調査：1部を取り出して調べる.

⟶ 信頼性と.時間・費用のバランスが大事.

標本の大きさ：標本の数のこと.

標本を調べるときは.無作為に個体をとり、調べ、戻す(復元抽出)

これをくり返す方法を独立な任意抽出という.

標本変量

母分布がきまっているため.

特性Xが数値で表されるとき、1つ目の標本の特性をX_1とする.

2つ目をX_2 … とくり返すと、$X_1, \cdots, X_n$は独立な確率変数になる.

これらを**標本変量**という. $(X_1, \cdots, X_n)$を**標本**という.

実際に抽出を行うと$x_1, \cdots, x_n$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.

例. 1000本のくじの中に、当りがどのくらい入っているか調べたい.

当り(1)の割合をpとすると、はずれ(0)は$1-p$になるので.

1回目に引いたときの確率分布は

X_1	0	1
確率	$1-p$	p

になるはず.

実際にくじを引くと、当り.はずれがわかるので. $X_1 = 0$ or 1 になる.

この値がx_1である.

標本$(X_1, \cdots, X_n)$に対し、次を定義する.

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

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

$= \dfrac{1}{n-1}\left(X_1^2 + \cdots + X_n^2 - n \cdot \overline{X}^2\right)$

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 1

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

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

問題 2 3

2 $\overline{X} = \frac{1}{7}(1.3+1.5+2.5+1.8+2.0+1.0+1.1) = 1.6$

$S^2 = \frac{1}{7-1}(1.3^2+1.5^2+2.5^2+1.8^2+2.0^2+1.0^2+1.1^2-7\cdot 1.6^2) \fallingdotseq 0.29$

3 $\overline{X} = \frac{1}{5}(1.67+1.72+1.70+1.72+1.69) = 1.70$

$S^2 = \frac{1}{5-1}((-0.03)^2+(0.02)^2+0+(0.02)^2+(-0.01)^2) = 0.00045$ である

定理 2.5.1. 母平均 μ, 母分散 σ^2 とすると. 標本平均 $\overline{X}$ は

(1) $E(\overline{X}) = \mu$, (2) $V(\overline{X}) = \frac{\sigma^2}{n}$ である

∵ 1つの標本 X_i は $E(X_i) = \mu$, $V(X_i) = \sigma^2$ より 定理1.7.2 から求まる.

定理 2.5.2. n が十分大きければ.

$\overline{X}$ はほぼ. $N(\mu, \frac{\sigma^2}{n})$ に従う.

定理 2.5.3.

$E(S^2) = \sigma^2$ である.

$V(X_n) = E(X_n^2) - E(X_n)^2$ より

$E(X_n^2) = \sigma^2 + \mu^2$

∵ $E(S^2) = E\left(\frac{1}{n-1}(X_1^2+\cdots+X_n^2-n\overline{X}^2)\right)$

$\frac{\sigma^2}{n} + \mu^2$

$= \frac{1}{n-1}(E(X_1^2)+\cdots+E(X_n)^2 - n\cdot E(\overline{X}^2))$

$= \frac{1}{n-1}(n\cdot(\sigma^2+\mu^2) - n\cdot(\frac{\sigma^2}{n}+\mu^2))$

$= \sigma^2$ である

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.

各階級C_iに属する個数N_iをその階級の標本度数という.

標本がn個なら　$N_1 + \cdots + N_k = n$　である.

また、標本の中で各階級が占める比率

$$P_i = \frac{N_i}{n}$$ を標本比率という.

これは、$P_1 + \cdots + P_k = 1$　をみたす

これらを表にしたものを度数分布表.

グラフにしたものを、ヒストグラム・という.

また、各級に至るすべての度数を足したものを累積度数という.