

ソフトウェアメトリクス 第1回:ソフトウェアメトリクスの基礎、 統計的手法の基礎

野中 誠
東洋大学

2010年4月12日



本講座の目的

- 「数字による管理」の実現
 - 品質の定量的評価と予測
 - 各種手法の有効性評価
 - プロジェクトの状況判断 など
- 習得項目
 - ソフトウェアメトリクスに関する基礎知識
 - データ分析技法
 - ソフトウェアエンジニアリング研究の知見 など



本講座で扱う主なトピックス

■ 測定技法(メトリクス)

- プロダクト(規模、複雑さ、機能、品質)
- プロセス(プロセスの質)

■ 評価・予測技法

- ソフトウェア規模・開発工数の見積り
- fault-proneモジュールの予測
- 成長曲線を用いた信頼性予測 など

■ データ分析技法

- 統計手法、統計解析ツール(EXCEL, R)

アプローチ

■ 体系的な学習

- 全体における各トピックの位置づけを明確化

■ アカデミックな視点を重視

- 研究論文を主な情報ソースとする
- 実践的ノウハウはあまり重視しない

■ 研究・実務の両視点に基づくディスカッション

- 研究で示された技法・知見を実践に適用するには？
- 実務に適用する際の課題は？

ソフトウェアメトリクスの基礎

なぜメトリクスが必要なのか？

- 「必要な情報」を得た上での意思決定 (informed decision)
 - ソフトウェア／プロジェクトといった抽象的・複雑な対象について、
 - そのマネジメントにおける様々な意思決定に役立つ情報を、
 - 一貫した方法で獲得し、論理的な判断を下し、リスクを把握し、
 - そして、前へ前へと進んでいきたい
- 「必要な情報」の獲得・伝達手段として
 - 主観的情報も、意思決定の際に用いている (例)説明の論理性を判断
 - 主観的情報は、情報収集の一貫性、効率性、説明性に問題がある
 - メトリクスを定めることで、必要な情報を、効果的に、一貫して収集できる
 - 組織やメンバーに説明可能な根拠情報として利用できる



定量的管理に取り組まない合理的理由は、極めて限定的である

L. M. Laird and M. C. Brennan, *Software Measurement and Estimation: A Practical Approach*, John-Wiley and Sons, 2006.
(野中・鷲崎訳: 演習で学ぶソフトウェアメトリクスの基礎、日経BP社(2009)) 訳者前書きより

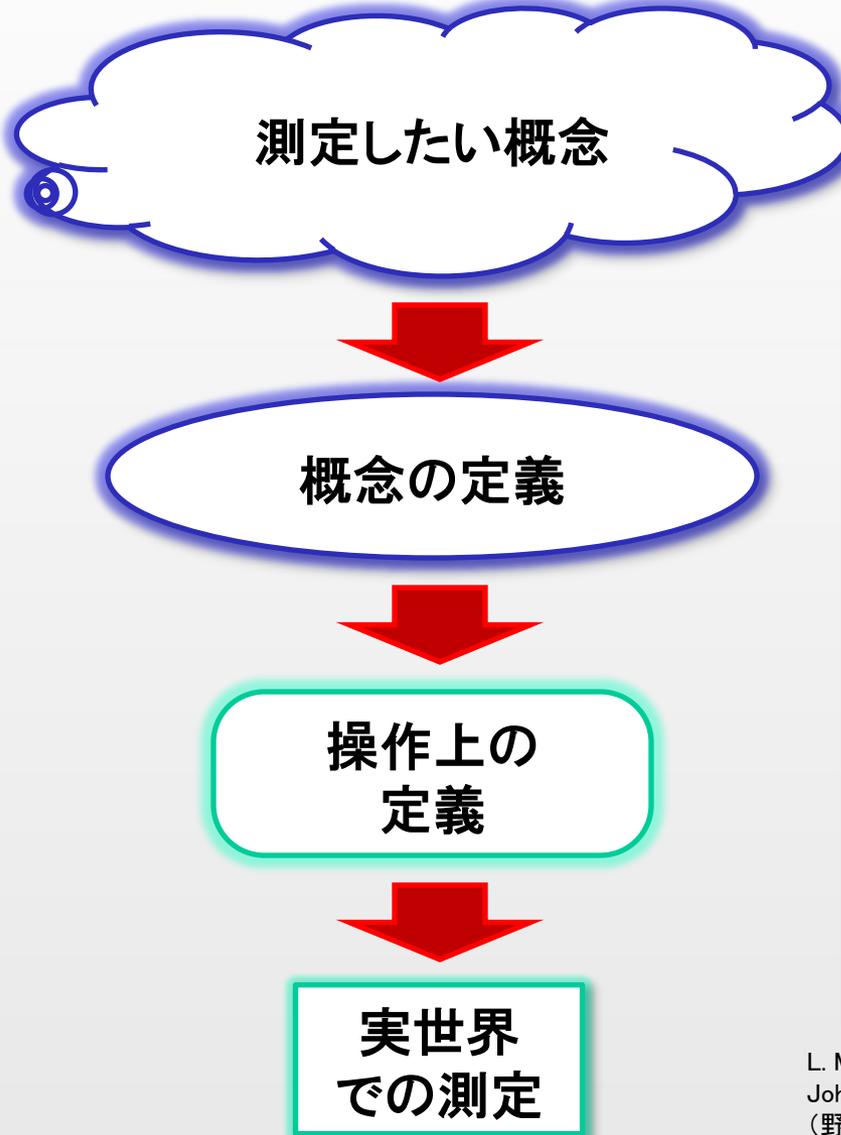
定量的管理／メトリクスの意義

- 予測・管理の基本
 - What you cannot measure is neither predict nor control.
測れないものは、予測したり管理したりすることができない (ケルビン卿)
- ソフトウェアエンジニアリングの基本
 - 体系的・規律的・定量的なアプローチを、ソフトウェア開発・運用・保守に適用するのが、ソフトウェアエンジニアリング (IEEE標準用語集)
- 評価・判断の裏付け
 - 適正なレビュー／テストが実施されたことを、定量データで示す
 - 例) レビュー工程完了の可否を、定量データを用いて判断する
 - 例) システムテスト完了の可否(出荷判定)を、定量データを用いて判断する
- 見える化
 - ソフトウェア／プロジェクトは、見て触って確かめることができない
 - 見るべき特性を「可視化」してはじめて、対象を理解・評価できる



日本的品質管理の特徴

- 全員参加
 - QC部門だけでなく、間接部門や管理部門も、品質向上に関わる
 - 経営トップから担当者まで、全員が品質向上にコミットする
- 品質第一
 - 短期的な利益を優先したのでは、品質は低下する
 - 不良や手戻りを減少させることで、生産性を向上する
- 消費者指向(マーケット・イン)
 - 顧客が欲しいものをしっかりと把握し、それを実現して提供する
 - 作り手側の思いこみによる製品の押しつけではダメ
- 事実・データ指向
 - 事実・データを分析し、その結果を意思決定へのインプットとする
- 後工程はお客様
 - 自分のやるべきことをしっかりとやって(自工程完結)、次の工程へ回す



測定したい概念

概念の定義

操作上の
定義

実世界
での測定

メトリクスの限界を知り、 意思決定に役立てる

- 「実世界で測定」できることは、「測定したい概念」のごく一部に過ぎない
- 過度の期待を抱いてはいけない
- メトリクスの限界を理解した上で、“informed decision” に役立てる

