



建設ICTマスター養成講座
基礎養成編 選択分野別ソフトウェア実習

表現技術検定（情報処理）

2020年07月29日(水)

一般財団法人最先端表現技術利用推進協会
State of the Art Technologies Expression Association
Tel.03-6711-1955 FAX.03-6894-3888
[mail: info@soatassoc.org](mailto:info@soatassoc.org) <http://soatassoc.org/>

Schedule

		項目	表現技術として要求される事項等
情報概論1	9:30～10:50	I 現代社会と情報リテラシー	<ul style="list-style-type: none"> ・資源としての情報 ・情報リテラシーとは ・情報とデータ ・情報システムとコンピュータ
		II 情報分析	<ul style="list-style-type: none"> ・情報分析の意義 ・抽象的な概念構造 ・チャート表現
休憩	10:50～11:00		
情報概論2	11:00～12:00	III データマイニング	<ul style="list-style-type: none"> ・データマイニングとは ・データマイニングの定義 ・知識発見のプロセス ・データマイニングの実際
		IV 情報社会と危機管理	<ul style="list-style-type: none"> ・情報セキュリティ ・技術的な情報セキュリティ対策
		V 情報社会と権利・法律	<ul style="list-style-type: none"> ・知的所有権(知的財産権) ・著作権の新しい流れ
昼食	12:00～13:00		

昼食	12:00～13:00		
情報概論3	13:00～14:20	VI 統計と情報	<ul style="list-style-type: none"> ・統計学の歴史 ・統計分析とデータ ・データのばらつきとヒストグラム ・基本統計量 ・相関関係
休憩	14:20～14:30		
情報概論3	14:30～15:20	VII 統計分析演習(Excel実習)	<ul style="list-style-type: none"> ・データの集計とグラフ作成 ・分散と標準偏差 ・標準得点と偏差値 ・回帰分析
質疑応答	15:20～15:30	質疑応答	
休憩	15:30～15:40		
検定	15:40～16:30	VIII 検定	<ul style="list-style-type: none"> ・講習内容についての検定を実施

はじめに

コロナ禍で分かった、日本社会の弱点

- ・政治家の脆弱な危機管理能力
- ・感染症に対するインフラ整備のおくれ
- ・社会のデジタル化のおくれ

I. 現代社会と情報リテラシー

人間と社会

- | | |
|--------|---------|
| ・旧石器時代 | 約500万年前 |
| ・農業化社会 | 約1万年前 |
| ・産業化社会 | 約200年前 |
| ・情報化社会 | 約30年前 |

情報化社会のキーワード

- ・情報のデジタル化 コンピュータ、スマートフォン
- ・情報のネットワーク化 インターネット

資源としての情報

資源

モノ・エネルギー + 情報

情報リテラシー: 情報を知的資源として使いこなす技能

情報リテラシー

■ 情報から価値を創造する技能

■ 情報を再生産する技能

- ・情報の収集 : 必要な情報を探索し入手する
- ・情報の分析 : 収集した情報を分析し理解する
- ・情報の構成 : 分析した情報に基づき新しい情報を構成する
- ・情報の発信 : 構成した情報を発表し公開する

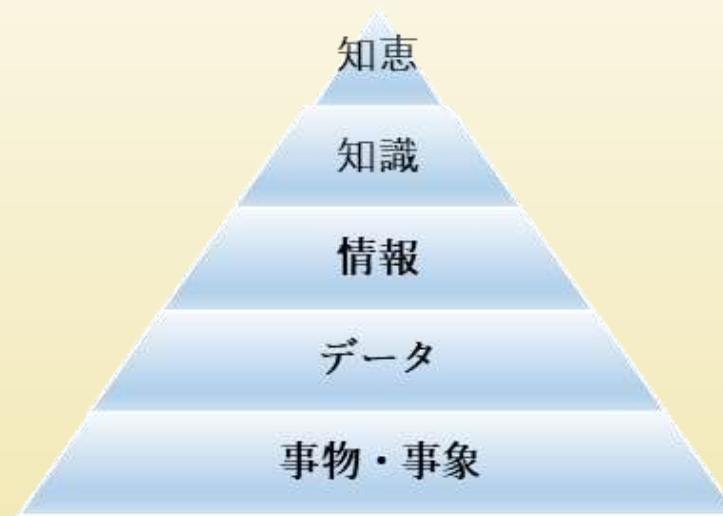
情報リテラシー

■ 情報から価値を創造する技能

■ 情報を再生産する技能

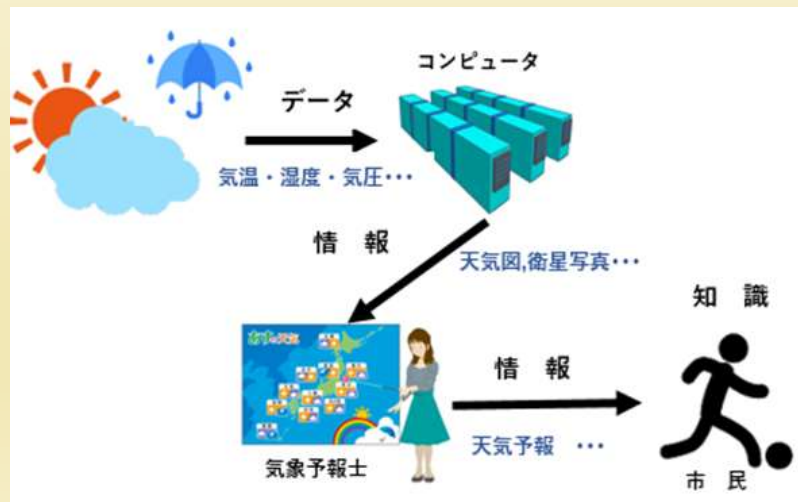
- ・情報の収集 : 必要な情報を探索し入手する
- ・情報の分析 : 収集した情報を分析し理解する
- ・情報の構成 : 分析した情報に基づき新しい情報を構成する
- ・情報の発信 : 構成した情報を発表し公開する

情報とデータ



情報ピラミッド

- ・事物・事象 : 五感による認識により見いだされるもの
- ・データ : 事物・事象を言葉や数字などに表現したもの
- ・情報 : 受け取った人にとって役に立つデータ
- ・知識 : 情報の受け手が情報を自分の中に体系的に蓄積したもの
- ・知恵 : 知識を昇華して、物事の理を悟り、適切に処理する能力



データ、情報、知識

情報システムとコンピュータ

■ 企業情報システム

- ・利潤追求
- ・企業活動の効率Up
- ・新しいビジネスへの進出

■ 公共情報システム

- ・住民サービス
- ・防災
- ・交通管制
-

■ 個人情報システム

- ・家庭用防犯
- ・AV(Audio/Visual)
- ・SNS
-

コンピュータの利用形態による分類

1) データの入出力手段による分類

- ・事務処理
- ・制御処理

2) 処理するデータの種類による分類

- ・文字・数値処理
- ・マルチメディア処理

3) 処理する単位とタイミングによる分類

- ・リアルタイム処理
- ・バッチ処理

4) 処理するコンピュータ数による分類

- ・集中処理
- ・分散処理

コンピュータの利用形態による分類

5) システムの公開性による分類

- ・オープンシステム
- ・クローズドシステム

II. 情報分析

情報分析の目的

- 評価的な問題解決
 - ・価値判断
 - ・問題認識
- 指令的な問題解決
 - ・意思決定
 - ・行動指示

抽象的な概念構造

■ 具体的なモノや事象を抽象的な概念構造に置き換える

- ・組織の構造 ⇒ チャート化
- ・構造物 ⇒ モデル化(骨組モデル、格子モデル..)
- ・コンピュータのシステム開発
特定の業務処理アプリケーション開発
業務分析 ⇒ 要求分析(仕様化) ... スキーマ設計

■ 具体的な対象世界から抽象的な概念構造を見つけ出す手順

- 1)対象世界の構成要素を列挙する
- 2)構成要素の状態や属性を明らかにする
- 3)構成要素間の関係を明らかにする
- 4)構成要素の状態変化やダイナミクスなど、動的な側面を明らかにする

チャート表現

■ モデル化

対象世界からモデル(抽象的な概念構造を表現したもの)を導出すること

■ モデル化の利点

- ・全体像が把握しやすくなる
- ・説明や説得がしやすくなる
- ・結果や効果が予想できるようになる
- ・未知の事物・現象に対処できるようになる
- ・作業の分担や協同が円滑に行えるようになる
- ・共通理解や知識共有が促進される
- ・問題点や課題が見えてくる

■ チャート

モデルを記述する図・表・グラフ等のこと

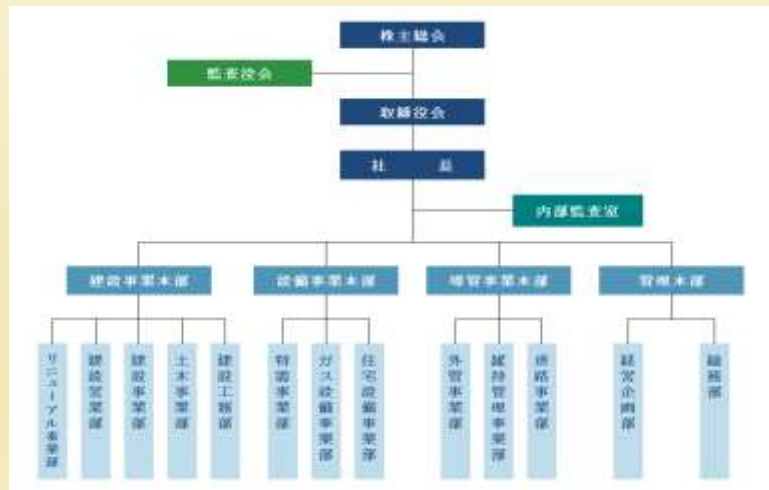
1) 構造・関係を記述するチャート

(a) -1 分類階層図

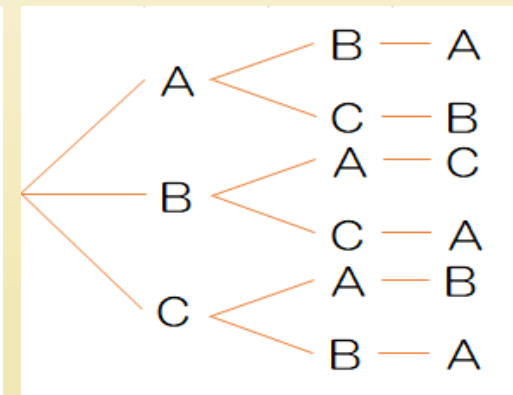
対象が多数の構成要素からなっており、階層構造をとるときに有効

(a) -2 樹形図

データやファイルなどの階層構造を記述する場合や順列や組み合わせを表現する場合に有効



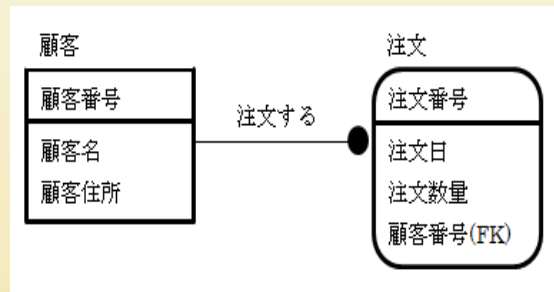
分類階層図の例



樹形図の例

(b) E-R図 (Entity Relationship Diagram)

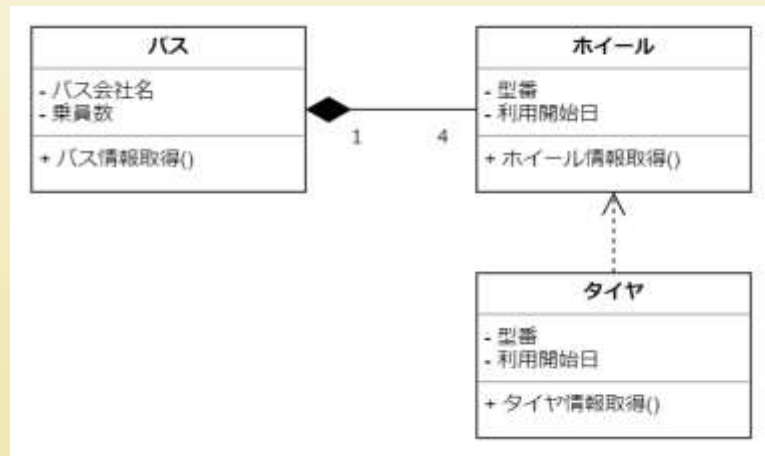
複数のエンティティ (実体) と、それらの間の関係を記述する



E-R図の例

(c) オブジェクト図

UML(Unified Model Language)というソフトウェア設計手法で提唱された図
E-R図とよく似ている / E-R図よりも記述量が多い



オブジェクト図の例

(d) KJ法

チャート表現ではなく、大量のデータから構造・関係を発見するための手法
ブレインストーミングによって収集した情報を整理する

■KJ法の手順

- 1.対象となる情報を集める
- 2.情報1件ごとに1枚の情報カードに書き込む
- 3.情報カード群を机などに並べ、類似のカード群をまとめ、グループ化する
- 4.グループに属するカードが多い場合には、さらにサブグループに分割する
- 5.グループやサブグループに表題（ラベル）を付け、ラベル間の関係を明らかにする

