

第1部 統計学の基礎と

基本的な解析手法

1. 統計学とは
2. 統計学の基本
 - 2.1 母集団とサンプル(標本)
 - 2.2 データ(data)
3. 集団の特性を示す統計量
 - 3.1 統計量(statistic)とは
 - 3.2 集団を代表する統計量 — 平均値など
 - 3.3 集団のばらつきを表す値 — 平方和、分散、標準偏差
4. ばらつき(分布)を表す関数
 - 4.1 確率密度関数
 - 4.2 最も重要な正規分布
 - 4.3 統計量の分布と基本的特性

実験計画法を活用するために、知らなければならない統計学の基礎を確認・整理しよう。

第1部 統計学の基礎と 基本的な解析手法

5. 基本的な統計解析手法

5.1 統計学的仮説検定(Statistical hypothesis testing)

- (1) 仮説検定の種類
- (2) 仮説検定の手順
- (3) 仮説検定における2種類の誤り
- (4) 仮説検定の実際

5.2 統計的推定(Statistical Estimation)

- (1) 点推定
- (2) 区間推定

実験計画法を活用するために、知らなければ
ならない統計学の基礎を確認・整理しよう。

第1部 統計学の基礎と 基本的な解析手法

5. 基本的な統計解析手法

5.3 相関と回帰

(1) 相関

(2) 回帰

5.4 分散分析(ANOVA: analysis of variance)

実験計画法を活用するために、知らなければ
ならない統計学の基礎を確認・整理しよう。

1. 統計学とは

統計学とは — 記述統計学と推測統計学

バラツキのあるデータから、応用数学の手法を用いて数値上の性質や規則性あるいは不規則性を見いだす。

統計的手法は、実験計画、データの要約や解釈を行う上での根拠を提供する学問であり、幅広い分野で応用されている。

記述統計学 (descriptive statistics)

記述統計学とは、収集したデータの要約統計量(平均、分散など)を計算して分布を明らかにする事により、データの示す傾向や性質を知ること。
母集団、標本という概念はない。

推測統計学 (inferential statistics)

標本データから、その性質を確率論的に推測する。
母集団は大きく、直接、調べられない、データは標本からしか得られないため、標本データから母集団を検定や推定などを行う。

2. 統計学の基本

2.1 母集団とサンプル

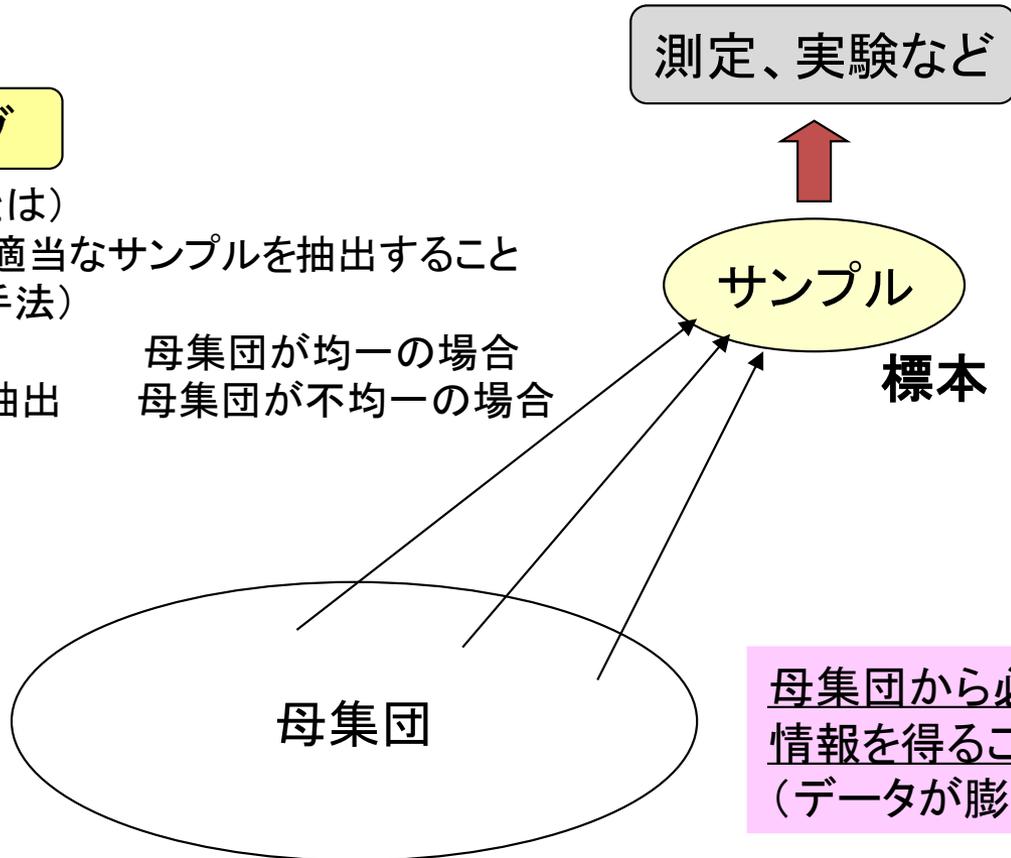
★★実験・解析の最重要ポイント

サンプリング

(サンプリングとは)
母集団から、適当なサンプルを抽出すること
(サンプリング手法)