L. M. Laird and M. C. Brennan, *Software Measurement and Estimation: A Practical Approach*, John-Wiley and Sons, 2006.

(野中・鷺崎訳: 演習で学ぶソフトウェアメトリクスの基礎、日経BP社(2009))



よい測定とは

- 妥当性(validity)の高い測定
 - 内容的妥当性(content validity)がある
 - 測定対象の属性を適切に表現できているか
 - 基準関連妥当性(criterion-related validity)がある
 - 測定結果によって外部事象の予測に役立つか
 - 構成概念妥当性(construct validity)がある
 - 抽象的概念や理論的概念を適切に測定しているか

- 信頼性(reliability)の高い測定
 - 測定結果に一貫性があり、安定している
 - ばらつき(variability)が小さい
 - 偏り(deviation)が小さいとは限らない



例題：何に問題があるか？

ファンクションポイント(FP)の測定実験：組織別

組織	測定者数	FP平均	SD	SD/平均	範囲
1	7	51.9	8.6	16.5	40 – 64
2	5	85.5	53.6	62.7	34 – 159
3	5	47.8	2.6	5.5	44 – 50
4	2	51.0	0.0	0.0	51 – 51
5	3	57.3	13.6	23.7	49 – 73
6	1	73.0			
7	1	26.0			

Low, G. C. and D. R. Jeffery (1990). "Function points in the estimation and evaluation of the software process." *Software Engineering, IEEE Transactions on* **16**(1): 64-71.



例題：何に問題があるか？

ファンクションポイント(FP)の測定実験：経験別(組織2は除外)

グループ	測定者数	FP平均	SD	SD/平均	範囲
熟練者	18	51.5	11.1	21.6	26 - 73
非熟練A	11	83.7	34.5	41.2	52 - 144
非熟練B	9	72.9	30.6	42.0	43 - 139

ソフトウェアメトリクス研究における基本的な問い

「正確に誤るよりも漠然と正しくありたい」

経済学者 ケインズ (J. M. Keynes)

我々の研究は、「正確に誤って」いないだろうか？

⇒ 理論モデルを高度化しないと論文が通らない！？

我々の研究は、「漠然と正しい」を満たしているか？

⇒ 現実を反映した、ロバストなモデルを提供しているか？



基本用語

■ メトリクス(metrics)

- a **quantitative measure** of the degree to which a system, component, or process possesses a given attribute (IEEE Std 610.12-1990)

- 定義された測定**方法**及び測定**尺度** (ISO/IEC 14598-1: 1998)

■ 測定量([名] measure)

- 測定の結果として値が割り当てられる**変数** (ISO/IEC 15939:2002)

■ 尺度(scale)

- 連続的もしくは離散的な**値**の順序集合、又は**分類**の集合で、それに属性を対応付けるもの (ISO/IEC 15939:2002)

『欠陥』に関連する用語 — IEEE標準の定義

■ エラー、誤り (error)

- 計算、観測、または測定された値または状態と、正しい、仕様化された、または理論的に正しい値または状態との差異

■ 障害 (fault)、バグ (bug)、欠陥 (defect)

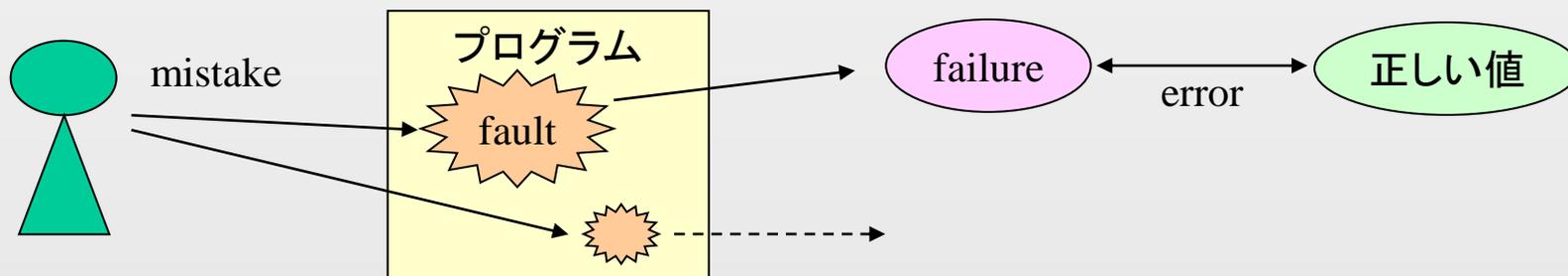
- 不正確なステップ、プロセス、またはデータ定義

■ 故障 (failure)

- 不正確な結果

■ 間違い (mistake)

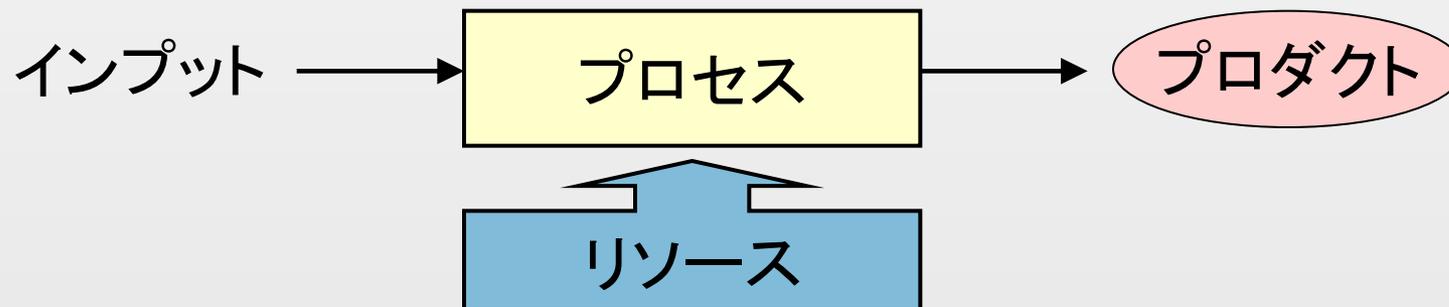
- 不正確な結果を作り出す人間の行為



IEEE Std 610.12-1990: *IEEE Standard Glossary of Software Engineering Terminology*, IEEE Computer Society (1990).

何を測定するのか

- プロダクト (product)
 - アクティビティから出力される成果物(文書含む)
- プロセス (process)
 - 入力を出力に変換するアクティビティの集合
- リソース (resource)
 - アクティビティの実施に必要なエンティティ



Fenton, N. E. and S. L. Pfleeger (1997). Software Metrics: A Rigorous and Practical Approach, PWS Publishing.

