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.

例：20歳の日本人全体，ある工場で作られる製品全体　など.

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.

特に次のものが重要.

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：製品の不良品の割合，160cm以上165cm以下の割合　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すればよいか.

方法1．全数調査：全部調べる.

→時間と費用がかかる

方法2　標本(サンプル)調査：一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.

・標本の数のことを標本の大きさ(サイズ)という.

・標本を調べるときは無作為に個体をとり．調べ、元に戻す.(復元抽出)

これをくり返す．この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つ目の標本の特性をX_1とする。

→ X_1の値は母分布に従うので、X_1は確率変数として扱える。

例. 製品の不良率が3%であるとする。Xを不良(1)、良(0)とするとき、

X_1のとる値と確率は、

X_1	1	0
確率	0.03	0.97.

とできる。これは離散型の確率変数。

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、$(X_1, \cdots, X_n)$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。

標本$(X_1, \cdots, X_n)$に対し、次を定義する。

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

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

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

標本標準偏差 : $S = \sqrt{S^2}$

これらも、確率変数として扱える。

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

$S^2 = \frac{1}{5-1}\left(3.2^2 + 1.5^2 + 3.5^2 + 2.8^2 + 3.0^2 - 5 \times 2.8^2\right) = 0.595$ 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 $\overline{X}=\frac{1}{5}(1.67+1.72+1.70+1.72+1.69)=1.70$

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

$=\frac{1}{4}(0.0018)=0.00045$ である.

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

(i) $E(\overline{X})=\mu$ (ii) $V(\overline{X})=\frac{\sigma^2}{n}$ である.

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

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2.5.5.

$E(S^2)=\sigma^2$ である.

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

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

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

$=\sigma^2$.

$V(X_i)=E(X_i^2)-E(X_i)^2$ より

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

P53,54の表、グラフをみるとわかりやすい.

問題. P55. 問1を解け.