- ・無作為抽出
- ・群に分けて抽出

母集団が均一の場合
母集団が不均一の場合



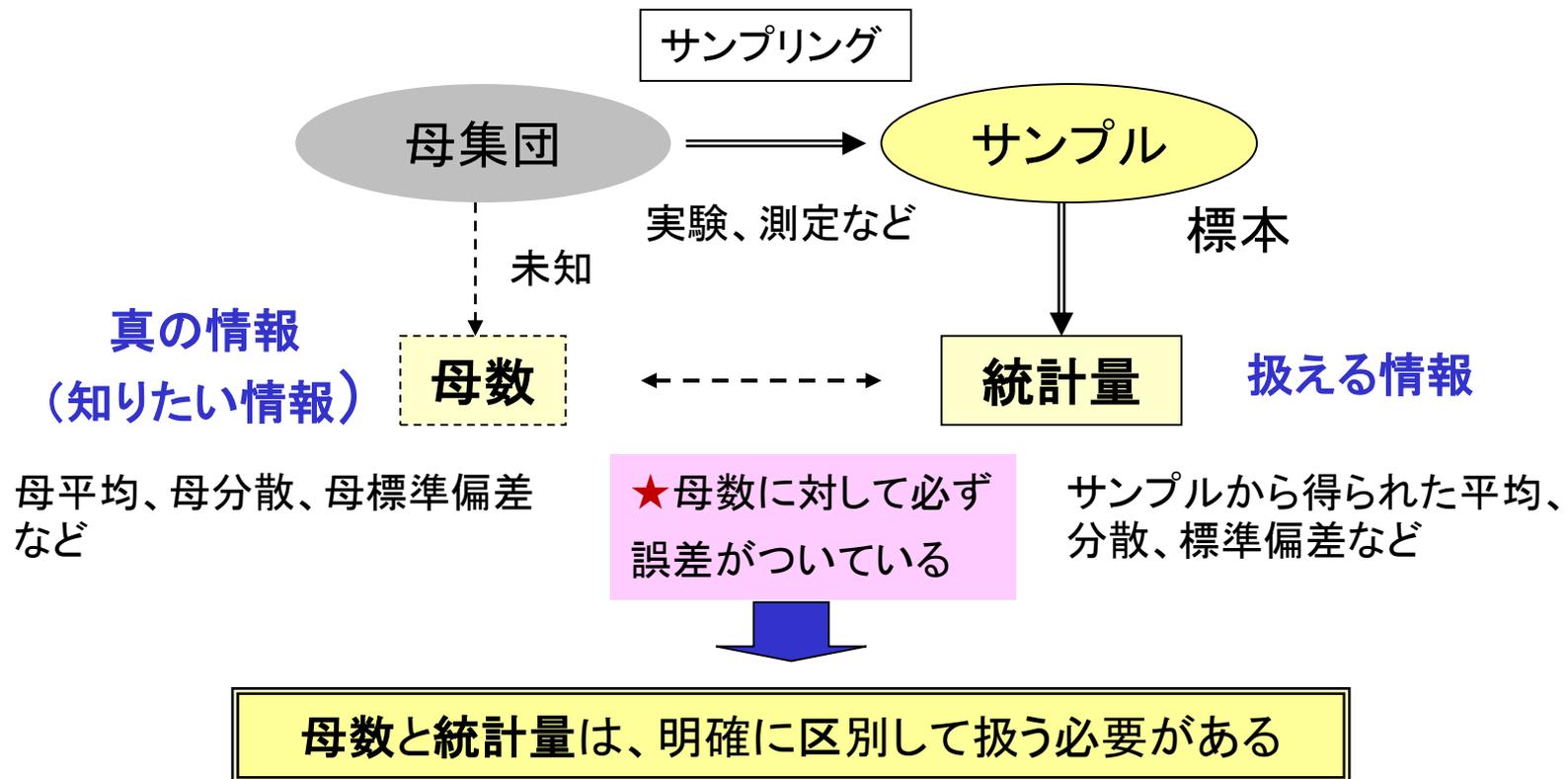
サンプルから
母集団の情報を得る

母集団から必要とする
情報を得ることは困難
(データが膨大など)

- ・母集団: 検討・考察の対象として妥当か?
- ・サンプル: 検討・考察の対象として母集団を適切に代表しているか?

2. 統計学の基本

2.1 母集団とサンプル



用語解説

母集団 (population)

- 1) 調査・研究の対象となる特性を持つすべてのものの集団。
 - 2) サンプルにより、処置をとろうとする集団。
- なお、母集団には無限母集団と有限母集団がある。

サンプル (標本: sample)

母集団からその特性を調べる目的を持ってとったもの。標本、試料という用語は同じ意味。

ランダムサンプリング (random sampling)

母集団を構成している単位体、単位量などがいずれも同じような確率でサンプルに入るようにサンプリングすること。

母数 (parameter)

- 1) 母集団分布の一族 $\{f(x; \theta_1, \theta_2, \dots, \theta_n)\}$ を考えるとき、その値を指定すれば分布が確定するような定数 $\theta_1, \theta_2, \dots, \theta_n$ 。
- 2) 広い意味では、確率分布によって定まる数値。
母平均、母分散、母標準偏差などは、全て母数。

統計量 (statistic)

サンプル (標本) から計算される値。サンプルから得られる平均、分散、標準偏差などは統計量である。

2. 統計学の基本

2.2 データ(data)