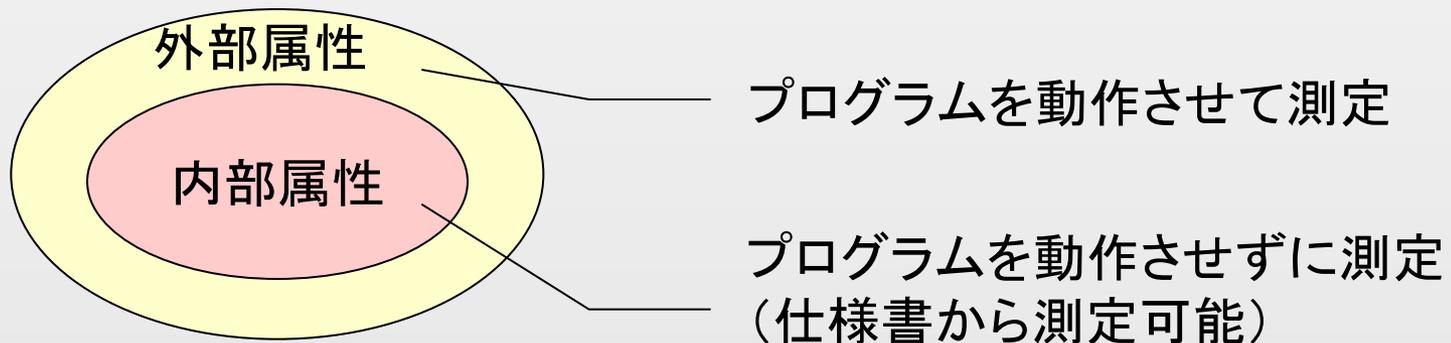
製品の内部属性と外部属性

■ 内部属性 (internal attributes)

- プロダクト単体で測定できる物理的／概念的特徴
- プログラムを動作させずに測定できる

■ 外部属性 (external attributes)

- プロダクトの振る舞いを通じて測定される特徴
- プログラムを動作させて測定する





プロダクト内部属性の例

■ 規模

- 長さ:ドキュメントのページ数、LOC
- 機能性:実現された機能、機能規模
- 複雑性:問題の複雑さ、アルゴリズムの複雑さ
- 再利用:再利用率

■ 構造

- 制御フロー:サイクロマティック数
- 情報フロー:fan-in/out
- モジュラリティ:結合度、凝集度
- オブジェクト指向に関して:C&Kメトリクスの属性群

Fenton, N. E. and S. L. Pfleeger (1997). Software Metrics: A Rigorous and Practical Approach, PWS Publishing.



プロダクト外部属性の例

- 信頼性 (reliability)
 - 発見欠陥数、潜在欠陥数の予測値
 - MTBF (Mean Time Between Failure)
- 使用性 (usability)
 - タスク遂行時間
- 効率性 (efficiency)
 - 応答時間、スループット、ターンアラウンド時間
- 保守性 (maintainability)
 - 変更作業の所用時間



プロセスの属性

■ 期間

- アクティビティの完了までに要した時間

■ 工数

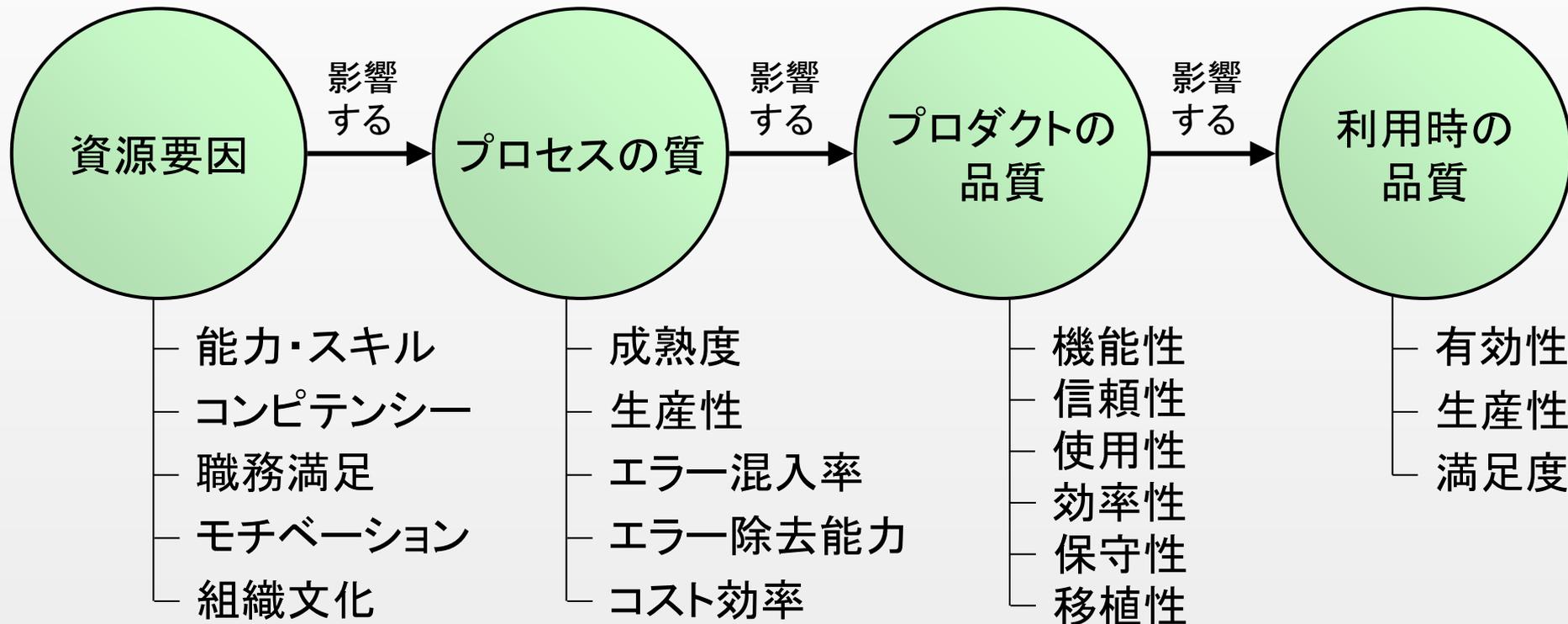
- アクティビティの遂行に要した工数

■ 事象の発生回数

- アクティビティの遂行中に生じた出来事の数
- 例) レビュー工程で発見した欠陥数

Fenton, N. E. and S. L. Pfleeger (1997). Software Metrics: A Rigorous and Practical Approach, PWS Publishing.

なぜプロセスに着目するのか？



プロセスの質は、プロダクトの品質に影響を及ぼす
資源要因は、プロセスの質に影響を及ぼす

分類	質問内容	成功		失敗		平均の差
		平均	S.D.	平均	S.D.	
職務満足	意欲的に仕事ができる環境づくり	3.6	1.07	2.7	0.52	0.93*
人間関係	ざっくばらんな雰囲気重視	3.5	0.97	2.7	0.52	0.83*
柔軟性・ 創意工夫	小さなことでも何か工夫を凝らす	3.6	0.52	2.8	0.41	0.77*
	やるべきことをすぐに実行	3.7	0.67	2.2	1.17	1.53*
	新しいマネジメントスタイルを柔軟に適用	3.3	0.82	2.3	0.52	0.97*
	慣習にとらわれず、最適なことを考える	3.7	0.95	2.7	0.82	1.03*
	臨機応変に行動	3.5	0.53	2.7	0.52	0.83*
	可能性や原因を広く考える努力	3.7	0.67	2.8	0.75	0.87*
情報共有	出来事や問題を全員が共有	4.2	0.92	2.3	1.03	1.87*
	指示の字面に現れない裏の意図を解釈	3.3	0.95	2.2	0.75	1.13*
こだわり	やっつけ仕事をせず、簡単に妥協しない	3.5	0.53	2.8	0.41	0.67*
決定重視	開発プロセス基準など公的ルール遵守を重視	4.3	0.67	2.8	0.98	1.47*
その他	謙虚、人間的に成熟	3.7	0.48	3.0	0.63	0.70*
	問題の構図を捉えてスマートな解決を図る	3.3	0.48	2.0	0.89	1.30*

* p<0.05

野中, 西 (2007): 「組織的要因がソフトウェア品質に与える影響」第26回ソフトウェア品質シンポジウム発表報文集

Copyright (C) 2007 National Institute of Informatics, All rights reserved.