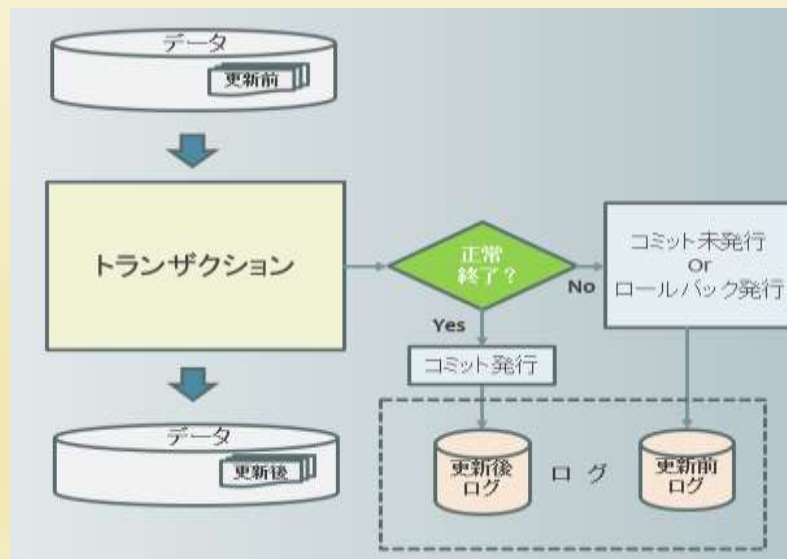


KJ法の例

2)動作・プロセスを記述するチャート

(a) フローチャート

処理順序が明白に決まっている事からを記述するのに適している

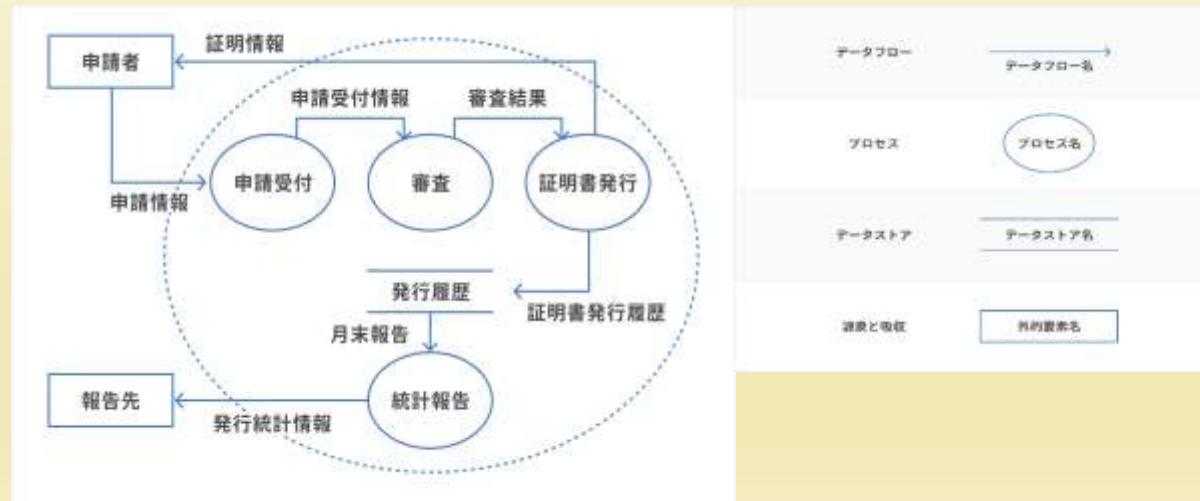


フローチャートの例

(b) DFD (Data Flow Diagram)

業務分析に有効で、システム開発の分野では広く使われている

- ・プロセス : 入出力のある処理.円または長円で記述
- ・データ源泉/データ吸収 : データの発生源と吸収先.長方形で記述
- ・データストア : データを記憶するファイル、平行な2本線で記述
- ・データフロー : データの流れ、矢印で記述



DFDの例

(c) 状態遷移図

対象がいろいろな状態に変化するとき、どのような状態をとり得るかを
列挙し、ある状態からほかの状態へ移るときに（状態遷移）どんな刺激
がもとになるかを整理するのに適している



状態遷移図の例

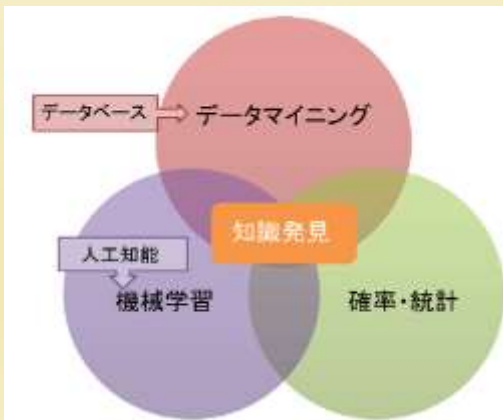
III. データマイニング

データマイニングとは

統計学、パターン認識、人工知能等のデータ解析の技法を大量のデータに網羅的に適用することで知識を取り出す技術のこと

機械学習とは

さまざまなアルゴリズムを用いてデータから反復的に「学習」することで、人間が探すべき場所を明示的にプログラムしなくても、コンピューターが自律的にデータから知識を導き出す技術のこと



データマイニング・機械学習・確率統計

データマイニングと機械学習

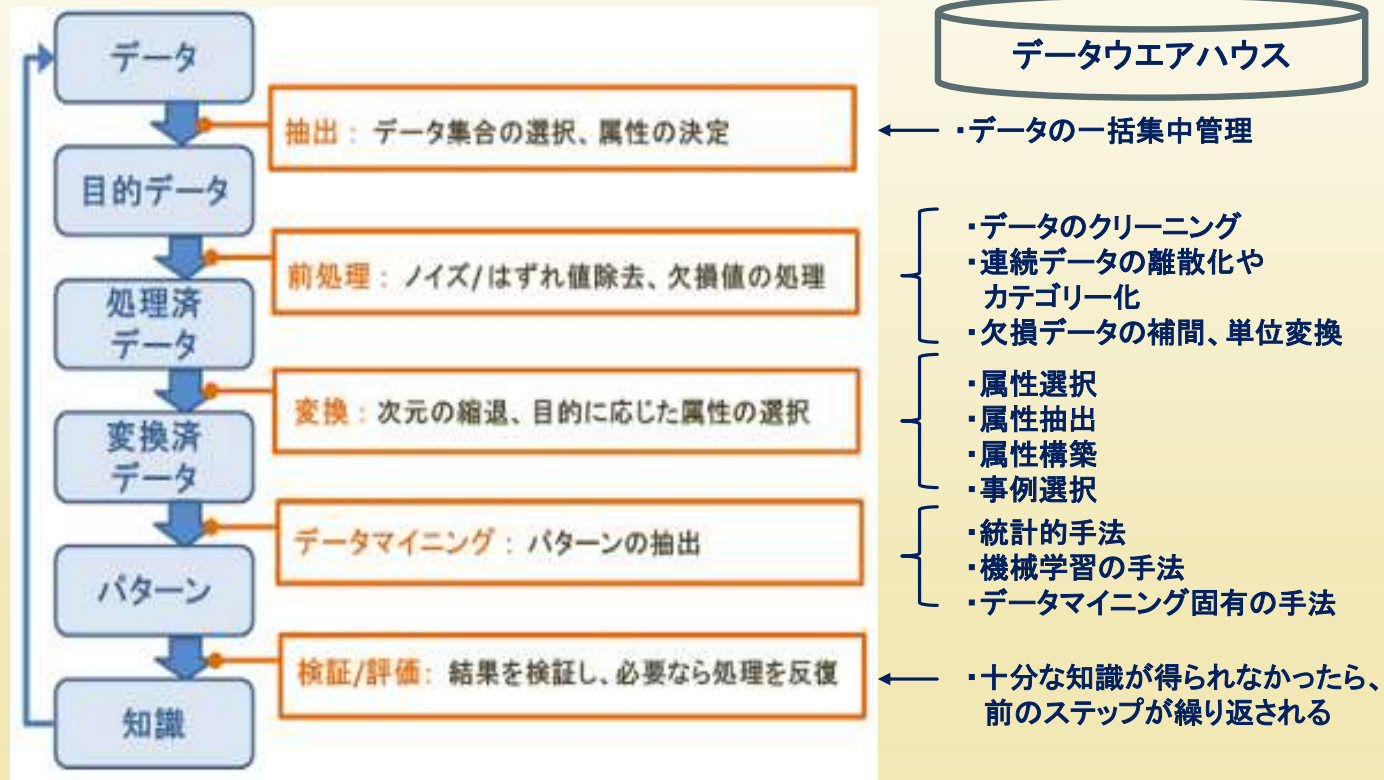
■ 共通点

- データから興味ある規則性、パターン、概念を見つけようとする技術
- データの不確かさの扱いについては統計学の手法に負っている
- 科学者の代替えとなるコンピュータシステムではなく、科学者を支援することを目標する技術

■ 異なる点

- データマイニングは、データが主役で、背後に意味のある構造の存在を仮定しないが、
機械学習は、データに潜む概念やメカニズム(モデル)の存在を仮定する
- データマイニングは、テラバイト 以上のデータを扱うことも要求されるが、
機械学習は、(現状の技法では)それほど大量のデータは扱えない。
- データマイニングの最終目的は、意思決定であり、単なる予測ではない
獲得した知識が実行可能であることが重要

データマイニングのプロセス



良い知識が発掘できるかどうかは、良い知識発見アルゴリズムに、良いデータを与えることができるかどうかにかかっている

データマイニングの実際(活用事例)

■ 金融分野

- ・生命保険の潜在的解約候補顧客発掘
- ・効果的なダイレクトメール宛先候補顧客
- ・過去の消費者ローン与信審査データからの半無人化ルール発掘
- ・ // 消費者ローン無人申し込み機の開発
- ・クレジットカード使用記録からの不正利用パターンの発掘

■ 流通・小売分野

- ・POSデータからの流通全般の業務知識の発掘
- ・ // 店内での販売促進用知識の発掘
- ・ // 有望顧客の洗い出しの発掘

■ 製造分野

- ・マーケティングへの適用
- ・社内の文書やマニュアルの検索システム

※現状では製造業固有のデータマイニングの適用事例は少ない

■ 通信分野

- ・膨大な通信ログから特徴的パターンを発掘し、
不正使用や不正アクセスの検出
- ・実時間でマイニングしなければならないスパムメールの判定など

※豊富なデジタルデータの蓄積があるので、適用範囲は広範

■ 製薬・医療分野

- ・化学化合物分子構造と人体への生理活性との相関解析
 - ・遺伝子発現形態と生理学的効果の相関解析
 - ・過去の膨大な論文,調査結果,治験情報などから得られる
科学的根拠に基づく診断・治療
- ※適用はまだ緒についたばかりだが、有望な見通しを示唆する成果が出はじめている。

IV 情報社会と危機管理

情報セキュリティとは

- ・インターネットやコンピュータを安心して使い続けられる
- ・貴重な情報が外部に漏れたり、ウイルスに感染したりしない
- ・蓄積したデータが壊れない、壊されない
- ・普段使っているサービスが急に使えなくなったりしない

情報セキュリティの三大基本理念 ～ CIA ～

①機密性 (Confidentiality)