(1) データとは

基礎的な事実を示すもので、処理によって意味を持つ前提で収集されたもの。
対象とするもの → 数字(数値データ)、言葉(言語データ)、図、表

○データはそのままでは使えない 統計処理をして検討、判断などを行う。

(* 真理の解明、主張などは、データを通して行うのが基本)

(2) 数値データ

統計で活用するのは、主に数値データ

数値データの分類: 計量値と計数値

- ・**計量値**(continuous variable) 測定機器で測定する値
長さ、重量、時間、温度など 連続的な値
- ・**計数値**(discrete variable) 数えられる値
部品個数、不良点数など 離散的な値

(注)本講座では、主に、計量値を対象とする

3. 集団の特性を現す統計量

3.1 統計量 (statistic) とは

サンプルから得たデータだけでは、その集団の様子はわからない
データを処理(四則演算、平方根などを用いる)して集団の性質を表すものを**統計量**と言う

その統計量により判断、考察などを行う

3. 集団の特性を現す統計量

3.2 集団を代表する統計量 — 平均値など

集団を代表する統計量

— 異なる集団を比較する時、何で比較するか。

(1) 平均値

測定値の総和をサンプル数で割ったもの
(集団を代表する値として最も一般的)

$$\bar{x} = \sum \frac{x_i}{n}$$

(2) 中心値(メディアン)

測定値を大きさの順に並べた時、その中央に当たる値
(データが左右対称にばらついていない場合などに効果的)

(3) 中点値

測定値の最大値と最小値の算術平均

3. 集団の特性を現す統計量

3. 3 ばらつきの程度を表す値 — 平方和、分散、標準偏差

集団の測定値がどのようにばらついているか(広がっているか)

— 平均値だけで集団を見ると、勘違いが起きる。 (平均値の罠)

(1) 平方和 → (平均値 - 各測定値)の自乗の総和

$$S = \sum (x_i - \bar{x})^2$$

(2) **分散** → 平方和を測定値の数で割ったもの 単位は測定単位の自乗

$$\text{母分散: } \sigma^2 = \frac{\sum (x_i - \bar{x})^2}{n} \quad \text{不偏分散(平均平方): } V(s^2) = \frac{S}{\phi} = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

(母集団の分散、通常は未知)

(統計量: 母集団の分散の推定値)

(3) **標準偏差** → 分散の平方根 単位を測定値の単位に合わせる

$$\text{母標準偏差: } \sigma = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}} \quad \text{サンプル標準偏差: } s = \sqrt{\frac{S}{\phi}} = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

(母集団の標準偏差、通常は未知)

自由度: $\phi = n - 1$

注意

補足解説

自由度 (degrees of freedom)

サンプル数 $(n) - 1$ で考えてよい。

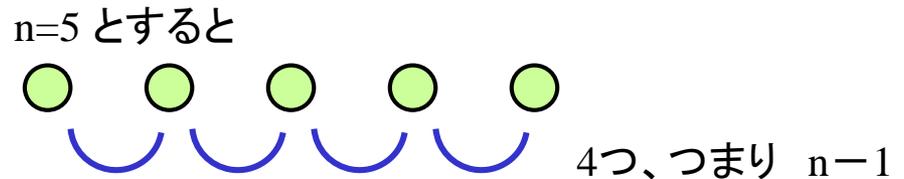
独立に変えることのできる変数の量。(誤差分散の個数に等しい。)

日本工業規格では、「カイ二乗分布、 F 分布、 t 分布などのパラメータ」と定義している。

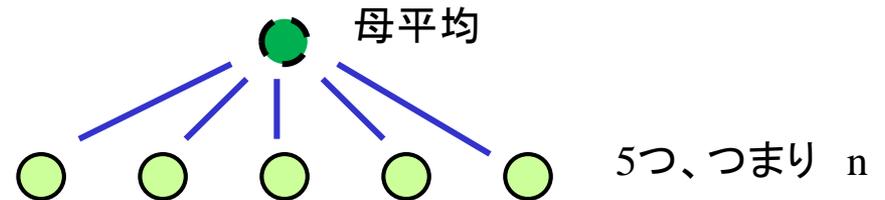
自由度とはどのような意味をもつか

ばらつきを作る個所の数

○ サンプルの情報だけから
ばらつきを考える場合



○ ばらつきの基準となるもの
(母平均) がわかっている場合



数式で考えると

$$\sum (x_i - \mu)^2 = \sum \{(x_i - \bar{x}) + (\bar{x} - \mu)\}^2 = \sum (x_i - \bar{x})^2 + \sum (\bar{x} - \mu)^2$$

期待値をとると

$$n\sigma^2 = E(S) + nV(\bar{x}) \quad V(\bar{x}) = \frac{\sigma^2}{n} \text{ であるから} \quad E(S) = (n - 1)\sigma^2$$

用語解説

母数、統計量、母数の推定量の記号一覧表(慣例)

サンプル数： n

	母数	母数の推定量	統計量	統計量の計算式
平均値	μ	$\hat{\mu}$	\bar{x}	$\bar{x} = \frac{\sum x_i}{n}$
中央値	—	—	\tilde{X} (メディアン)	—
分散	σ^2	$\hat{\sigma}^2$	$V(s^2)$ (不偏分散)	$V = s^2 = \frac{S}{\phi} = \frac{\sum (x_i - \bar{x})^2}{n-1}$
標準偏差	σ	$\hat{\sigma}$	s (標本標準偏差)	$s = \sqrt{V} = \sqrt{s^2}$
範囲	—	—	R	—