Software Measurement and Estimation: A Practical Approach

演習で学ぶ ソフトウェア メトリクスの基礎

【ソフトウェアの測定と見積もりの正しい作法】

Linda M. Laird and M. Carol Brennan
リンダ・M・ライルド・M・キャロル・ブレナン著
Makoto Nomura and Hiromori Washizuki
野中 誠・鷲崎弘寛訳

169題の演習を通じて
ソフトウェアの
欠陥・工数・規模が
“見える”!

「数字による管理」を確立し、
進化させるための実践的ガイド

日経BP社 定価/本体2800円+税

目次

1. イントロダクション
2. 何を測定するか
3. 測定の基礎
4. ソフトウェア規模の測定
5. 複雑性の測定
6. 工数の見積もり
7. 欠陥に学ぶ
8. ソフトウェアの信頼性測定と予測
9. 応答時間と可用性
10. 進捗の測定
11. アウトソーシング
12. ソフトウェア技術者のための財務指標
13. ベンチマーキング
14. 経営陣にメトリクスをうまく伝えるには

日経BP社より好評発売中！

統計的手法の基礎

(典型的な)統計的手法の学習項目

典型的な統計学の教科書が扱う全範囲

1. 表・グラフ
2. 統計学の基本概念
3. 基本統計量(平均、標準偏差など)
4. 相関係数・回帰直線(記述統計における)
5. (時系列解析法)
6. 確率と確率変数
7. 二項確率分布
8. 正規確率分布
9. 標本分布(中心極限定理、標本平均、標準誤差)
10. 推定・検定(大標本、小標本)
11. 分散分析
12. 相関分析・回帰分析(推測統計における)
13. 多変量解析
14. 質的データの分析(分割表の検定)
15. ノンパラメトリック手法

データ解析の第1ステージで
重点を置くべき範囲

多くの大学が確率・統計の講義で
取り扱う範囲

データ解析の第2ステージで
重点を置くべき範囲

一般的な統計学の学習で得られる知識と、
データ解析に役立つ知識の間には、
ややギャップがある

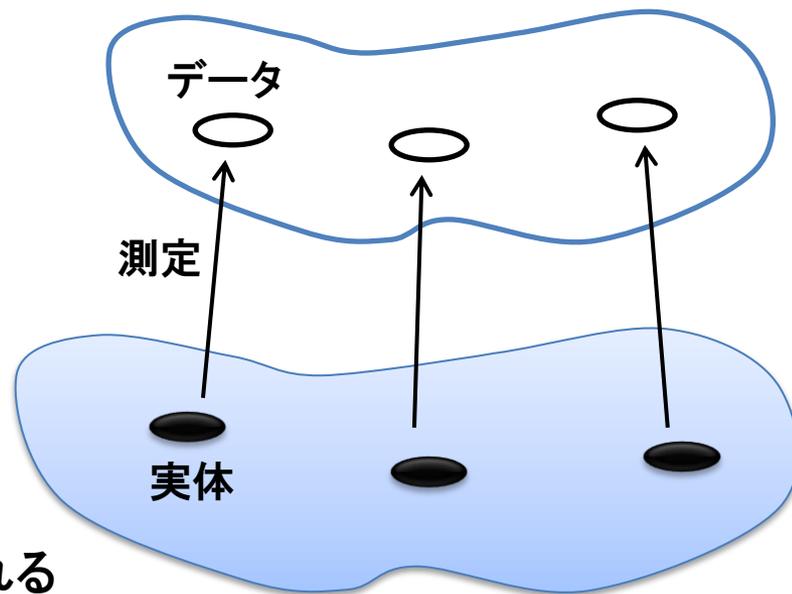
全範囲を学習すればギャップは埋まるが、
初学者には大変

データ解析という目的に軸足を置き、
統計学基礎を概観した方が実践的

データとは何か？

データ：実体を、数値などの記号によって表現したもの

- **実体 (entity)**
 - 測定が行われる個々の対象
- **測定 (measurement)**
 - 定められた測定手法に従って、変数に値を割り当てる行為
- **データ (data)**
 - 測定によって得られた値
 - 4つの測定尺度 (scale) のいずれかで表される
 - 名義尺度、順序尺度、間隔尺度、比率尺度
- **変数 (variable)**
 - 実体の特性(個体差や時間経過により値が変化する)
 - 質的変数 (qualitative/categorical) と量的変数 (quantitative/numerical) がある
 - 量的変数は、さらに、次に分類される
 - 離散変数 (discrete) 有限の数値、カウント可能な数値をとる変数
 - 連続変数 (continuous) 無限の数値



データを表す尺度

尺度の種類	概要	例	許容される演算
名義尺度	大小関係のない名義的な分類の集合	性別、企業コード	分類、数え上げ 最頻値
順序尺度	順序のある分類の集合	成績評価	順位付け、中央値 ※ 一定条件を満たせば、 間隔尺度に準じた演算も可
間隔尺度	等間隔性が保証された順序数の集合 絶対的なゼロを持たない	温度(摂氏・華氏)、 カレンダータイム	加減算 平均、分散、標準偏差など
比率尺度	間隔尺度の特性に加えて、 原点が定まった数値の集合	売上高、従業員数、 発生件数	四則演算 変動係数など

(補足)リッカート尺度

質問項目に対する回答者の態度を測定する簡便な方法

5段階や7段階の尺度がよく用いられる

(例) 嫌い - 少し嫌い - どちらでもない - 少し好き - 好き

リッカート尺度は、上記4つのうちどれに該当すると考えられるか？

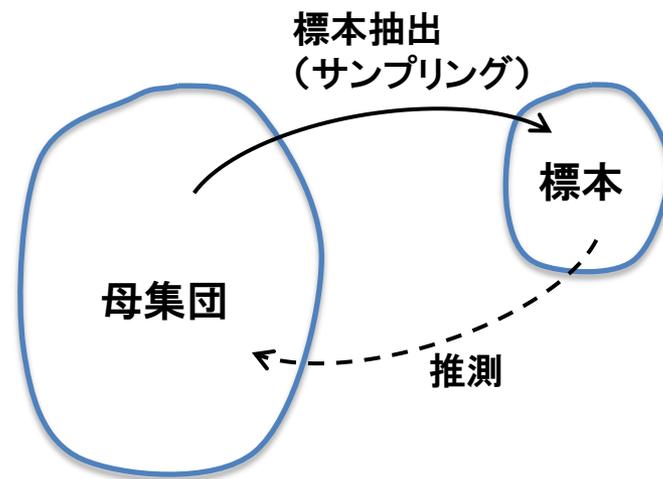
母集団と標本

■ 母集団 (population)

- 分析者が関心を持つ測定値のすべての集合
- 母集団の特性値を母数 (parameter) と呼ぶ

■ 標本 (sample)

- 母集団の部分集合
- 通常は、無作為抽出により標本を得る
- 標本の特性値を統計量 (statistic) と呼ぶ



■ 記述統計学 (descriptive statistics)

- 母集団の特性を要約・記述するのに用いる統計手法

■ 推測統計学 (inferential statistics)

