09/20

参考文献

実験計画法入門：鷲尾泰俊、日本規格協会、2700 円、ISBN 4-452-50330-5、1997

入門統計解析法：永田靖、日科技連、3045 円、ISBN 4-8171-0266-7、1992

1. 実験計画法と統計解析

（学習内容）

・実験計画法とは何か

・統計解析の目的について理解する

1.1 実験計画法とは

実験：理論・仮説の検証、条件と結果の因果関係の把握

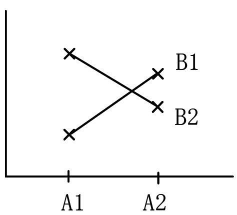
どのような実験を行ったらいいか？　実験データをどう取り扱えばいいか？

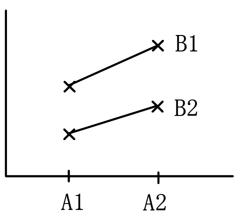
（例）

ある化学製品の製造において、強度を高めるために反応温度と触媒量の最適な条件を知りたい。

反応温度（A）：200℃　300℃　…

触媒量（B）：　5%　6%　…





	B1	B2
A1	○	○
A2	○	○

Aの効果、Bの効果それぞれについて2つずつデータ

Bを固定して実験、つぎに
Aを固定して実験

	B1	B2
A1	○	
A2	○	○

Aの効果、Bの効果について
２つずつデータを得るには
3x2=6回の実験が必要

・実験目的に対してどのような実験が効果的か（少ない実験回数で、正しい情報を得る）

・誤差を含む実験データに対してどのように解析すればよいか

　→　実験計画法

実験計画法（DOE: design of experiments）：

1920 年代　イギリス　Ronald A.Fisher（遺伝学者、統計学者）により創始

農業試験場にて、種子の品種による収穫の差異を研究。種子、土地、気候、肥料など多くの要素が影響する実験を、統計解析に基づき効率的に行う実験方法を考案。

1.2 データと統計解析

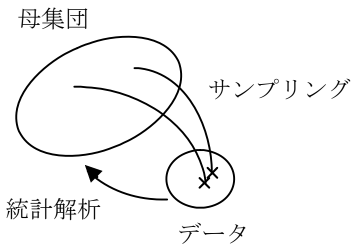
(1) 母集団とデータ

（例）ガラス工場で 100 枚のガラス板製品の板厚を計測した。

　　　大学生 100 人に好きな TV 番組を質問した。

・サンプリング（sampling）：データを取ること

・母集団（population）：サンプリング対象となる集団（要素の集まり）



＊ データの取得の目的は、母集団についての情報を得ること

...「（このガラス工場で）（機械Ａで）製造されたガラス板の板厚の平均は 7.15mm である」

「番組 A は（大学生に）（工学部学生に）人気がある番組である」

母集団の構成要素が等しい確率で現れるようなサンプリングが必要　→ランダムサンプリング

統計解析：データから母集団に関する情報を推測する

・計量値：長さ、重さのような連続量のデータ

・計数値：不良品の数のように個数を数えて得られるデータ

（例）板厚 7.15mm→計量値

板厚 7.10～7.20mm のガラス板の数→計数値

(2) 統計量、統計値（statistic）

：母集団に関する情報を推測するためにデータから計算される量（値）

データ数 n 個　x_1, x_2, x_3, …, x_n

・平均　$\bar{x}=(x_1+x_2+\cdots+x_n)/n=\sum_{i=1}^{n}x_i/n$

・メジアン（中央値）：データを大きさの順に並べた中央値

・平方和　$S=\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum(x_i^2-2\bar{x}x_i+\bar{x}^2)=\sum x_i^2-2\bar{x}\sum x_i+n\bar{x}^2$

$$=\sum x_i^2-2(\sum x_i)^2/n+n(\sum x_i/n)^2=\sum_{I=1}^{n}x_i^2-(\sum_{i=1}^{n}x_i)^2/n$$

・分散　V=S/(n-1)

$(x_1-\bar{x}),(x_2-\bar{x}),\cdots(x_{n-1}-\bar{x})$　があれば　$(x_n-\bar{x})$　は自動的に求まる。情報量は n-1

・標準偏差　$s=\sqrt{V}$

・範囲　$R=x_{max}-x_{min}$

(3) 母集団に関する情報

・母集団分布（population distribution）：母集団を構成する要素のばらつき

→確率として考える

・確率（probability）：ある事象 A が起こる確からしさ　Pr{A}

・確率変数（random variable）：どのような値をとるか確率が決まっている変数 x

計量値→連続型確率変数、計数値→離散型確率変数

（例）サイコロをふって出る目を x　　$Pr\{x=1\}=1/6$,　$Pr\{x=2,4,6\}=1/2$

データはサンプリングによって得られた確率変数の実現値である

どのようなデータが得られるか ＝ 確率変数がどのような値をとるか

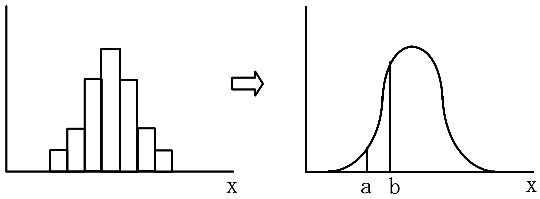
・確率分布（probability distribution）: 確率変数がどのような値を取るかという法則性

・確率密度関数（probability density function）：

連続型確率変数 x がとる値の確率分布を関数 $f(x)$ で表したもの

ヒストグラムの総面積を１とし、区間幅→0 としたもの

$\int f(x)dx=1$



・母平均（population mean）μ ：母集団の平均値

期待値（実現値を無限個得たときの平均）で定義できる

x が離散型確率変数のとき　　$\mu=E(x)=\Sigma x_i f_i$

x が連続型確率変数のとき　　$\mu=E(x)=\int xf(x)dx$

・母分散（population variance）σ^2

期待値を用いて次式で定義される

$\sigma^2=V(x)=E[(x-E(x))^2]$

$=E[x^2-2xE(x)+(E(x))^2]=E(x^2)-2\mu E(x)+\mu^2=E(x^2)-\mu^2=E(x^2)-E(x)^2$

・母標準偏差（population standard deviation）σ

$\sigma=\sqrt{V(x)}$

（期待値の性質）

x を確率変数、a, b を定数とすると

$E(ax+b)=aE(x)+b$

（証明）$E(ax+b)=\int(ax+b)f(x)dx=a\int xf(x)dx+b\int f(x)dx=aE(x)+b$

→母数（population parameter）

母数と統計量の区別

＊ 統計解析では得られたデータから母集団の確率分布の特性を得る