平方和： $S = \sum (x_i - \bar{x})^2$

自由度： $\phi = n - 1$

用語解説

確率変数 (random variable)

どのような値をとるかが、ある確率法則によって決まる変数。

とることができる値が離散的な連続的であるかによって、それぞれ離散型確率変数、連続型確率変数という。

確率変数は、大文字で表記する。

3. 集団の特性を現す統計量

【例題3-1】

ある軸部品について、サンプルを10個、取り出し、直径(mm)を測定したら、以下の値であった。

平均(母平均の推定値)、不偏分散(母分散の推定値)を計算せよ

No	直径(mm)
1	18.7
2	20.3
3	18.8
4	19.4
5	19.6
6	20.1
7	19.0
8	19.4
9	19.9
10	19.2

【解答3-1】

1) 平均値を求める

$$\bar{x} = \sum \frac{x_i}{n} = 19.4$$

(母平均と推定)

2) 不偏分散を求める

$$V(s^2) = \frac{S}{\phi} = \frac{\sum (x_i - \bar{x})^2}{n-1} = 0.29$$

(母分散と推定)

(エクセルの関数)

=AVERAGE(数値1,数値2)

=VAR.S(数値1,数値2)

4. ばらつき(分布)を表す関数

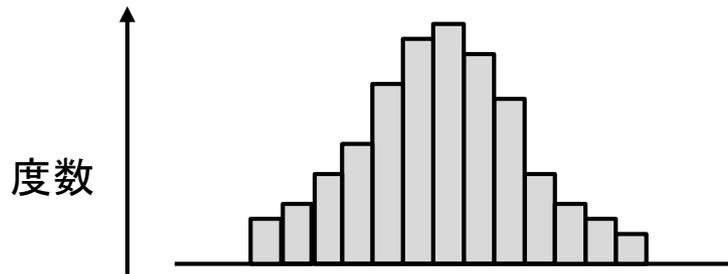
4.1 確率密度関数 - ばらつき(分布)の姿を表す関数

(1) 確率(probability)

事象が起きる確からしさ $0 \leq P \leq 1$

- ・事象が絶対に起きない $P=0$
- ・事象が必ず起きる $P=1$
- ・事象が2回に1回起きる $P=0.5$

(2) ばらつき(分布)の姿を現すヒストグラム



特性の最大値と最小値の間
を等区間(Δt)に分割

各区間に入る度数を棒グラフ化 → ばらつき(分布)の姿を可視化

(3) 確率密度関数(probability density function)

区間 $\Delta t \rightarrow 0$ にしたのが、確率密度関数: $f(x)$
変数 x が区間 $[a,b]$ に入る確率

積分しないと数値が出ない関数

$$P(a \leq x \leq b) = \int_a^b f(x)dx \quad \int_{-\infty}^{\infty} f(x)dx = 1$$

4. ばらつき(分布)を表す関数

4.2 最も重要な正規分布

正規分布

確率密度関数: $y = f(x) = \frac{1}{\sqrt{2\pi}\sigma} \times \exp[-\frac{1}{2} \times (\frac{x - \mu}{\sigma})^2]$

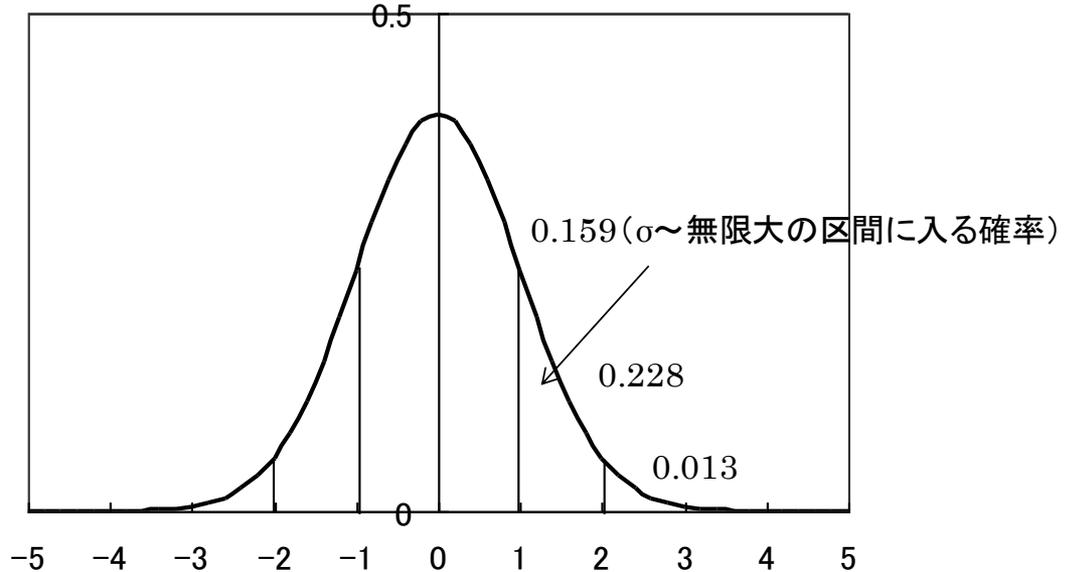
平均値 0 標準偏差 1 の正規分布

- ・計量値の分布
- ・偶然のばらつきの分布は正規分布となる

記号 $N(\mu, \sigma^2)$

標準正規分布: $N(0, 1^2)$

(平均値 0、標準偏差 1)