- 母集団の特性を、標本が持つ情報を用いて推測する統計手法

グラフを読むポイント

■ 尺度

- 横軸と縦軸は、何について測定した項目が記載されているのか？

■ 分布の位置

- 横軸上におけるデータの中心はどのあたりか？
- 複数の分布を比較する場合、それらの中心は同じか、異なるか？

■ 分布の形状

- 「ピーク」は1つか、複数か？ ピークをとる値はいくつか？
- 分布は左右対称か？ 左右に偏っていないか？
 - 右(左)に偏る … ピークの値よりも右側(左側)に多くの測定値が含まれている状態
 - 例) サラリーマン世帯の年収や貯蓄額は、一般に、右に偏った分布である

■ 外れ値

- 他の測定値よりも極端に大きく外れた値はないか？
- それは、測定誤りによるものか？ それとも何か別の理由によるものか？

■ (2変数の場合)変数同士の関係

- 右上がり／下がりの直線関係か、曲線関係か？
- 関係の強さはどの程度か？
- 層別の必要性はないか？

基本統計量

■ 中心を表す統計量

- 算術平均: 母集団平均 μ 標本平均 $\bar{x} = \sum x_i / n$
- 中央値(メジアン): データ数が偶数個の場合は、中央2つのデータの平均で表す

■ 散らばりを表す統計量

- 分散
 - 母分散 $\sigma^2 = \sum (x_i - \mu)^2 / N$
 - 標本分散 $s^2 = \sum (x_i - \bar{x})^2 / n - 1$

- 標準偏差
 - 母標準偏差 $\sigma = \sqrt{\sigma^2}$
 - 標本標準偏差 $s = \sqrt{s^2}$

- 範囲 ... 最大 - 最小

- 四分位点

- 第1四分位点(Q1) ... 25パーセント点
- 第3四分位点(Q3) ... 75パーセント点

電卓で計算するときは、次の式が便利

$$s^2 = \frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n - 1}$$

なぜ標本分散を求めるときは $n-1$ で割るのか？

■ もっとも抽象度の高い説明

- 母分散を推定するにあたって、 n よりも $n-1$ で割った方が偏りのない推定量(不偏推定量)になるから

■ もう少し理論に踏み込んだ説明

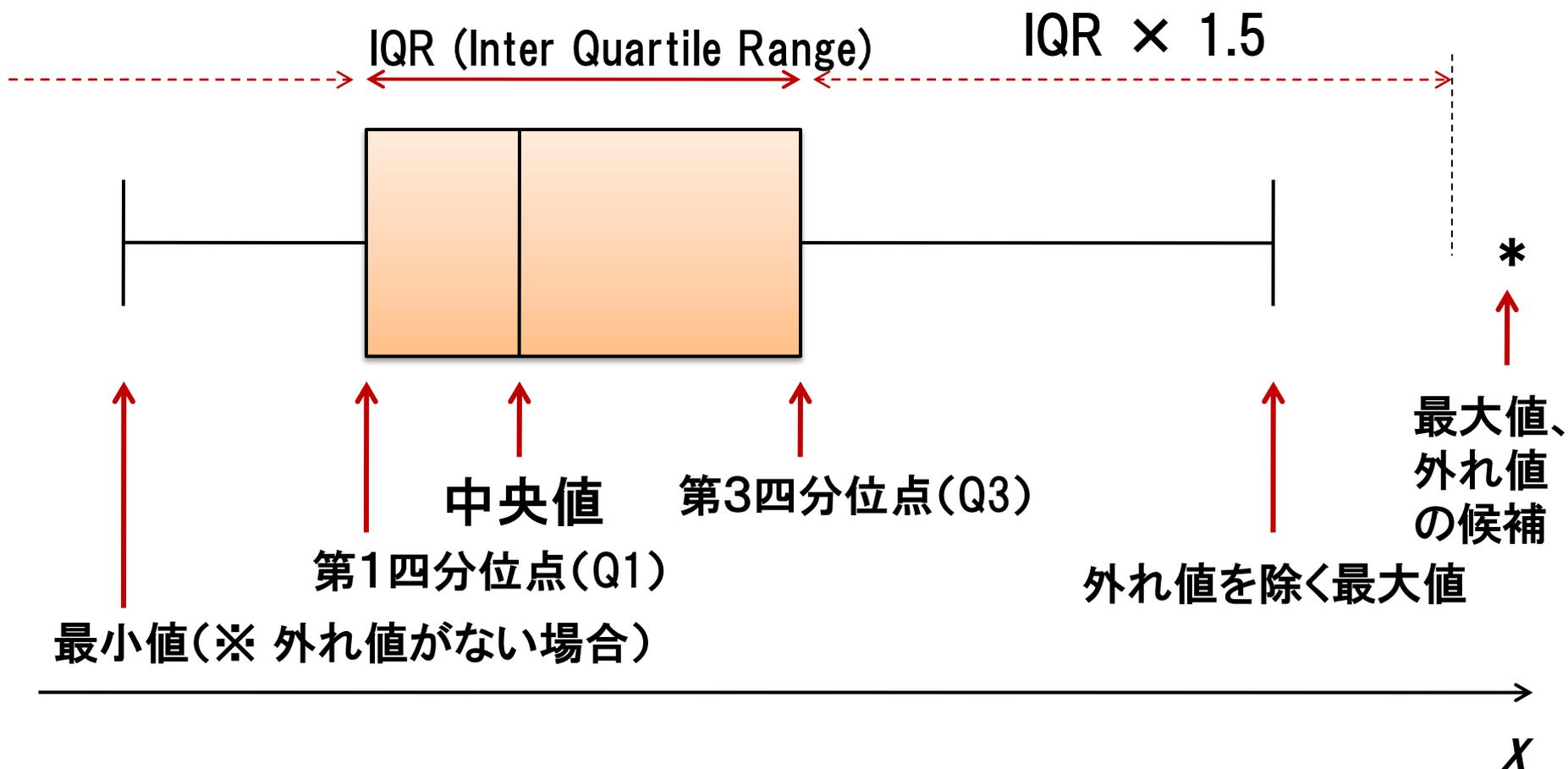
- 標本平均は、測定値の合計をデータ数で割った値である
- 標本分散は、偏差二乗和の平均値と見なすことができる
- 偏差の合計は必ずゼロになるため、標本分散を求める式における情報量は、 $n-1$ とするべきである

■ さらに踏み込んだ説明

- 確率変数の期待値の定義に基づいて証明(省略)

5数要約と箱ひげ図(boxplot)

- 5数要約 … 最小値、Q1、中央値、Q3、最大値
- 箱ひげ図



相関係数(correlation coefficient)

- **ゴール:2変数の関係を、 $-1 \leq r \leq 1$ で表現したい**
 - 右上がり \Rightarrow 正 右下がり \Rightarrow 負
 - 相関が強い \Rightarrow 絶対値が 1 に近い 相関が弱い \Rightarrow 絶対値がゼロに近い

■ 2変数の散布図から

- 共分散(covariance)