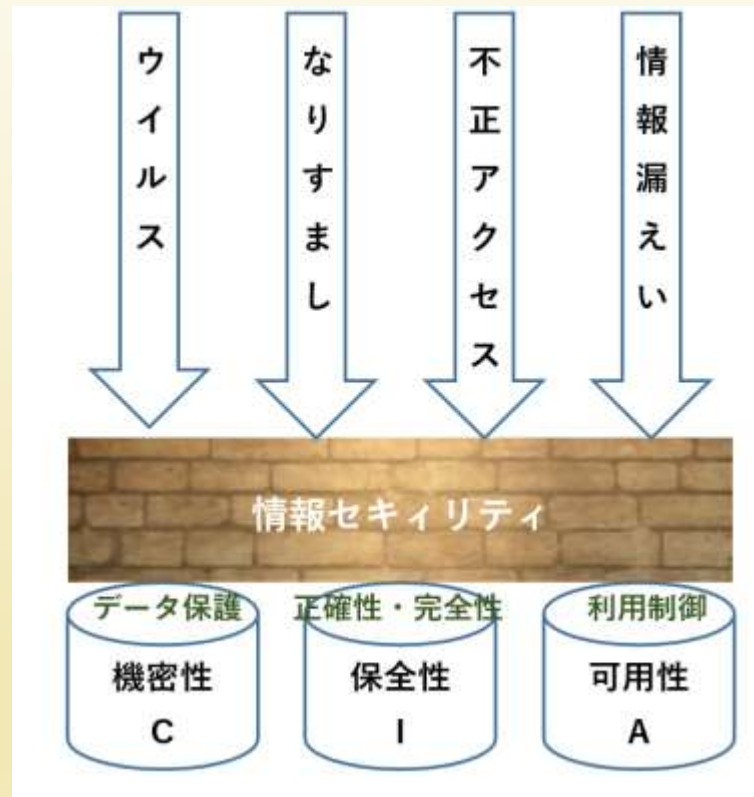
許可された者だけがアクセスできる

②保全性 (Integrity)

情報の内容が正確に保たれる

③可用性 (Availability)

必要なときに必要な情報資産にアクセスできる



情報セキュリティの三大基本理念

情報セキュリティ対策

- 技術的な対策
- 金銭的な対策
- マネジメントの対策
- 教育の対策
- 精神的な対策
- 法的な対策

技術的な情報セキュリティ対策

1)個人認証技術

2)暗号化技術

①共通鍵暗号方式

②公開鍵暗号方式

3)電子署名

1)個人認証技術

■ ICカード 、 ID&パスワード

■ 生体認証技術（バイオメトリクス）

・指紋 ・声紋 ・血管形状 ・眼球の虹彩模様 等



代表的なバイオメトリクス

2)暗号化技術

①共通鍵暗号方式

暗号化と復号に同じ鍵を用いる暗号方式

代表的な暗号アルゴリズム : DES(Data Encryption Standard)

・鍵のサイズが大きいほど、解読されにくい (1024bit)

※共通鍵方式の問題点 : 暗号鍵が他人に漏れると解読されてしまう

①公開鍵暗号方式

暗号化と復号に用いる鍵が異なる暗号方式

代表的な暗号アルゴリズム：

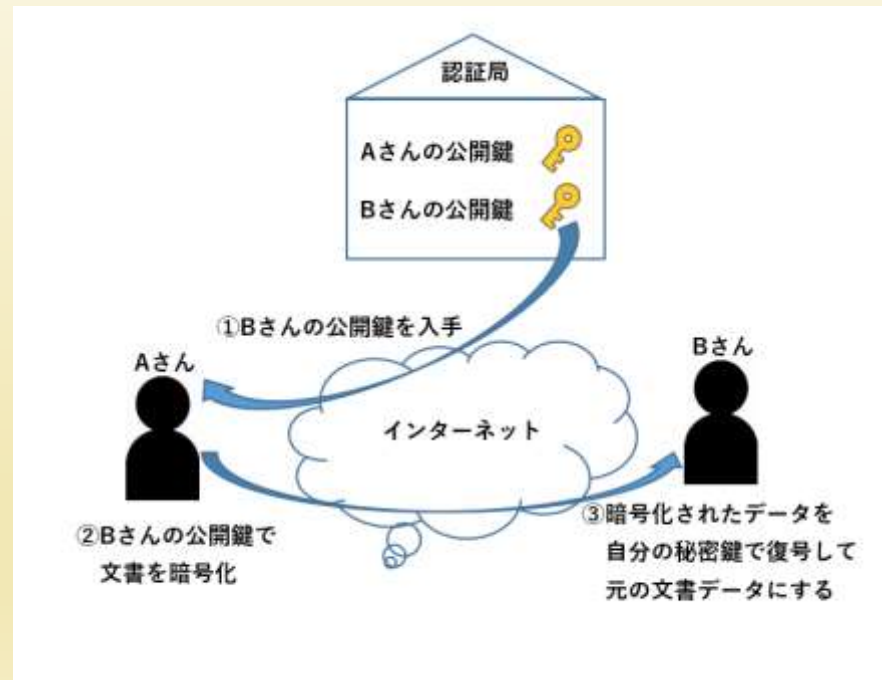
RSA(Rivest-Shamir-Adleman cryptosystem)

- 巨大な素数同士をかけ合わせた整数を素因数分解するのが困難であることを利用
- 公開鍵暗号は対になる二つの鍵を用いる
 - 暗号化する鍵を公開する(公開鍵)
 - もう一方の秘匿する(秘密鍵)で複合する
 - 暗号化に使った公開鍵では復号できない
 - 一方の鍵からもう一方の鍵を割り出すのが困難
 - 巨大な素数同士をかけ合わせた整数を素因数分解するのが困難であることを利用

※公開鍵方式は、自分の秘密鍵を他人に教えないですむため、安全性がより高い

PKI (Public Key Infrastructure) : 公開鍵暗号基盤

CA (Certification Authority) : 各人の証明書を保管し、要求に応じて公開鍵発行

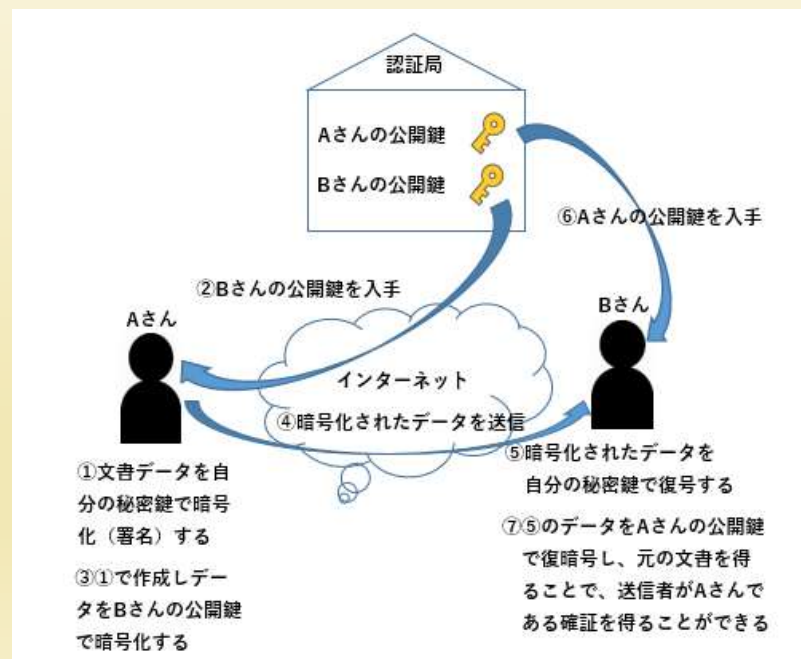


公開鍵暗号方式

2)電子署名

公開鍵暗号方式 : 暗号化／**公開鍵** 復号／**秘密鍵**
 電子署名 : 暗号化／**秘密鍵** 復号／**公開鍵**

※公開鍵で復号できたということは、秘密鍵の持主が暗号化したということ
 ※電子署名は「なりすまし」を防ぐ技術



電子署名

V 情報社会と権利・法律

知的所有権(知的財産権)とは

- 情報は複製が可能
- 情報を利用すると／利益を生み出す／被害を与える／娯楽を得る／
- 情報は資産である
- 資産(情報)は法的に守られる
- 情報の使用権 = 知的所有権(知的財産権)

知的所有権の種類

- 著作権
- 特許権, 実用新案権
- 商標権, 意匠権
- パブリシティ権
- e.t.c.

1)著作権

著作物 ： 著作権が対象とする物

著作権法 第2条

「思想又は感情を創作的に表現したものであって、文芸、学術、美術または音楽の範囲に属するものをいう」

著作権者の許可を要する行為

「複製, 上演, 演奏, 公衆送信等, 口述, 展示, 上映, 頒布, 貸与, 翻訳, 翻案等」

(a) 著作者人格権

原著作者に認められる権利

- | | | |
|---------|------|------------------------------|
| ①公表権 | ………… | 未公表の著作物を公表する権利 |
| ②同一性保持権 | ………… | 著作物の改変を禁止する権利 |
| ③氏名表示権 | ………… | 原著作者の本名やペンネームなどの氏名表示を義務づける権利 |

(b) 著作権行使ができない場合

以下の場合には、利用者が情報を利用できる(許可不要)

- ①批評のために引用を行う場合
- ②著作物を購入した人が自分で利用するためにバックアップを作る場合
- ③教科書などへの利用を行う場合
- ④授業の最中に使う場合
- ⑤点字にする場合

2)特許権, 商標権, 意匠権

著作権と対象とする物が異なる

- | | | |
|------|-------|--------------|
| ①特許権 | | 対象は、工業的な創意工夫 |
| ②商標権 | | 対象は、商業的な創意工夫 |
| ③意匠権 | | 対象は、美観的な創意工夫 |

※著作権との違い

- ・登録が必要(著作権は登録不要)
- ・登録と登録維持には登録料の支払いが必要(著作権は支払い不要)
- ・国際条約がないので各国に届出が必要(著作権はベルヌ条約加盟国へ直ちに有効)

3)個人情報保護法（個人情報の保護に関する法律）

個人情報を保有・利用するすべての団体や事業者に対し、
取得や保存・利用に関する義務や、違反時の罰則などを定めている
小規模な事業者や町内会のような団体も対象

■個人情報の種類

- ・生存する個人の氏名や生年月日、住所、電話番号など
個人の特定・識別に用いることができるもの
- ・顔写真、所属先のメールアドレス、金融機関の口座番号
- ・DNA配列、指紋、声紋、顔貌、虹彩など身体に固有の特徴を符号化したデータ
- ・マイナンバー、パスポート番号、運転免許証番号、公的な識別番号・符号

■違反した場合

最大で6ヶ月以下の懲役または30万円以下の罰金

4)パブリシティ権

有名人などの写真や声、名前などを他人が勝手に利用することを禁じる権利

著作権の新しい流れ

著作者が自ら積極的に権利の一部を解放して、その著作物を上手に利用してもらおうという動きが活発になっている

(1) クリエイティブ・コモンズ (Creative Commons : CC)

考え方 : この条件を守れば私の作品を自由に使ってよいですよ

「All Rights Reserved」	著作権をすべて留保
「Some Rights Reserved」	中間
「No Rights Reserve」	パブリックドメイン (PD)



著作権マーク

(2) 自由利用マーク

文化庁が推進する

「著作者が自分の著作物を他人に自由に使ってもらってよい」
と考える場合に、その意思を表示するマーク

- 「プリントアウト・コピー・無料配布」OKマーク



- 「障害者のための非営利目的利用」OKマーク



- 「学校教育のための非営利目的利用」OK



(3) フリーソフトウェアとオープンソース

ソフトウェアについても、自由にコピーを許ことでさまざまな人の叡知を集めて良いものを作ろうという動きもある

(例) Linux (コンピュータのOS)

GNU (グニュー)

「一般公衆利用許諾契約書」GPL (General Public License)

- ・複製物を自由に頒布できる
- ・ソースコードを一部改変して新しいプログラムの一部として使用できる
- ・改変プログラムは第3者へ無償で利用許諾しなければならない

ナイチンゲールと統計

フローレンス・ナイチンゲール
(Florence Nightingale, 1820~1910年)

イギリスの看護師

社会起業家

統計学者

ランプの貴婦人



クリミアの天使

白衣の天使

看護教育学者

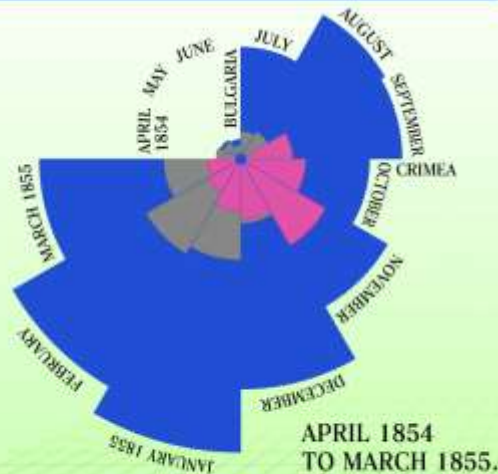
近代看護教育の母

Perry Pictures
/Library of Congress

ナイチンゲールと統計

Polar-Area Diagram (鶏のとさかグラフ)

Diagram of the Causes of Mortality
in the Army in the East (1858)