- ・平均値と分散が決まれば、分布が決まる。
- ・変数変換して、平均値:0、標準偏差:1とすれば、標準正規分布が適用できる。

↔	$-\sigma \sim \sigma$ の区間に入る確率	0.683	95%の範囲
↔	$-2\sigma \sim 2\sigma$ の区間に入る確率	0.954	-1.96σ~1.96σ
↔	$-3\sigma \sim 3\sigma$ の区間に入る確率	0.997	

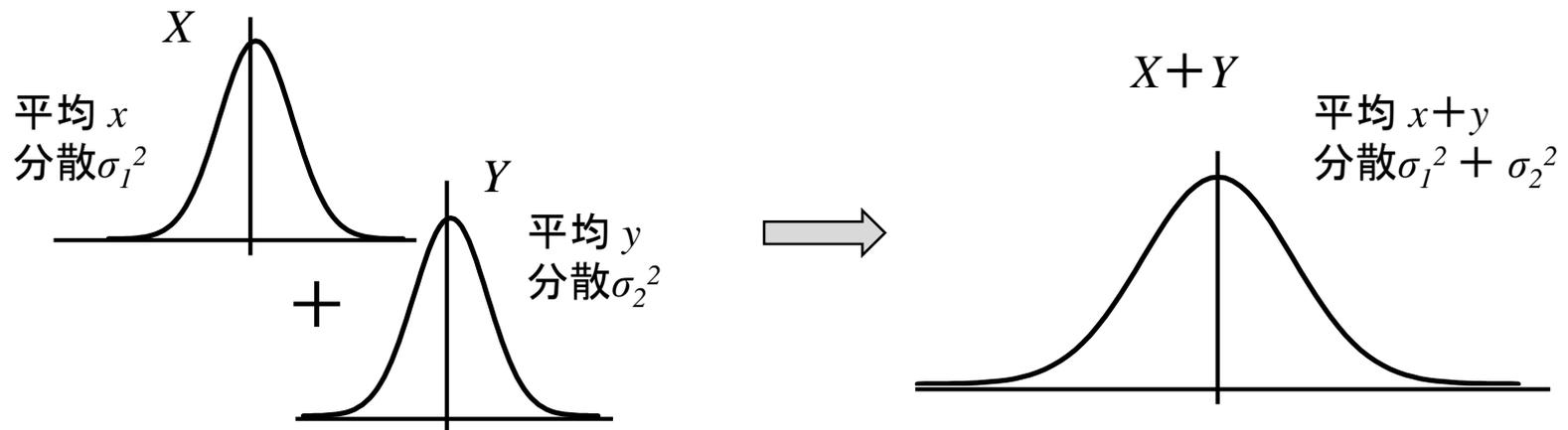
4. ばらつき(分布)を表す関数

4. 2 最も重要な正規分布

○独立な正規分布の平均値と分散の加法性 ー同時分布

独立に正規分布に従う X (平均 x , 分散 σ_1^2) と Y (平均 y , 分散 σ_2^2) について

確率変数 $Z=X+Y$ の分布は正規分布(平均 $x+y$, 分散 $\sigma_1^2 + \sigma_2^2$) に従う。



4. ばらつき(分布)を表す関数

4. 2 最も重要な正規分布

○独立な正規分布の平均値と分散の加法性 ー同時分布

(補足1)

独立に正規分布に従う X (平均 x , 分散 σ_1^2) と Y (平均 y , 分散 σ_2^2) について
確率変数 $Z=X-Y$ の分布はどうなるか。



正規分布(平均 $x-y$, 分散 $\sigma_1^2 + \sigma_2^2$)に従う。 (分散に引き算はない。)

(補足2)

部品A($X \pm a$)、部品($Y \pm b$)を組み合わせるとどうなるか。



$(X+Y) \pm (a+b)$ ではない。

$(X+Y) \pm \sqrt{a^2 + b^2}$ となる。

(ばらつきは分散(二乗)で加算される。)

4. ばらつき(分布)を表す関数

4.3 統計量の分布と基本的特性

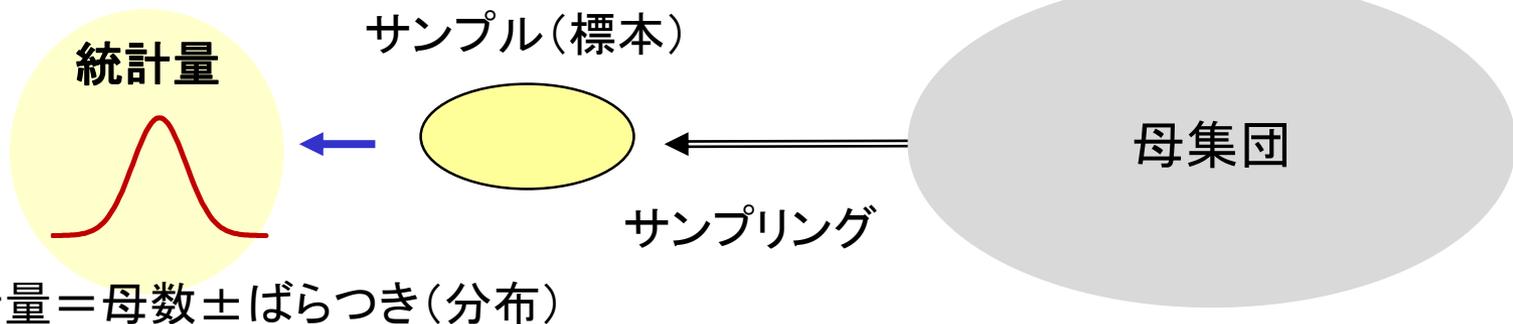
★統計解析の重要ポイント

母集団の母数(真の値)を調べる(検定・推定する)ためには、

サンプル(標本)の統計量进行分析する

➡ 統計量の分布を知ることが必要

母数(未知)