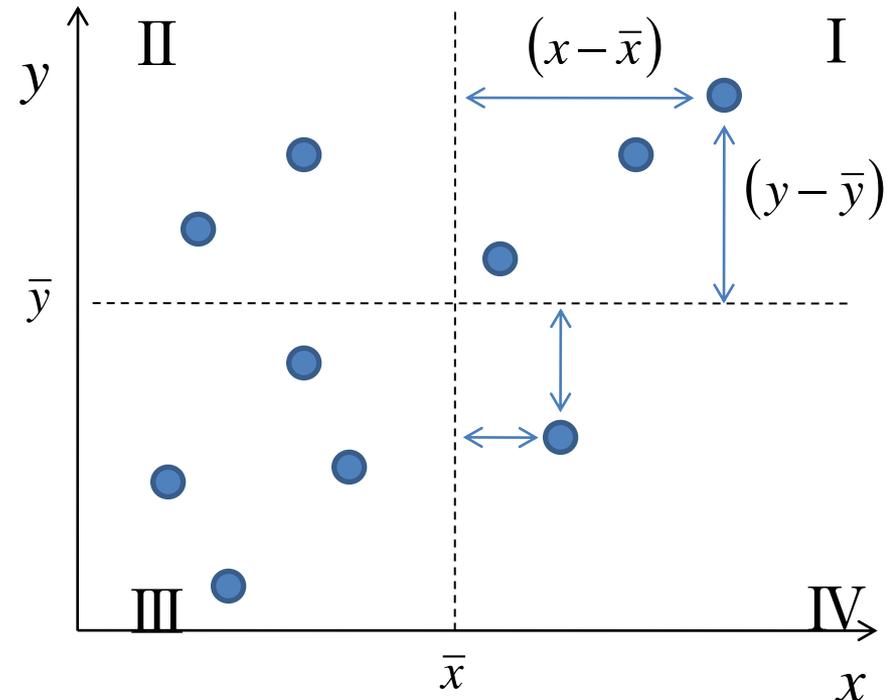
$$s_{xy} = \frac{1}{n-1} \sum (x - \bar{x})(y - \bar{y})$$

- 象限 I、IIIにデータが多ければ+
- 象限 II、IVに " -

- **相関係数**(correlation coefficient)

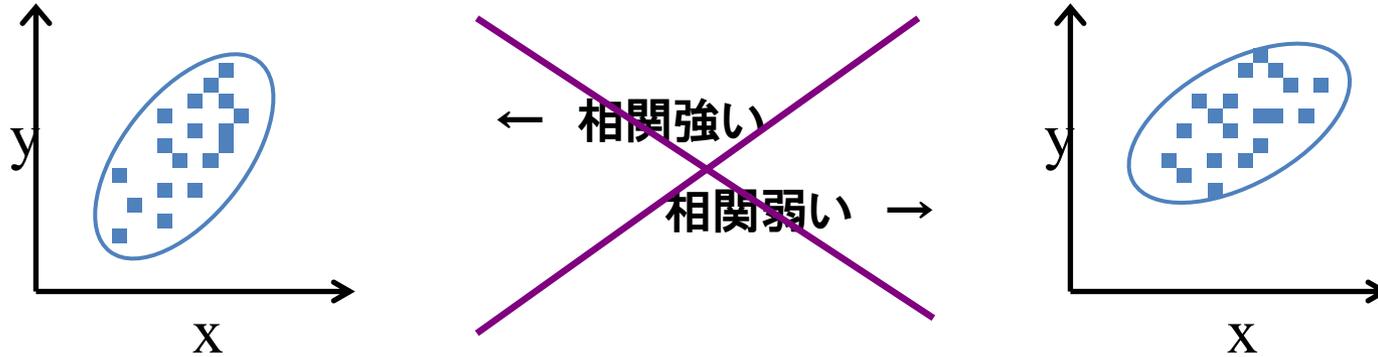
$$r = s_{xy} / s_x s_y$$

s_x : x の標準偏差
 s_y : y の標準偏差

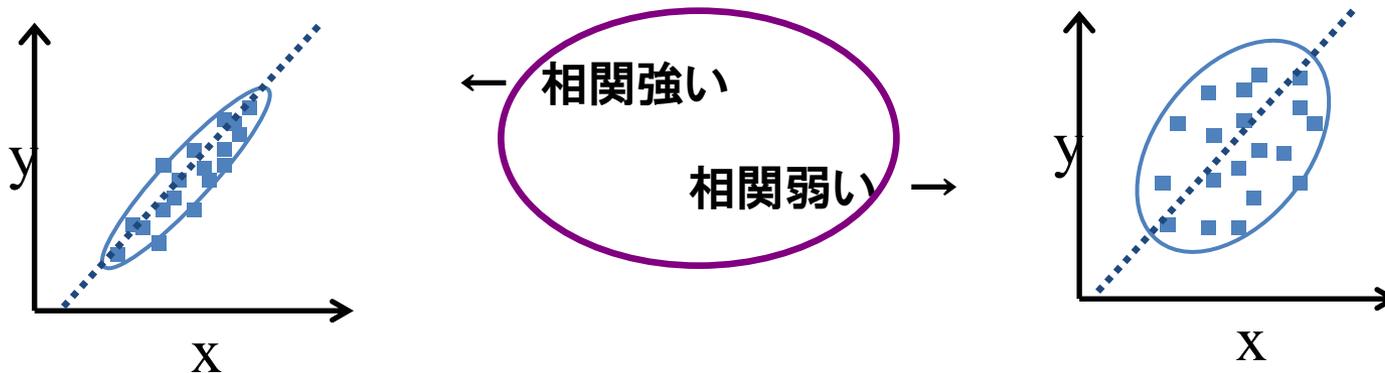


相関係数の意味

- 相関の強さ \neq 散布図の傾斜の強さ

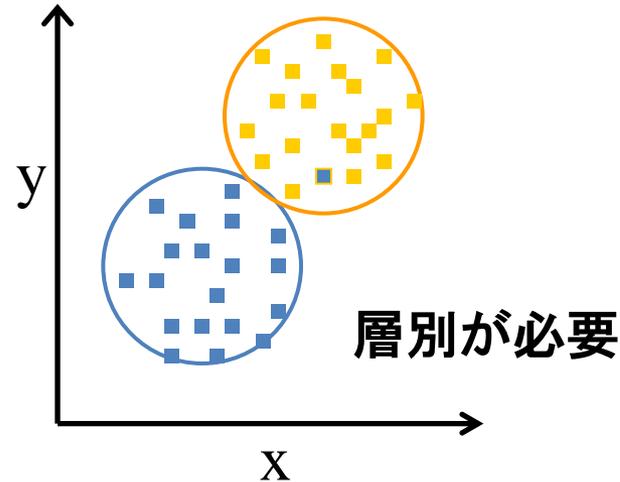
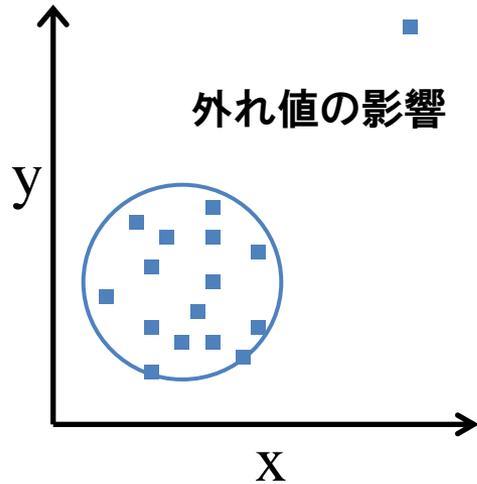