陸軍の衛生・保健医療改革

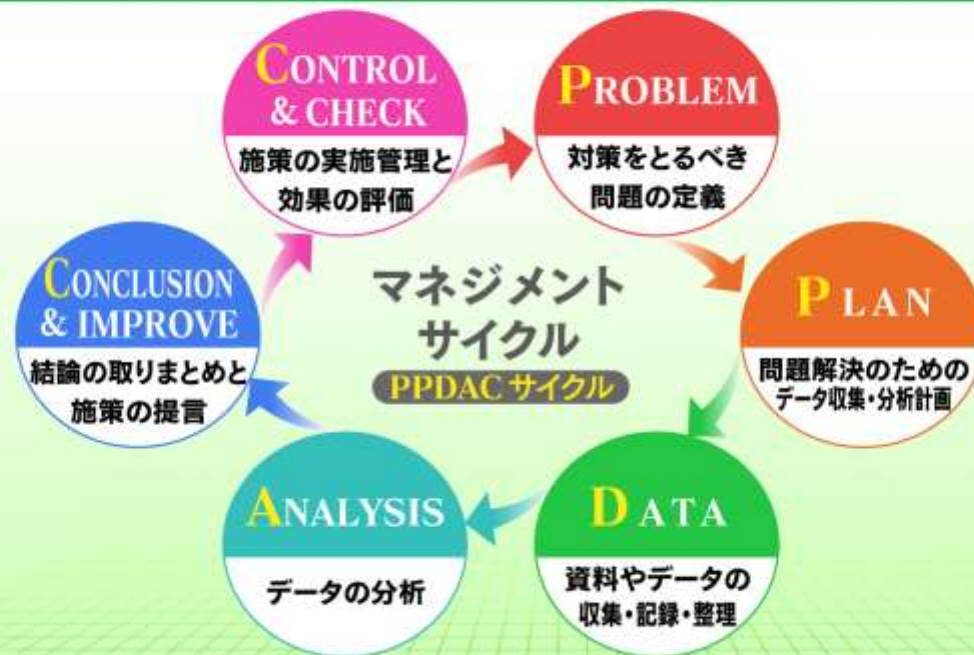
データによる説得

- クリミア戦争従軍兵士の原因別死亡データ
- データを集計し、死亡率を計算
- グラフを考案

神の御心を知るには
統計学を学ばなければならない

統計的マネジメントサイクル PPDACサイクル

ナイチンゲールが実践した統計的マネジメントのサイクル



統計の意味

統計『統^すべて計る』

集団の大量観察

- | | | |
|----------------------------------|-------|-------------------|
| ■ 国や社会の姿を映し出す | 『鏡』 | 現在の状況の 計測 |
| ■ 進むべき方向を示す | 『羅針盤』 | 将来の到達点の 予測 |
| ■ 経済や社会の内部構造に迫り,
そのメカニズムを解明する | 『内視鏡』 | 目標値への 制御 |

2005.6.10 内閣府経済社会統計整備推進委員会『政府統計の構造改革』

Statistics ← State (国家)

自然現象・モノの生産管理・輸送・収益管理・
人的資源管理・経営・リスク管理 など

マネジメントにおける科学的な意思決定

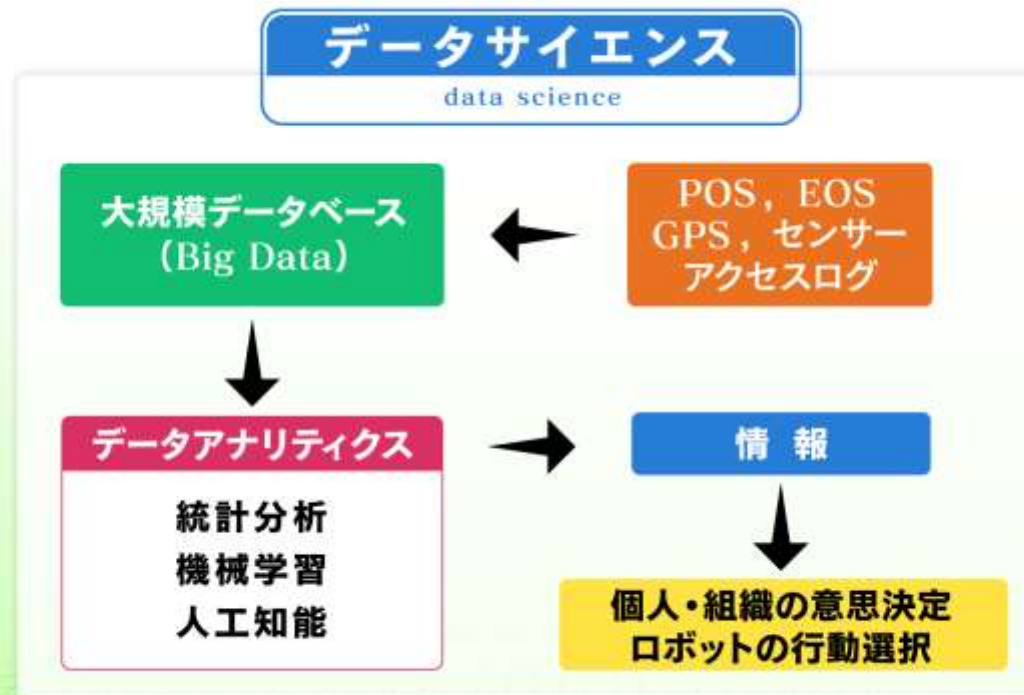
データとは？ 客観的事実

データに基づく管理・マネジメント
KKD から FACT CONTROL へ

情報とは？ 個人や組織の意思決定に役立つ

エビデンス(科学的証拠)に基づく
(EBPM, EBM, EBP, ...)
説明責任 ➡ Evidence-based Society

データの取得から意思決定のプロセス



統計的分析手法

統計的分析手法(データ分析法)

- 統計数値や統計グラフを作成し,データの持つ情報を表現する



計量化と視覚化

- データを集め,集めたデータを分析・解釈し,背後にある現実の構造に関して結論を導き出すための方法



知識創造

統計の適用分野

統計の適用分野

医療と医薬

遺 伝 学
臨 床 試 験
疫 学
製 薬
健 康 科 学

環 境

農 業
自 然 保 護
森 林 破 壊
野 生 動 物 保 護

行 政

企 画 立 案
リ ス ク 管 理
行 政 評 価
公 的 統 計 ・ E B P M

自 然 科 学

気 象
災 害
化 学
物 理

社 会 科 学

経 済
社 会 ・ 犯 罪
集 団 行 動 分 析
文 化 計 量 学
S N S

ビ ジ ネ ス

経 済 分 析
経 営 分 析
マ ー ケ テ ィ ン グ
情 報 シ ス テ ム
ゲ ー ム ・ ギ ャ ン プ ル
ス ポ ー ツ

質的変数(データ)と量的変数(データ)

質的変数(データ)

少数個のカテゴリーのどれかが
応答値となる観測項目

性 別
(男性・女性)

職 種
(事務職・技術職・管理職など)

改善度
(著効・有効・やや有効・無効・悪化)

量的変数(データ)

数量値(数直線上の値)として
記録される観測項目

売上高

生産指数

給与総額

在庫率

質的データと量的データ

■ テレビ番組視聴の
アンケート結果

性別	スポーツ	ドラマ
男	みる	みない
男	みる	みない
男	みる	みない
女	みない	みる
女	みる	みる
⋮	⋮	⋮

■ 証券会社の売上等のデータ

企業名	資本金(万円)	従業員数(人)	売上(百万円)
A証券(株)	20000	76	7277
B投資信託(株)	80000	87	16408
C証券(株)	13000	140	3970
D証券(株)	1215000	2789	97134
E証券(株)	25020	105	3972
⋮	⋮	⋮	⋮

■ 取引内容を示したデータ

担当者名	得意先名	業種名	商品名	平均単価	数量	金額
熊原	セルフ山上	ミニスーパー	ケチャップ	230	3	690
熊原	セルフ山上	ミニスーパー	烏龍茶	215	3	645
渡辺	サキコーポ	スーパー	ドライパン粉	330	3	990
渡辺	サキコーポ	スーパー	焼肉のタレ	169	3	507
渡辺	サキコーポ	スーパー	カップうどん	79	3	237
⋮	⋮	⋮	⋮	⋮	⋮	⋮

データが測定された物差し(尺度)の違い

質 的	名義尺度	区別のみ 順序はない	性別, 居住地, 職業区分, ハイ・イイエ	分 類
	順序尺度	区別に加えて 順序もある 差は取れない	5段階の評価 (満足・やや満足・どちら でもない・やや不満・不満)	分 類 クラスの併合 累 積
量 的	間隔尺度	差の大きさに意味が ある(等単位性) 比には意味がない	摂氏(華氏)温度, 知能指数, 偏差値, ...	足し算や 引き算が可 (合計, 平均)
	比率尺度	差も比も意味がある (絶対ゼロがある)	身長, 体重, 得点, ...	掛け算や 割り算が可

分布(distribution)

データのばらつき方の確率的なパターンもしくは、
パターンを数量的に把握したもの



データが集中している範囲,バラツキの大きさ,データの値や範囲を
指定した場合,そこに全体の何%のデータが含まれるか・・・etc.
を教えてくれるもの



現状の把握・予測・管理・マネジメント

質的データと量的データ

取引内容を示したデータ

担当者名	得意先名	業種名	商品の種類	平均単価	数 量	金 額
石 崎	山上商店	ミニスーパー	ケチャップ	230	15	3,450
石 崎	山上商店	ミニスーパー	烏龍茶	215	5	1,075
渡 辺	サキコーポ	スーパー	ドライパン粉	330	20	6,600
渡 辺	サキコーポ	スーパー	焼肉のタレ	169	40	6,760
渡 辺	サキコーポ	スーパー	カップうどん	79	20	1,580
渡 辺	石内食品	食料品店	桃缶詰	135	18	2,430
渡 辺	石内食品	食料品店	たくあん	275	45	12,375
渡 辺	山口商店	食料品店	ラーメンみそ味	68	100	6,800
渡 辺	FF長野店	コンビニ	ドライパン粉	330	48	15,840

ばらつき

データを集計して、値の起こり易さのパターン(分布)を探る

度数分布表(量的データの場合)

データ値

取引金額の階級 単位(千円)	階級値	取引件数 (度数)	累積件数 (累積度数)	構成割合(%) (相対度数)	累積構成割合(%) (累積相対度数)
0~20	10	3	3	3.23	3.23
20~40	30	7	10	7.53	10.75
40~60	50	11	21	11.83	22.58
60~80	70	12	33	12.90	35.48
80~100	90	19	52	20.43	55.91
100~120	110	16	68	17.20	73.12
120~140	130	10	78	10.75	83.87
140~160	150	6	84	6.45	90.32
160~180	170	5	89	5.38	95.70
180~200	190	4	93	4.30	100.00
合 計		93		100	

度数分布表(量的データの場合)

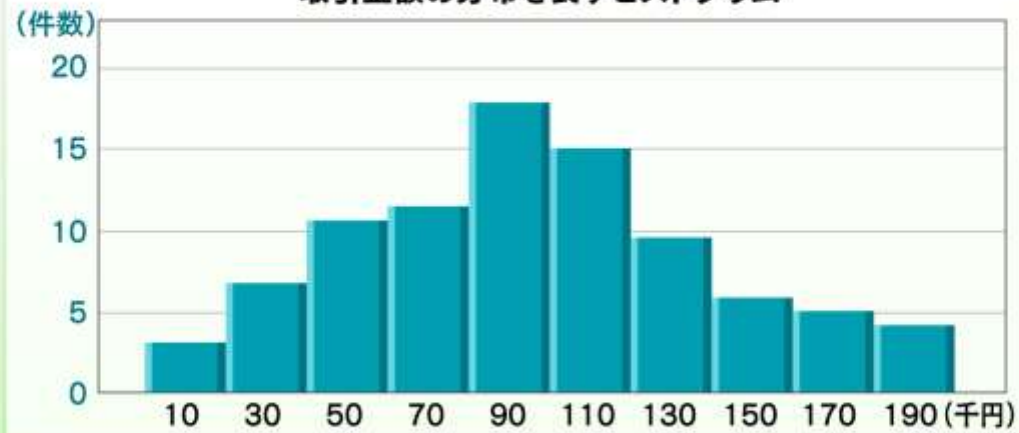
データ値

取引金額の階級 単位(千円)	階級値	取引件数 (度数)	累積件数 (累積度数)	構成割合(%) (相対度数)	累積構成割合(%) (累積相対度数)
0~20	10	3	3	3.23	3.23
20~40	30	7	10	7.53	10.75
40~60	50	11	21	11.83	22.58
60~80	70	12	33	12.90	35.48
80~100	90	19	52	20.43	55.91
100~120	110	16	68	17.20	73.12
120~140	130	10	78	10.75	83.87
140~160	150	6	84	6.45	90.32
160~180	170	5	89	5.38	95.70
180~200	190	4	93	4.30	100.00
合 計		93		100	