統計量の分布

→ 母数に対する統計量のばらつきの状況(分布)

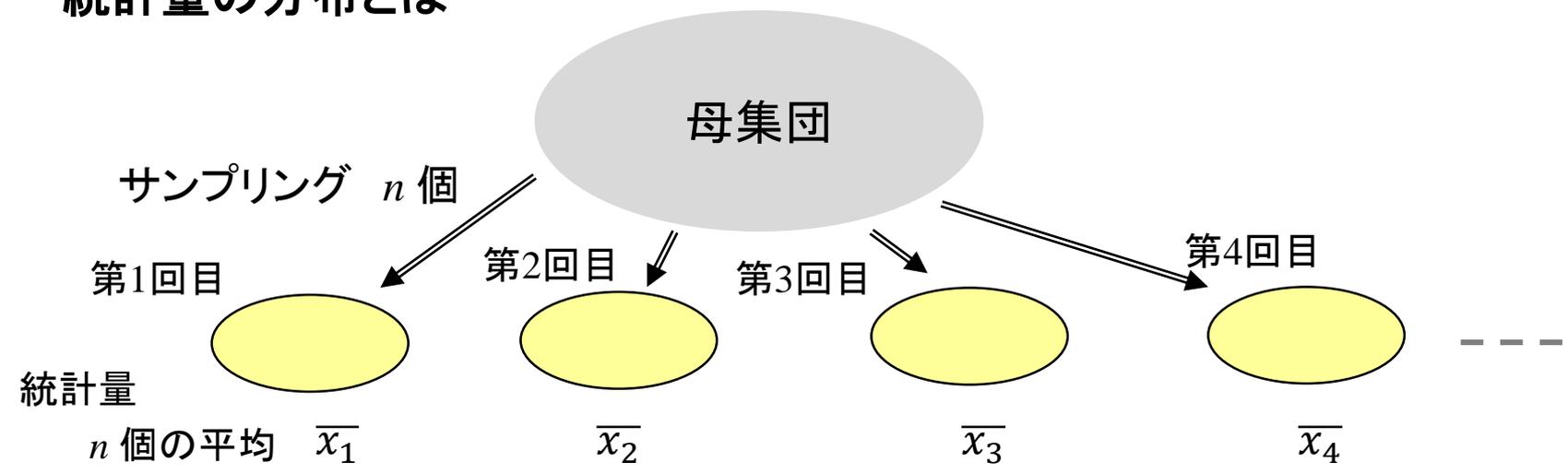
(前提) — ランダムな分布とランダムなサンプリング

- 母集団は正規分布をしている。
- 母集団は無限母集団で、サンプル(標本)はランダムサンプリングで抽出した。

4. ばらつき(分布)を表す関数

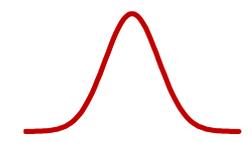
4.3 統計量の分布と基本的特性

統計量の分布とは



上記のサンプリングと統計量の計算を無限回、行った場合、

$\bar{x}_1 \dots \bar{x}_\infty$ ばらつき(分布)の様子 \Rightarrow **統計量(サンプル n の平均値)の分布**



サンプル数: n によって
分布の形状が変わる。

4. 3 統計量の分布と基本的特性

(1) サンプル平均値 (\bar{x}) の分布

母集団 (平均値: μ 、標準偏差: σ)

→ サンプルの大きさ n から得られる平均値 \bar{x} とその標準偏差の分布

・ サンプル平均値の平均値 ($\bar{\bar{x}}$) → μ (母集団の平均値)

・ サンプル平均値の標準偏差 (s) → $\sigma \times \frac{1}{\sqrt{n}}$ (母集団の標準偏差の $\frac{1}{\sqrt{n}}$ 倍)

< サンプル平均値に関する法則、定理 >

○ サンプル平均値の標準偏差が $\frac{1}{\sqrt{n}}$ 倍となる。 $V(\bar{x}) = \frac{\sigma^2}{n}$

→ 大数の法則 (law of great numbers): サンプル数が増加すると母平均に近づく。

○ 母集団が正規分布でなくても サンプル平均値の分布は正規分布となる。

→ 中心極限定理 (central limit theorem)

4. 3 統計量の分布と基本的特性

(1) サンプル平均値 (\bar{x}) の分布

母集団(平均値: μ 、標準偏差: σ)

→ サンプルの大きさ n から得られる平均値 (\bar{x}) の分布

○ 母分散 (σ^2) が既知 → 平均値の分布は **正規分布**
検定、推定は **正規分布**

サンプル数が十分に多ければ ($n > 50$ 程度)、標本分散は母分散と同じとみなすことができる。

○ 母分散 (σ^2) が未知 → サンプルから得た不偏分散 (V) を用いる。
平均値の分布は **t 分布**
検定、推定は **t 分布**

4. 3 統計量の分布と基本的特性

(1) サンプル平均値 (\bar{x}) の分布

t 分布 (t distribution)