- 相関の強さ(弱さ)は、1本の直線から点が全体としてずれている程度

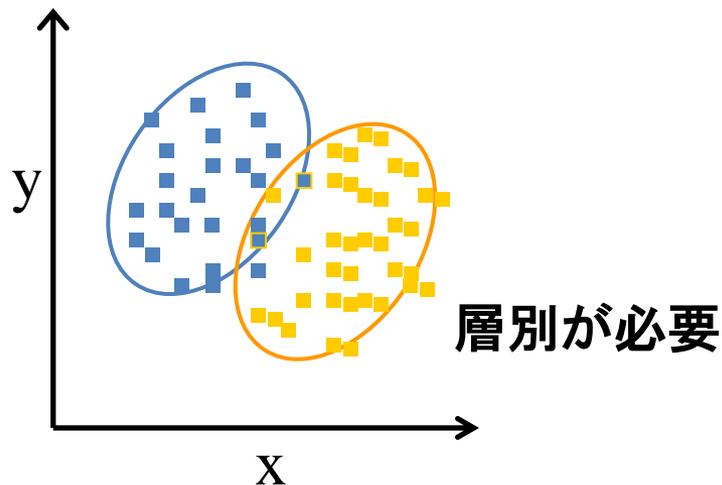


相関係数の留意点

- 実際は無関係なのに、相関係数の上では相関関係があるように見える場合



- 実際は関係があるのに、相関係数の上では無相関に見える場合

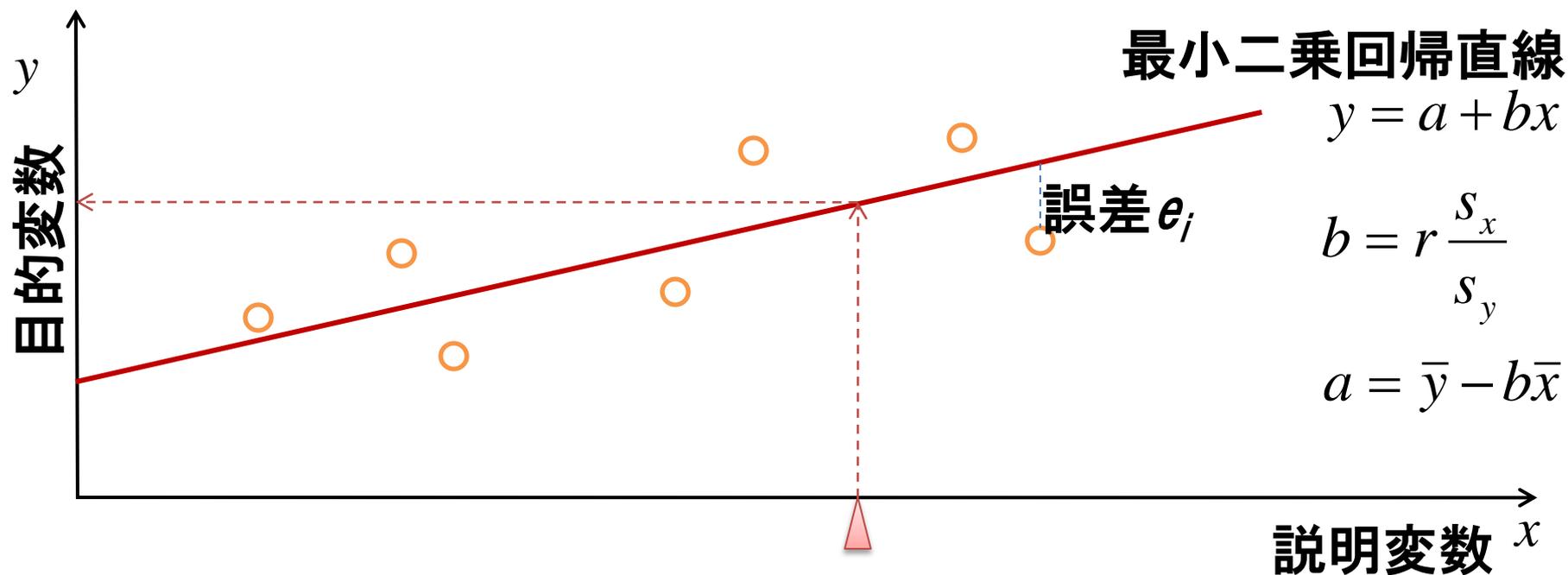


回帰直線(regression line)

説明変数(従属変数)の値を用いて、目的変数(独立変数)の値を予測する

誤差 e_i の二乗和を最小化する直線 $y = a + bx$ を求める

a と b について、それぞれ偏微分することで a と b が求まる



回帰分析の適用場面

■ 構造推定

- 興味対象の特性 y を支配する要因をモデル化

■ 制御

- 特性 y を狙った値にするために制御変数を定める

■ 予測

- 目的変数 y を y 以外の値から予測する

■ 変動要因解析

- 目的変数 y の変動に関する情報を得る

回帰分析における変数の扱い

- 説明変数に対する吟味が必要である
- データは実験的にとることが望ましい
- 変数間の相関は弱いことが望ましい
- 得られた回帰式の合理性を検証する必要がある
- 説明変数の値を自由に動かせることが必要である
(制御の場合)
- 説明変数は、目的変数よりも前に観測可能でなければならない

正規分布(normal distribution)

■ 連続確率変数の確率分布が持つ性質

- 連続確率分布と横軸に囲まれた面積は 1 である
- $P(x = a) = 0$ したがって $P(x \geq a) = P(x > a)$ である

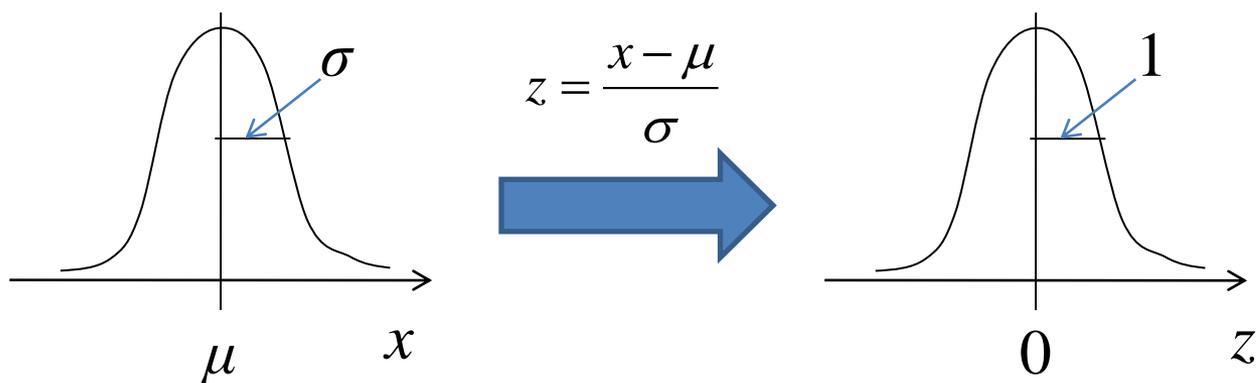
■ 正規確率分布(正規分布)

- 次の関数で与えられる確率分布
- $N(\mu, \sigma^2)$ で表すことが多い
- 左右対称である、単峰形である、平均値 = 中央値 である

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

■ 標準化(標準正規分布への変換)