ヒストグラム

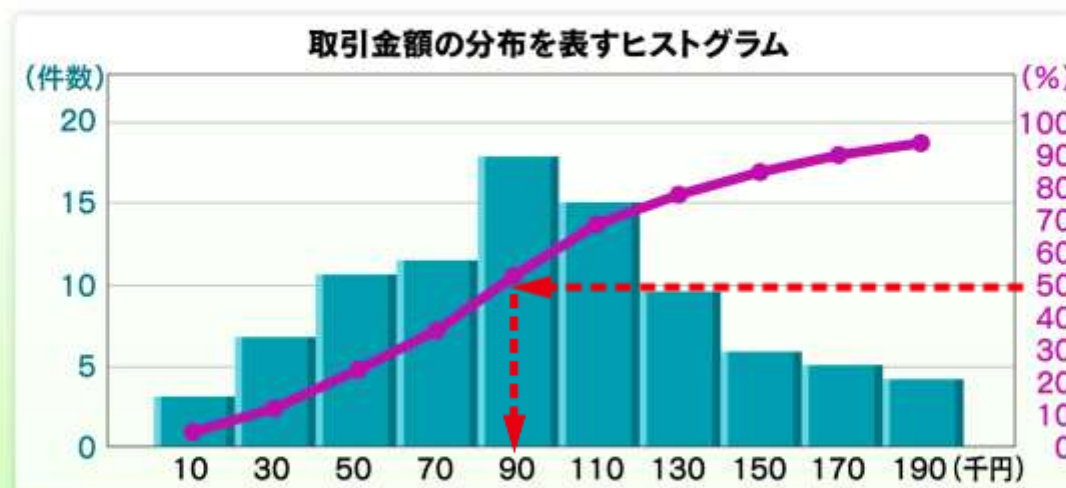
量的データの分布を表すグラフ

取引金額の分布を表すヒストグラム



ヒストグラム

量的データの分布を表すグラフ

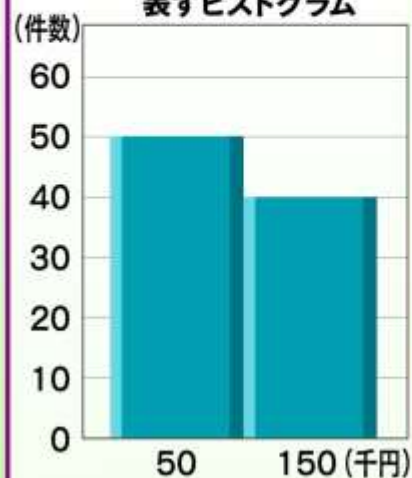


中央値 (データを半分に分ける値)

階級幅の取り方(悪い例)