(定義)

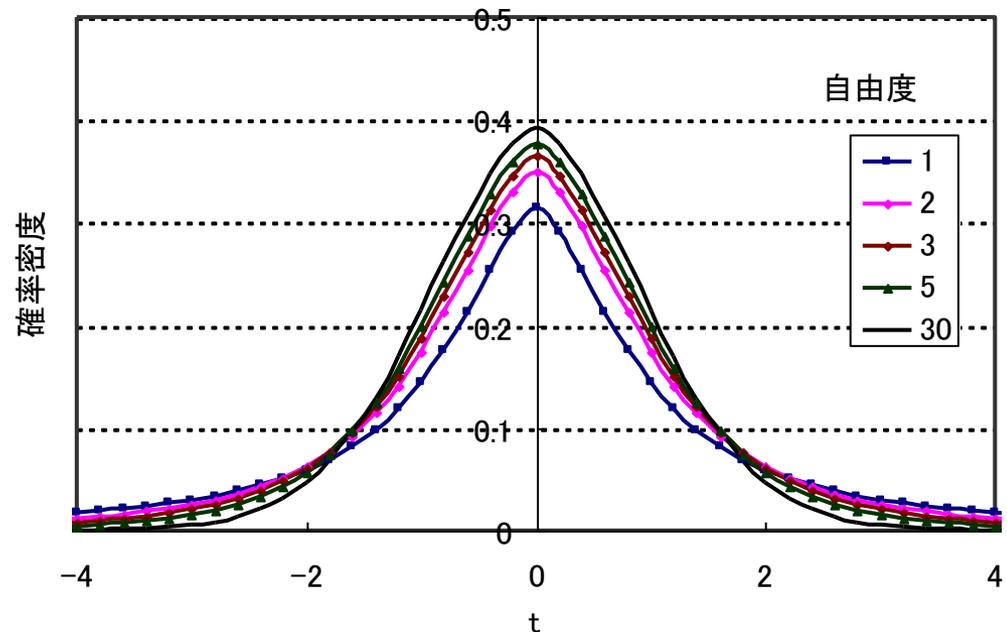
$$t = \frac{(\bar{x} - \mu)}{(s / \sqrt{n})}$$

ここで $s = \sqrt{V} = \frac{\sqrt{S}}{\phi}$ (サンプルの標準偏差)

t は自由度: $\Phi = n - 1$ の t 分布 をする。 ($n \rightarrow \infty$ で正規分布)

(応用)

母分散が未知のとき、
母平均に対する検定、推定に
用いられる。



4.3 統計量の分布と基本的特性

(2) サンプル分散 (s^2) の分布

母集団(平均値: μ 、分散: σ^2)

→ 試料の大きさ n から得られる分散 (s^2) の平均値とその標準偏差の分布

・分散 (s^2) の平均値 → $\frac{(n-1)}{n} \times \sigma^2$ (注) 正規分布とはならない

・分散 (s^2) の標準偏差 → $\frac{\sqrt{2}(n-1)}{n^2} \times \sigma^2$

(留意点)

○分散の推定値(母集団の分散(σ^2)に最も近い値)は、 $\frac{n}{(n-1)} \times s^2$

不偏分散(unbiased variance)

サンプルから求める分散は通常、不偏分散を用いる。

不偏分散: $V = \frac{n}{(n-1)} \times s^2 = \frac{S}{\phi}$ (平方和 / 自由度)

自由度: $\phi = n - 1$

(n ではないことに注意)

4. 3 統計量の分布と基本的特性

(3) χ^2 の分布

カイ二乗分布 (chi-square distribution)

(定義)

$$\chi^2 = \frac{S}{\sigma^2} = \sum_i \frac{(x_i - \bar{x})^2}{\sigma^2} = \sum_i \left\{ \frac{(x_i - \bar{x})}{\sigma} \right\}^2$$

標準正規分布 ($N(0,1)$: 平均 0、標準偏差 1) の分布から n 個 (自由度) のサンプルを抜き取り、二乗の和をとった場合の分布。

それぞれの正規分布が独立であり、平均、標準偏差が異なる場合、上式で表す。

・母平均が既知の場合

分散が等しければ、 χ^2 は自由度: $\Phi = n$ の
カイ二乗分布をする。

・母平均が未知の場合

分散が等しければ、 χ^2 は自由度: $\Phi = n - 1$ の
カイ二乗分布をする。

4. 3 統計量の分布と基本的特性

(4) 二つのサンプルで、不偏分散の比率 (V_1/V_2) の分布

F分布 (F distribution)

(定義)

自由度がそれぞれ ϕ_1 、 ϕ_2 であるカイ二乗変数 $\chi^2_{\phi_1}$ 、 $\chi^2_{\phi_2}$ とが互いに独立であるとき、その比

$$F(\phi_1, \phi_2) = \frac{\chi^2_{\phi_1} / \phi_1}{\chi^2_{\phi_2} / \phi_2} \quad \text{がF分布とされる。}$$

ここで、 $\chi^2_{\phi_1}$ 、 $\chi^2_{\phi_2}$ が、それぞれ正規分布の母集団からサンプルしたもので、その不偏分散を s^2_1 、 s^2_2 とすれば、

$$F(\phi_1, \phi_2) = \frac{s^2_1 / \sigma^2_1}{s^2_2 / \sigma^2_2}$$

もし、二つの正規分布の母集団の分散が等しければ、($\sigma^2_1 = \sigma^2_2$) であれば、母数に関係なく、不偏分散のみの関係式となる。

$$F(\phi_1, \phi_2) = \frac{s^2_1}{s^2_2} = \frac{V_1}{V_2} \quad \text{(分散分析で用いる)}$$