- $N(\mu, \sigma^2)$ を $N(0, 1^2)$ に変換する

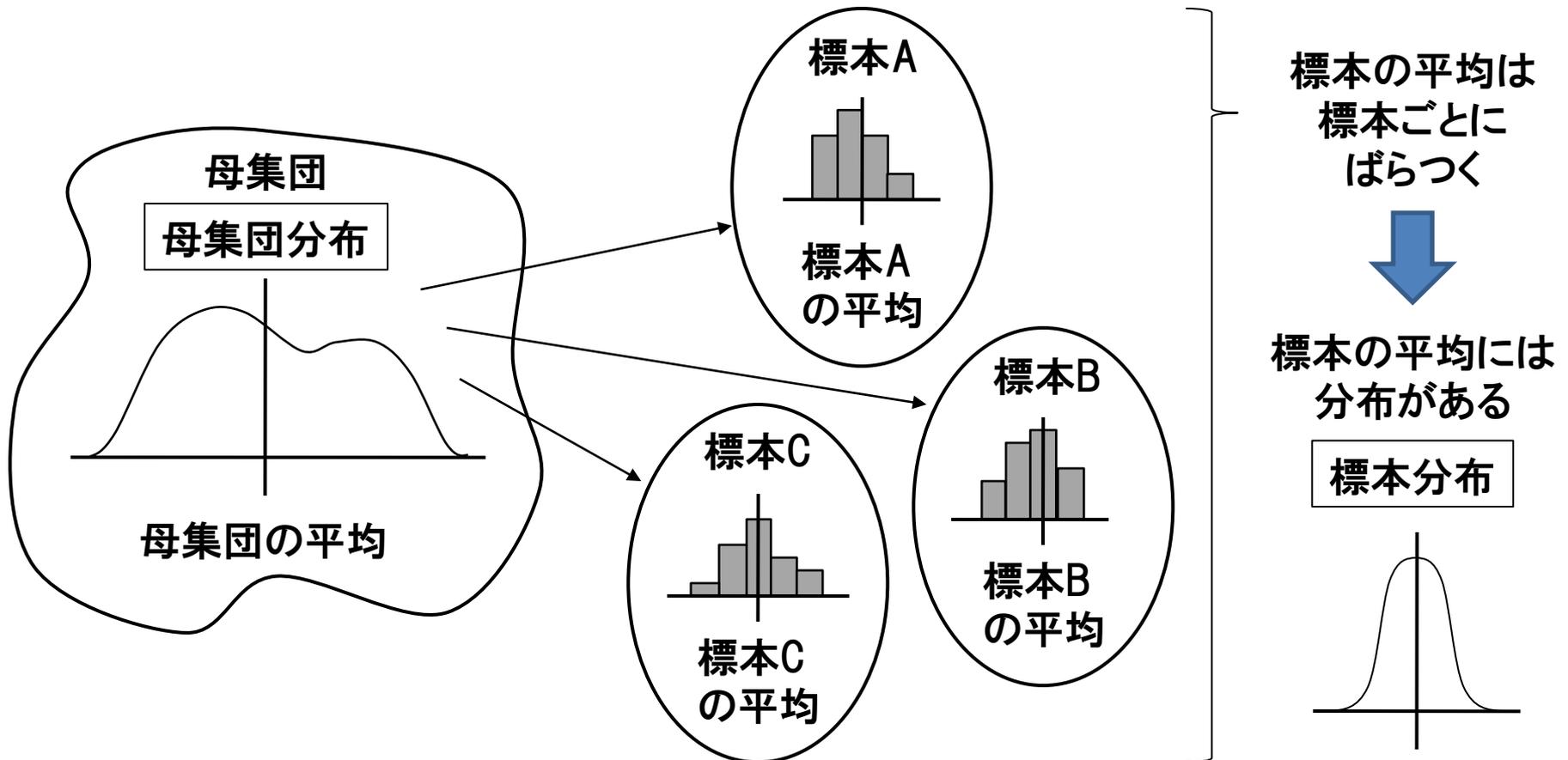


- 標準正規分布表から、特定の確率を求めることができる

標本分布(sampling distribution)

■ 統計量の標準分布

- 母集団から標本サイズ n の無作為抽出を繰り返し行ったとき、統計量の値がとり得る確率分布



統計的検定(Statistical Testing)

■ 基本的な考え方 ～ 体重測定を例に（片側検定の場合）

– 2つの仮説

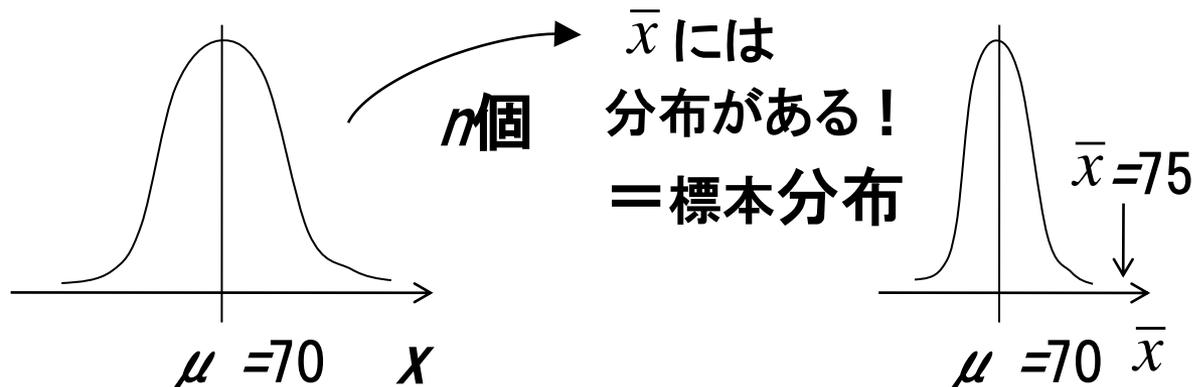
• **帰無仮説** H_0 : 野中の平均体重 μ は、70kgである

• **対立仮説** H_a : 野中の平均体重 μ は、70kgを超えている

– 標本サイズ n で、標本平均 \bar{x} を求める

《帰無仮説下の母集団分布》

《帰無仮説下の標本分布》



– n 個のデータから \bar{x} を求めた結果、帰無仮説下では滅多に起きない事象が観測された(この確率を p 値と呼ぶ)

– 帰無仮説が成り立っているのは不自然と考えるのが妥当

– したがって、帰無仮説を棄却する

多変量解析の分類

原因		結果		目的変数	
		結果	目的変数	間隔尺度	名義尺度
説明変数	間隔尺度	重回帰分析	判別分析		
	名義尺度	数量化理論 I 類	数量化理論 II 類		

- ロジスティック回帰分析
 - 重回帰分析の一種
 - 目的変数の値を、0～1の連続値として表現 → 確率を表すのに利用

多変量解析の手順

1. データの理解、妥当性確認
2. 変数とモデルの選択
3. 予備分析の実施
 - グラフ(ヒストグラム、散布図)
 - 基本統計量(平均、標準偏差、範囲など)
 - 変数変換が必要であれば実施
4. 多変量モデルの構築



データの理解と妥当性評価

- 何に関するデータか？
- データが収集された時期はいつか？
- なぜデータが収集されたのか？
- 誰が収集したのか？
- 各変数の定義は何か？
- データ入力者はそれを正しく理解していたか？
- 各変数の測定単位は何か？
- 各変数に入力された値の定義は何か？
- 欠損値はないか？ ゼロ記入があればその意味は？

重回帰分析の留意点：サンプルサイズ

- しばしば、「一説明変数あたり30個のデータが必要」と言われる！
- 実際には、
 - 有意水準 α （通常は0.05）
 - 効果量（効果変動／誤差変動、社会科学分野では一般に0.15）
 - 検出力（第二種の過誤にならない確率、一般には0.80）を指定すると、説明変数の数に応じたサンプルサイズの下限が決まる
- 上記条件の場合、

■ 説明変数1個	→	サンプルサイズ54個
■	2	→ 67個
■	3	→ 76個
■	4	→ 84個
■	5	→ 91個

参考 <http://www.danielsoper.com/statcalc/calc01.aspx>