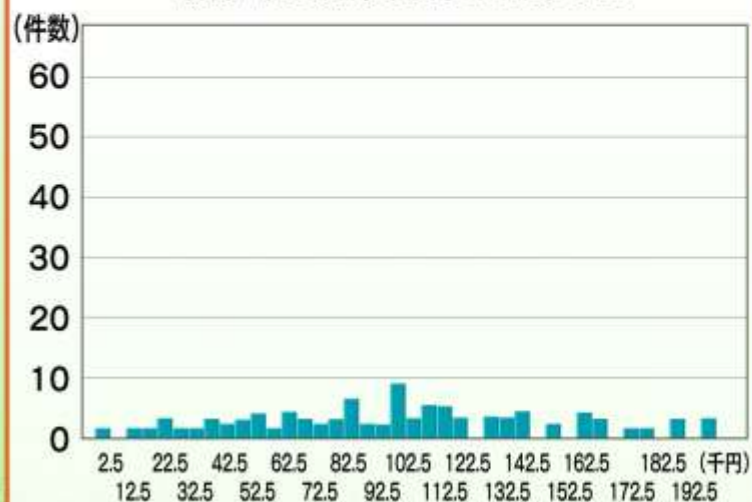
幅が大きすぎる

取引金額の分布を表すヒストグラム



幅が小さすぎる

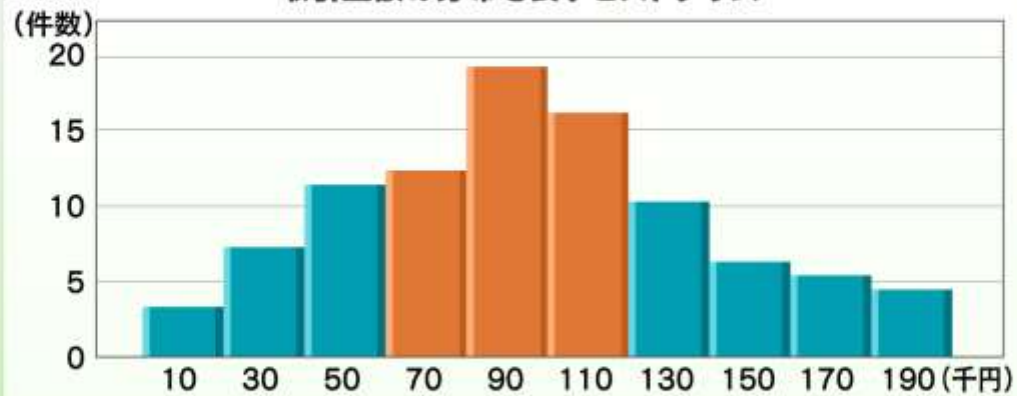
取引金額の分布を表すヒストグラム



階級幅の取り方

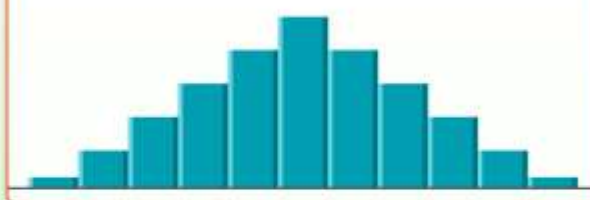
分布の山が見える適度な幅を設定

取引金額の分布を表すヒストグラム

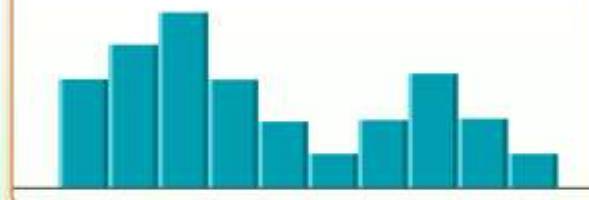


代表的なヒストグラムの形

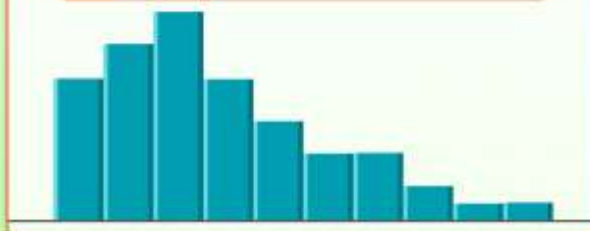
A 単峰性(山が一つ)で左右対称な形



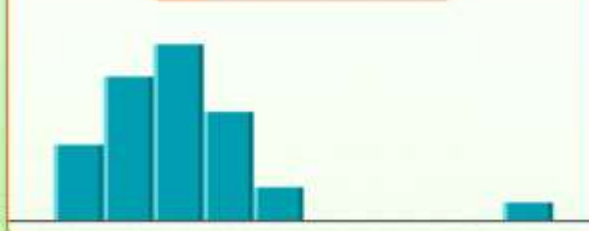
B 多峰性(山が二つ以上)を示す形



C 非対称な(左右対称でない)形

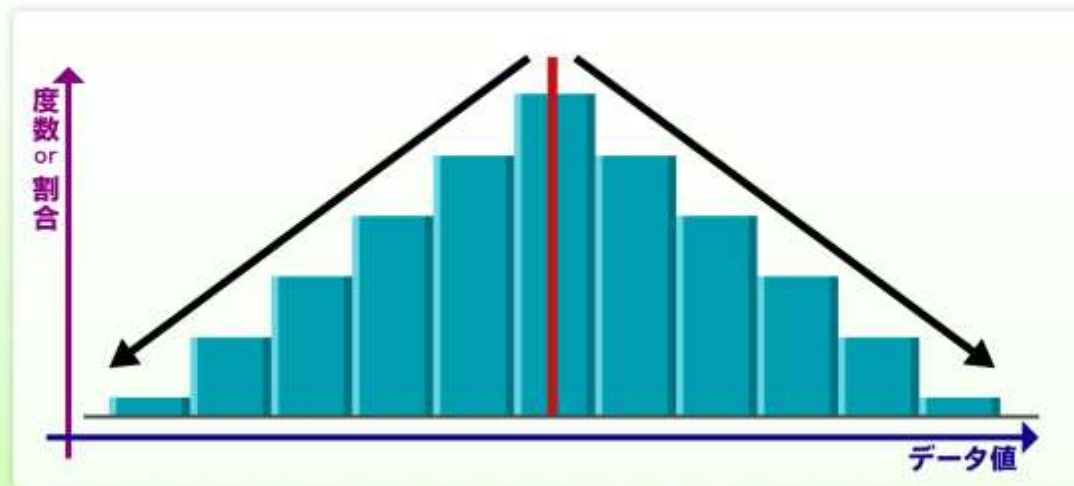


D 外れ値のある形



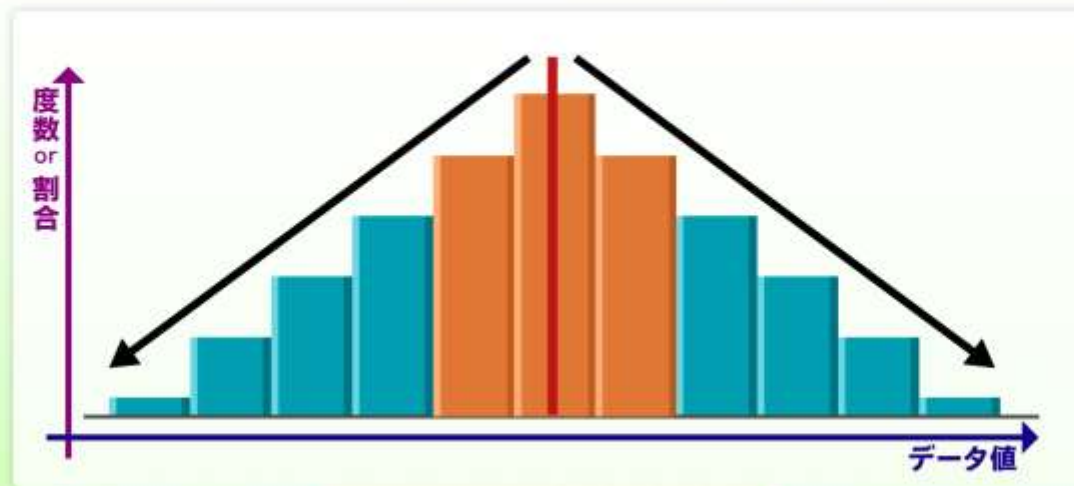
(A) 単峰性(山が一つ)で左右対称な形

同質な集団の代表的なデータのばらつき方



(A) 単峰性(山が一つ)で左右対称な形

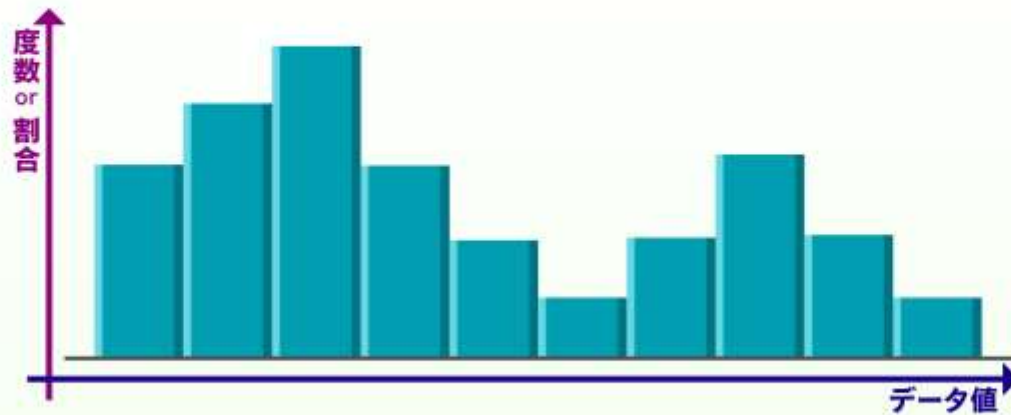
同質な集団の代表的なデータのばらつき方



分布の中心傾向は1ヶ所

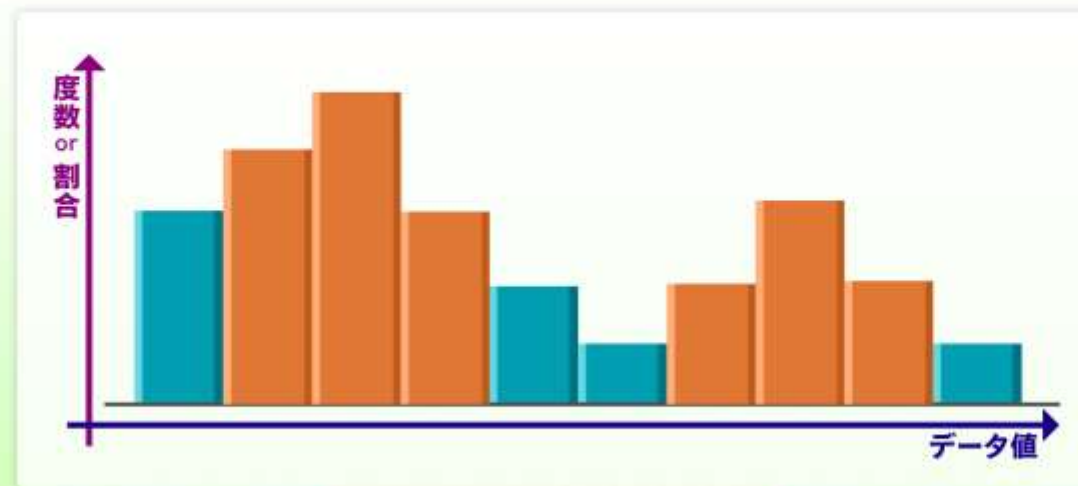
(B)多峰性(山が二つ以上)を示す形

データが集中している部分(峰、山)が2つ以上→異質な集団の混在の可能性



(B)多峰性(山が二つ以上)を示す形

データが集中している部分(峰、山)が2つ以上→異質な集団の混在の可能性



分布の中心傾向は2ヶ所

多峰性を示すヒストグラムの例



山の理由(分類基準)の探索

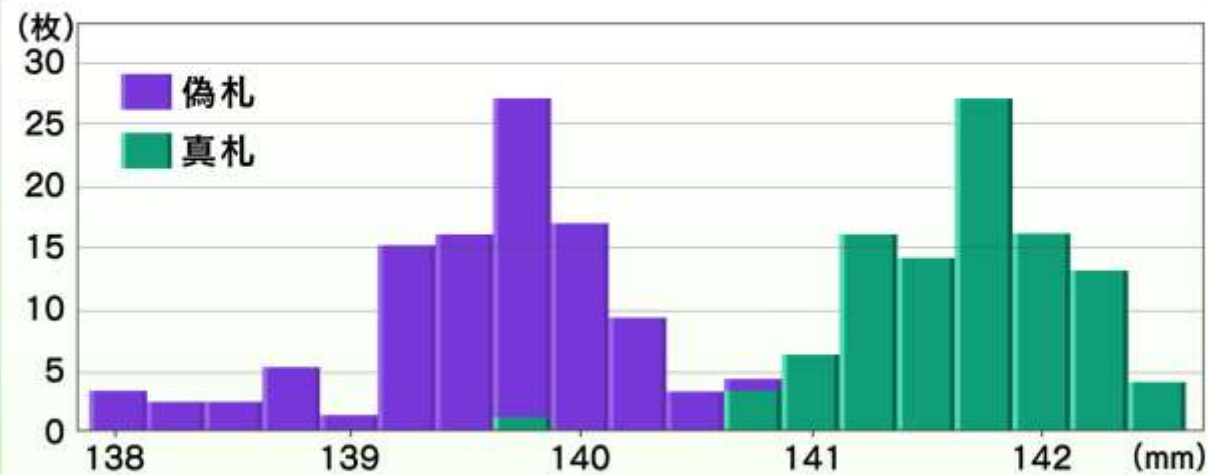
多峰性を示すヒストグラムの例



山の理由(分類基準)の探索

多峰性を示すヒストグラムの例

紙幣の絵柄の長さの分布(旧スイスフラン)



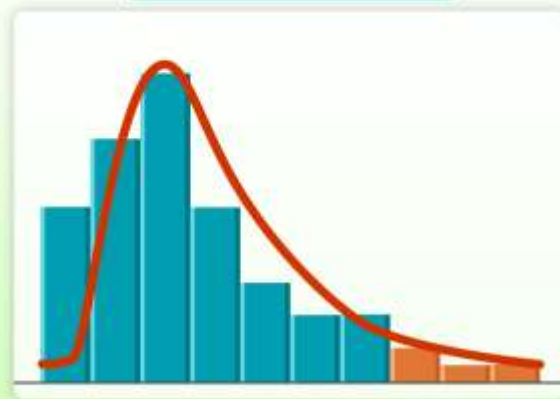
山の理由(分類基準)の探索

資料: Flury, B. and Riedwyl, H. (1988) Multivariate Statistics, A Practical Approach, Cambridge University Press.より

(C) 非対称な(左右対称でない)形

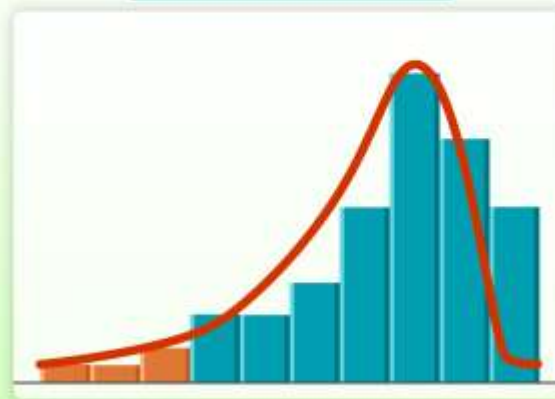
ゆが
歪んだ分布

右に歪んだ分布



値の大きな偏ったデータ

左に歪んだ分布



値の小さな偏ったデータ

貯蓄現在高階級別世帯分布

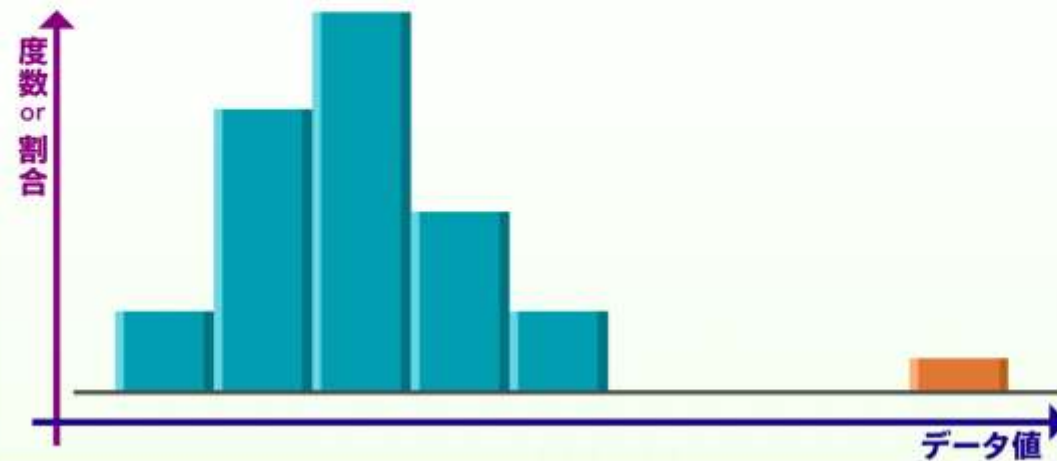
二人以上の世帯(2016年)



面積が度数や割合に対応する(階級幅が等間隔でない場合)

(D)外れ値(outlier)のある形

データの大部分が含まれるデータ値の区間から
かけ離れたところに位置するデータ



ヒストグラムを読む視点

多峰性の分布

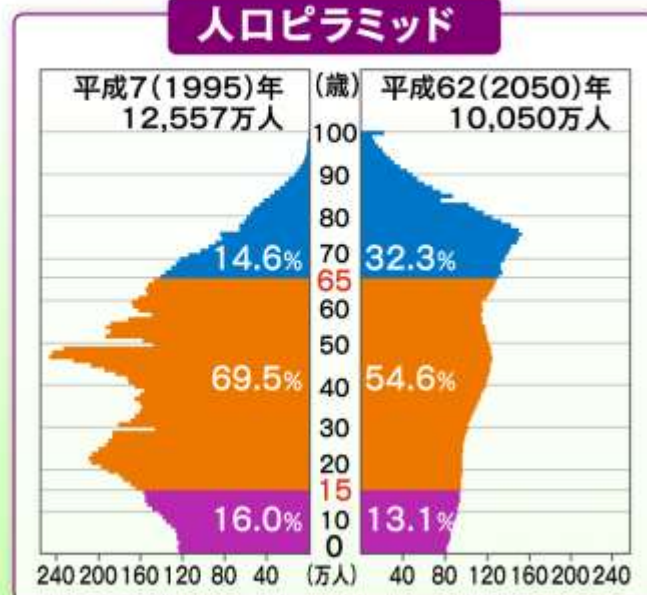
■ 分類 → 単峰性

単峰性の分布

- 分布の幅(範囲)
- 中心傾向(中心範囲)
- 外れ値・分布の裾(偏ったデータ)の異質性

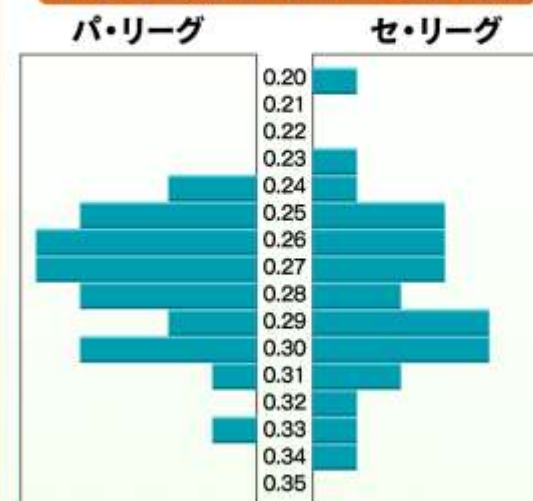
2つの集団の比較(背中合わせのヒストグラム)

人口ピラミッド



出典:第78回 人口問題審議会「少子化と人口減少社会を考える」

打者の打率分布(2016年)

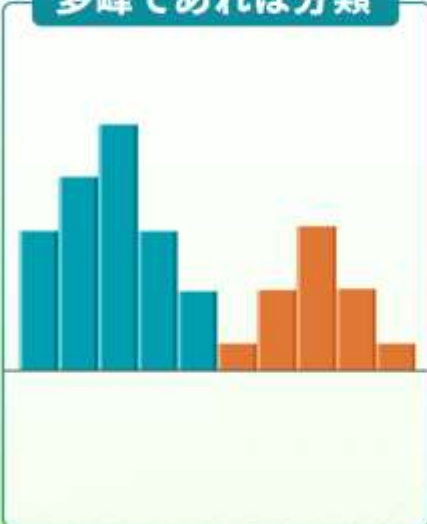


資料:「日本野球機構」公式サイトより

分布を読む視点

単峰か多峰か？

多峰であれば分類

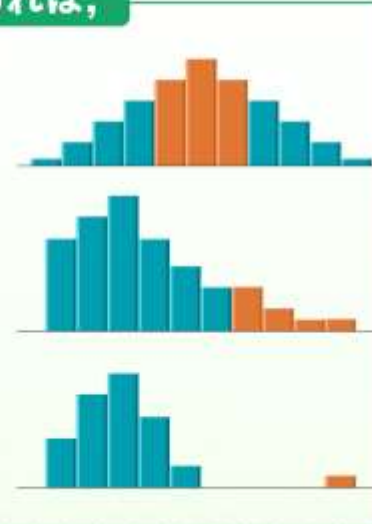


単峰であれば、

- 中心の位置(傾向)は？
- ばらつきの大きさは？
- 対称か非対称か？
- 外れ値はあるのか？



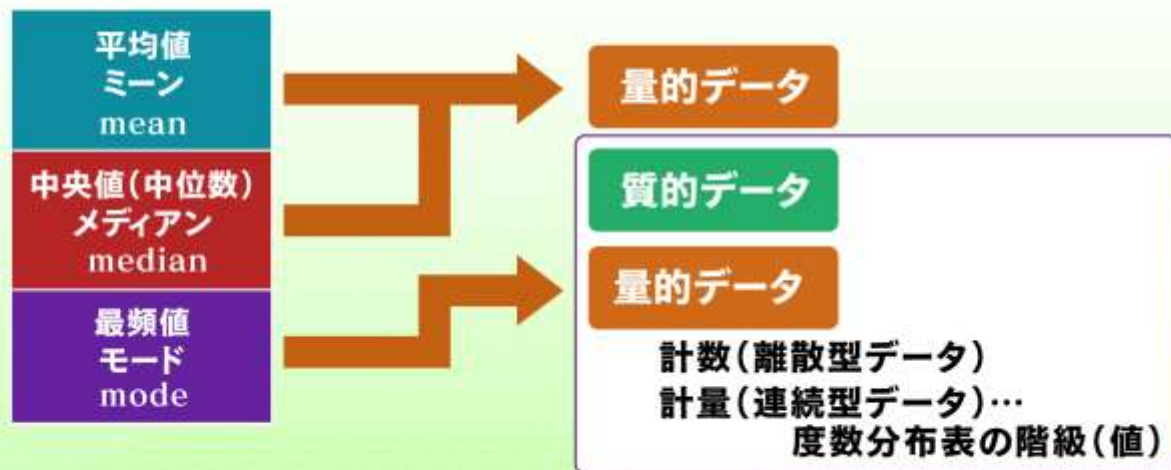
基本統計(量)