4. 3 統計量の分布と基本的特性

(4) 二つのサンプルで、不偏分散の比率 (V_1/V_2) の分布

F分布 (F distribution)

二つの試料の不偏分散 $V_1 = \frac{S_1}{(n_1 - 1)}$ $V_2 = \frac{S_2}{(n_2 - 1)}$ として $F = \frac{V_1}{V_2}$ とするとき

F は 自由度: $\Phi_1 = n_1 - 1$ 、 $\Phi_2 = n_2 - 1$ の F 分布となる。

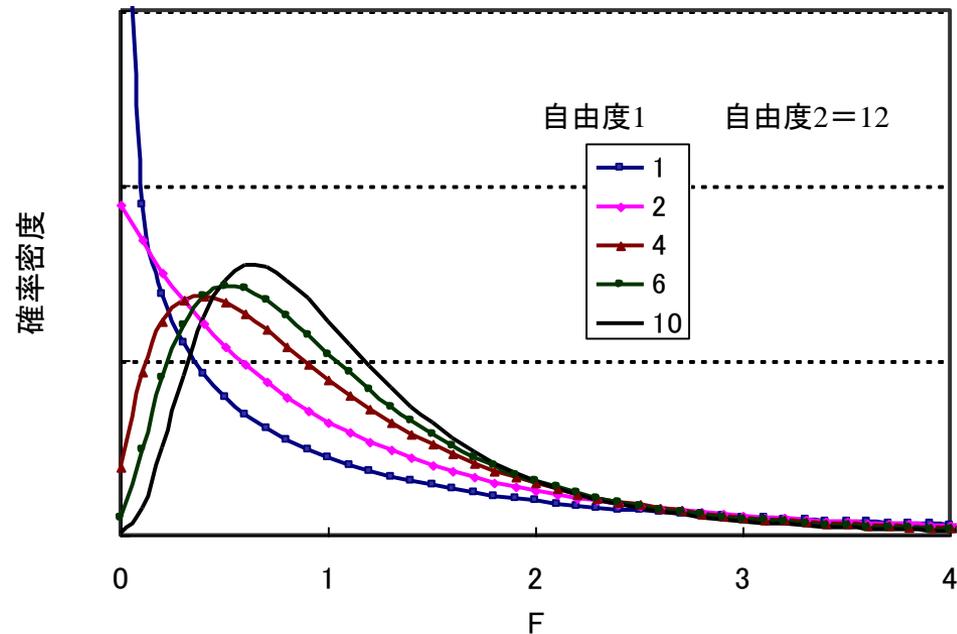
$$F(\phi_1, \phi_2, \alpha) = \frac{1}{F(\phi_2, \phi_1, 1 - \alpha)}$$

(α : 信頼度)

(応用)

○等分散性の検定: 正規分布に従う2つの群の「標準偏差が等しい」という帰無仮説の検定。

○分散分析: 正規分布に従う複数の群(標準偏差は等しいと仮定する)で、「平均が等しい」(つまり同じ母集団に由来する)という帰無仮説の検定。



4. 3 統計量の分布と基本的特性

(4) 二つのサンプルで、不偏分散の比率 (V_1/V_2) の分布

F分布 (F distribution)

(補足)

1) 二つの確率変数 U と V がそれぞれ独立に、カイ二乗分布に従うとき、二つの確率変数の比として定義される確率変数。

$F = \frac{U / \phi_1}{V / \phi_2}$ は、自由度 ϕ_1 、 ϕ_2 の F 分布に従う。

4. 3 統計量の分布と基本的特性

[統計量の分布のまとめ]

分布	応 用
正規分布	<ul style="list-style-type: none"> ・正規分布に従う2つの集団の平均値の検定(母分散が既知) ・平均値の区間推定(母分散既知) ・基本となる分布 統計解析は、基本的に本分布を前提とする ・自由度の概念はない
t 分布	<ul style="list-style-type: none"> ・正規分布に従う二つの集団の平均値の検定(母分散が未知) ・平均値の区間推定(母分散未知) ・n(サンプル数)\rightarrow無限大 で正規分布
χ^2 分布	<ul style="list-style-type: none"> ・二つのグループの独立性の検定、カテゴリ別の比率が基準の比率と同じかの適合度の検定。
F 分布	<ul style="list-style-type: none"> ・正規分布に従う二つの集団の標準偏差の検定。 ・3群以上のデータがある正規分布に従うかを1回で検定(分散未知) \rightarrow 分散の比較から、平均値の違いの検定(分散分析)

5. 基本的な統計解析手法

当研究所 販売商品
電子書籍セット(CD 3枚組)を
ご覧下さい

[まとめ]

知りたい情報(母数:平均値・分散)は、サンプル(標本)を通して得る。

サンプルから得た統計量は、真の値(母数)とは言えない。
必ず、誤差がついている(ばらついている)。



統計量が、真の値(母数)に対して、ばらつく様子(分布関数)を表すのが統計量の分布関数。

- ・正規分布
- ・ t 分布
- ・ F 分布

→ 複数の統計量の違いの有無、大小関係を調べるのが検定。

→ 真の値(母数)を推測するのが推定。

終わり