基本統計(量)basic statistics

平均値 ミーン mean	分 散	標準偏差 SD シグマ	変動係数 CV	わいど 歪 度	せんど 尖 度
中央値(中位数) メディアン median	四分位数 (四分位点)	四分位 範囲	最大値 MAX	最小値 MIN	範囲 レンジ
最頻値 モード mode	これらの値を読んで分布の形状を再現する				

分布の中心の位置を測る指標



平均 \bar{x} (エックスバー)

$$\bar{x} = \frac{\text{データ値の合計}}{\text{データの数}} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i$$

ケースの番号	データ値	記号
1	5	x_1
2	7	x_2
3	3	x_3
4	37	x_4
5	8	x_5
合計	60	x

例

$$\begin{aligned}
 \bar{x} &= \frac{5+7+3+37+8}{5} \\
 &= \frac{60}{5} \\
 &= 12
 \end{aligned}$$

中央値(中位数) \tilde{x} (エクステルダー)

データを大きさの順に並べたときにちょうど中央にくる値

データが **奇数個** のとき

3 5 **7** 8 37

↑
中央値

データが **偶数個** のとき

3 5 **○** 7 8

↑
中央値

$$\tilde{x} = \frac{5+7}{2} = 6$$

中央値(中位数) \tilde{x} (エクステルダー)

$x_{[1]}, x_{[2]}, \dots, x_{[n]}$

n 個のデータを大きさの順に並べたもの

$(x_{[1]} \leq x_{[2]} \leq \dots \leq x_{[n]})$

ケースの番号	データ値	記号
3	3	$x_{[1]}$
1	5	$x_{[2]}$
2	7	$x_{[3]}$
5	8	$x_{[4]}$
4	37	$x_{[5]}$

$$\tilde{x} = \begin{cases} x_{\left[\frac{n+1}{2}\right]} & (n \text{ が奇数}) \\ \frac{x_{\left[\frac{n}{2}\right]} + x_{\left[\frac{n}{2} + 1\right]}}{2} & (n \text{ が偶数}) \end{cases}$$

平均値は、データの値が釣り合う重心の位置



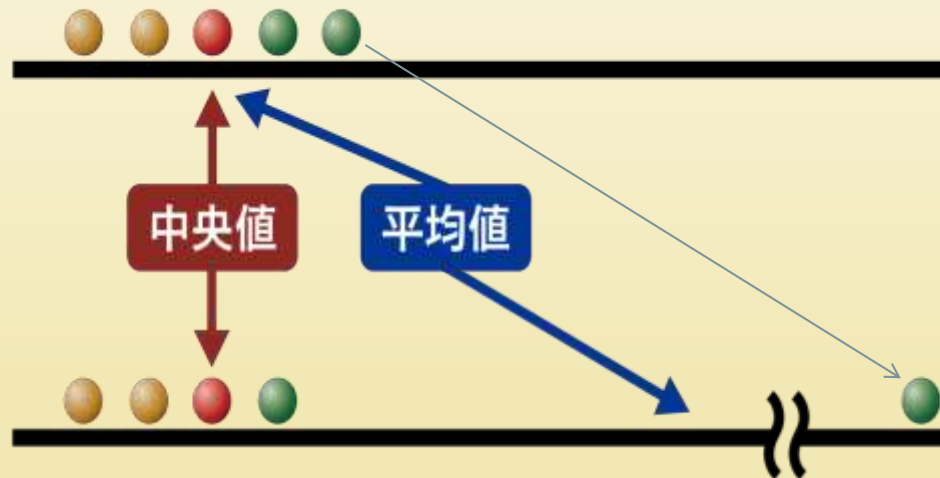
中央値は、データを半分に分ける位置(データの値)



中央値(中位数)の性質

ゆが
分布の歪み(偏り)や外れ値に対する頑健性(ロバストネス)がある

robustness

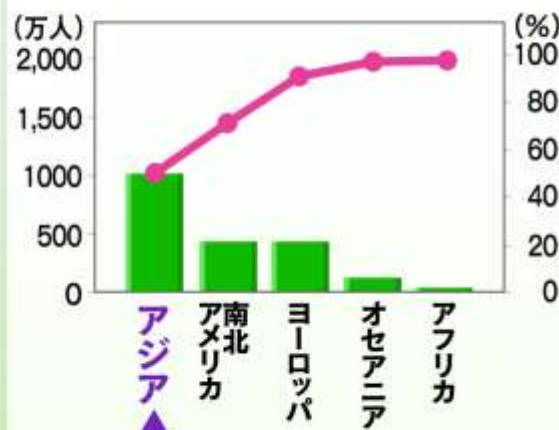


1つのデータ移動により平均値は大きな影響を受けるが中央値は影響を受けない

最頻値(モード, mode)

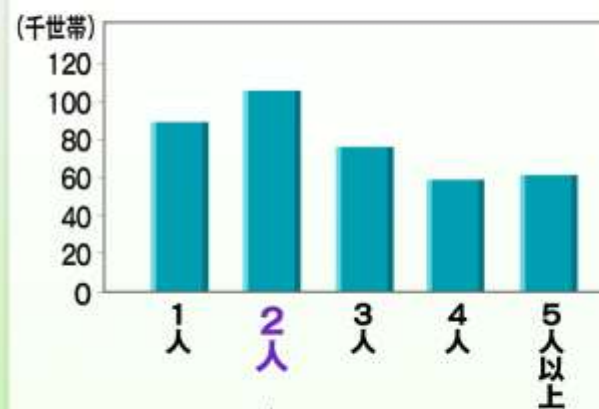
度数(その値をとるデータ数)が最も大きいデータ値

海外旅行先の分布を表すパレート図

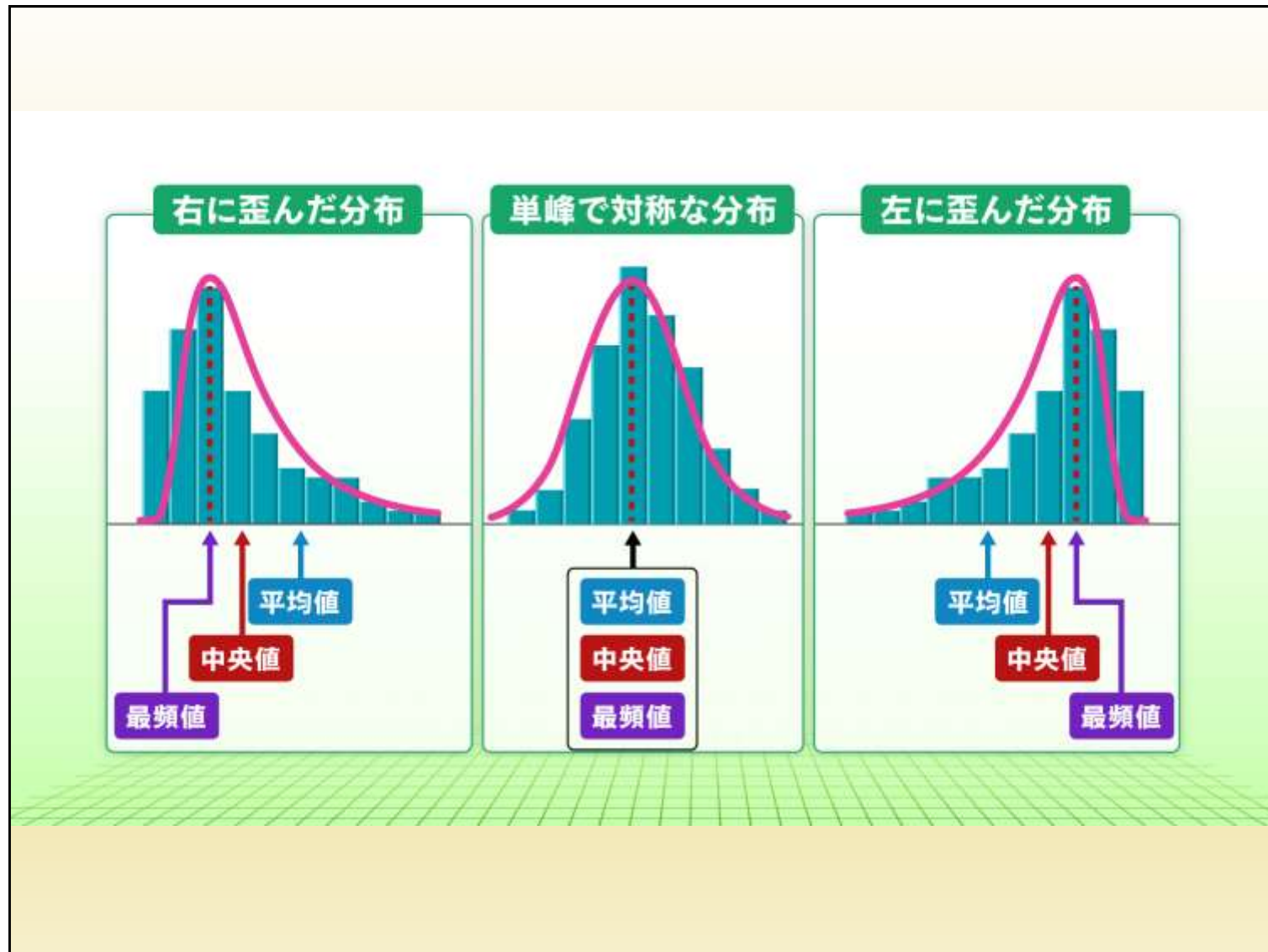


最頻値

世帯人員の分布(秋田県:平成17年)



最頻値



貯蓄現在高階級別世帯分布

二人以上の世帯(2016年)



面積が度数や割合に対応する(階級幅が等間隔でない場合)

貯蓄現在高の推移表

1世帯当たり平均

年 次	貯蓄現在高 (万円)
平成17年	1,728
平成18年	1,722
平成19年	1,719
平成20年	1,680

総額も
減少

資料：総務省統計局「家計調査」

多峰性の場合の平均は？

代表値としての意味は持たない

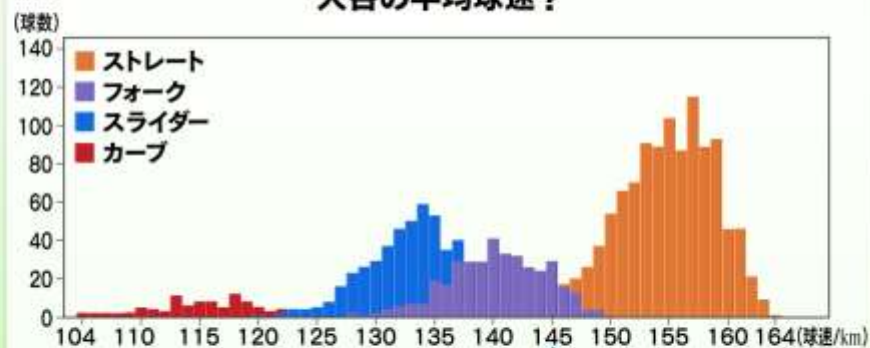
日本人の平均身長？



平均値

男女別の平均身長は意味がある

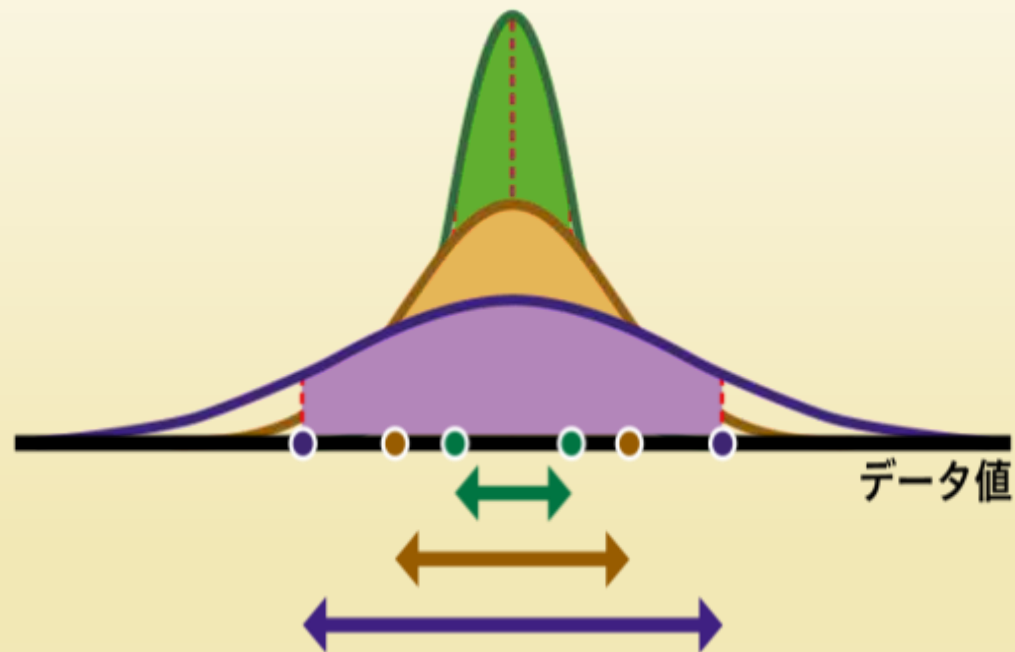
大谷の平均球速？



平均値

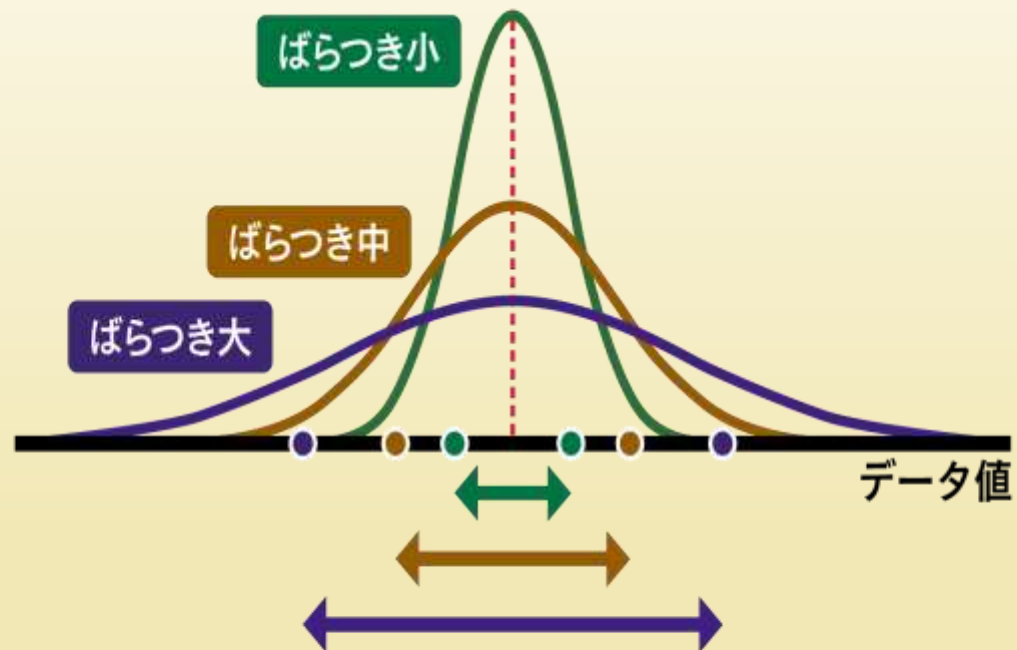
データ提供: データスタジアム

データのばらつきの大きさ



1つだけの基本統計量では、データの分布(ばらつき)を要約できない

データのばらつき大きさ



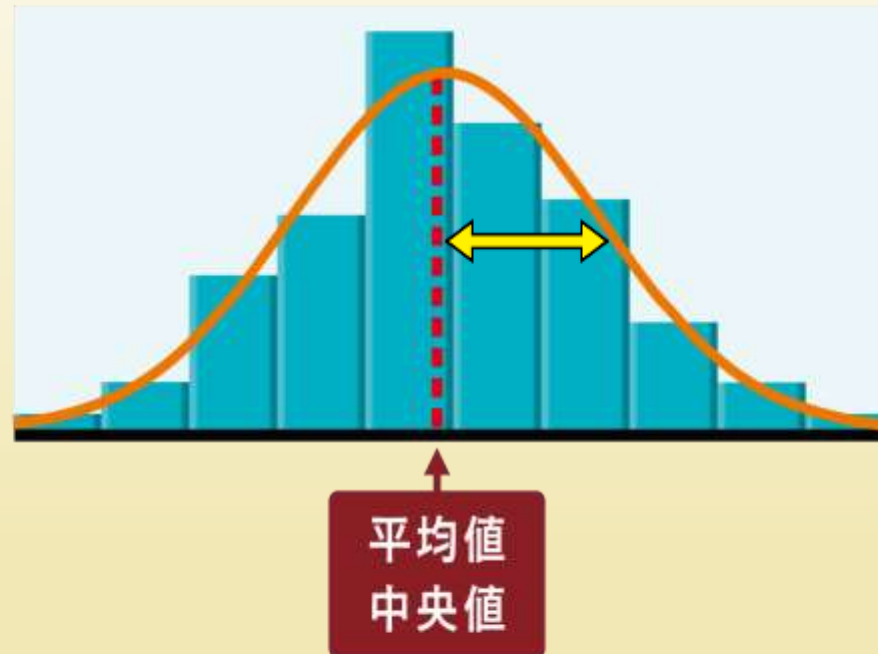
基本統計(量) basic statistics

平均値 ミーン mean	分散	標準偏差 SD シグマ	変動係数 CV	わいと 歪度	せんと 尖度
中央値(中位数) メディアン median	四分位数 (四分位点)	四分位 範囲	最大値 MAX	最小値 MIN	範囲 レンジ
最頻値 モード mode	これらの値を読んで分布の形状を再現する				

分布(単峰性)から情報を読むポイント

- 全体の過半数のデータが従う
中心的(標準的)な傾向
- 標準から外れた
少数個のデータの特徴

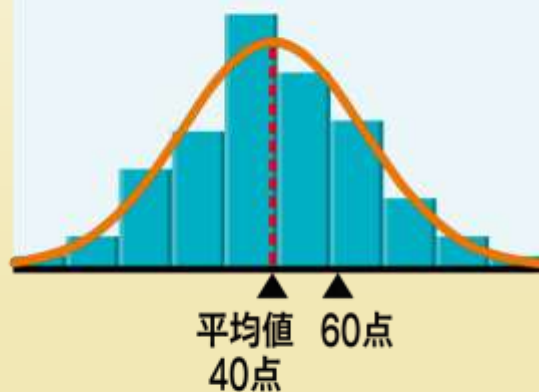
単峰で対称な分布



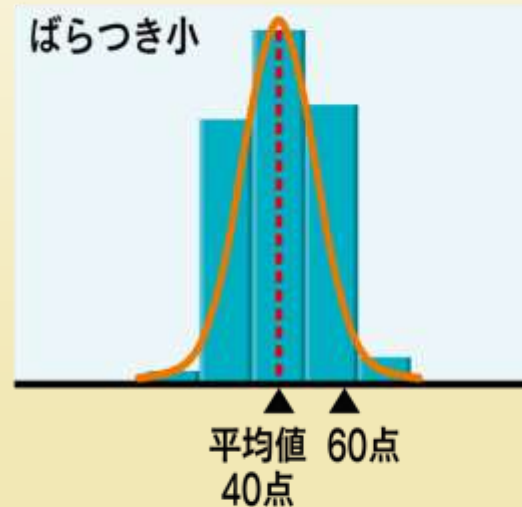
平均からの差(偏差)

数学・・・60点(全体の平均 40点)
 $60\text{点} - 40\text{点} = +20\text{点}$ 大or小?

ばらつき大



ばらつき小



偏差, 分散, 標準偏差

ケース番号	データの値(万円)	平均からの偏差(万円)	偏差の2乗(偏差平方)
1	9万円	-2万円	4
2	6万円	-5万円	25
3	12万円	+1万円	1
4	18万円	+7万円	49
5	10万円	-1万円	1
合計	55万円	0	偏差平方和 80
平均	11万円	0	分散 16

$$\text{標準偏差} = \sqrt{\text{分散}} = 4\text{万円}$$

偏差, 分散, 標準偏差

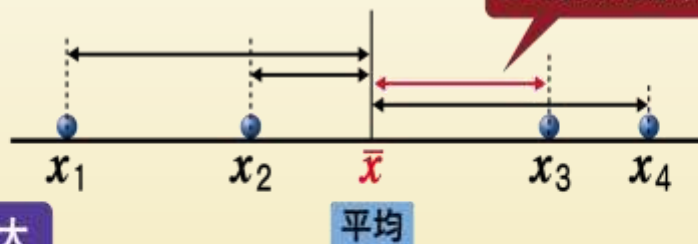
ケース番号	データの値(万円)	平均からの偏差(万円)	偏差の2乗(偏差平方)
1	x_1	$x_1 - \bar{x}$	$(x_1 - \bar{x})^2$
2	x_2	$x_2 - \bar{x}$	$(x_2 - \bar{x})^2$
3	x_3	$x_3 - \bar{x}$	$(x_3 - \bar{x})^2$
4	x_4	$x_4 - \bar{x}$	$(x_4 - \bar{x})^2$
5	x_5	$x_5 - \bar{x}$	$(x_5 - \bar{x})^2$
合計	x_{\cdot}	0	$\sum_{i=1}^5 (x_i - \bar{x})^2$
平均	\bar{x}	0	分散 $\sum_{i=1}^5 (x_i - \bar{x})^2 / 5$

$$\text{標準偏差} = \sqrt{\text{分散}} = \sqrt{\sum_{i=1}^5 (x_i - \bar{x})^2 / 5}$$

分散の大小

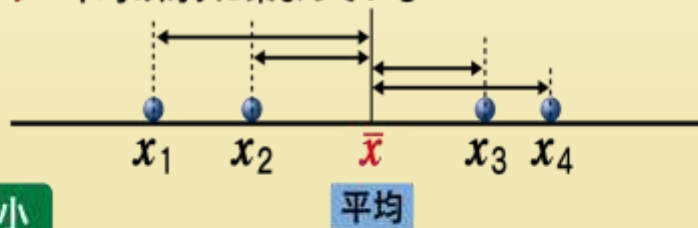
ばらつきが大きいデータ：平均から離れたデータが多い

この長さの2乗が $(x_3 - \bar{x})^2$



分散, 標準偏差が大

ばらつきが小さいデータ：平均の周りに集まっている



分散, 標準偏差が小

分散 ばらつき的大小

各データ値と平均値との距離の2乗の合計をデータ数 n で割った値

$$\begin{aligned}s^2 &= \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n} \\ &= \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2\end{aligned}$$

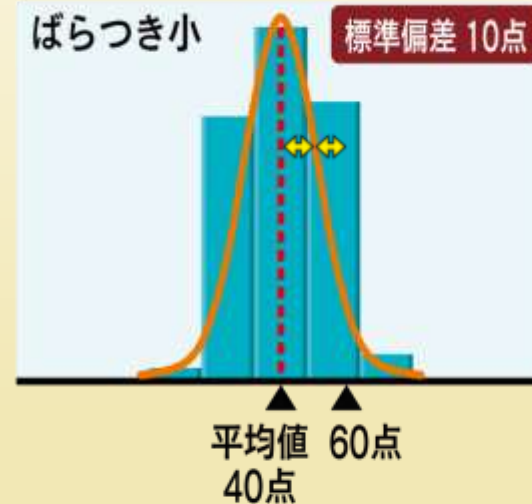
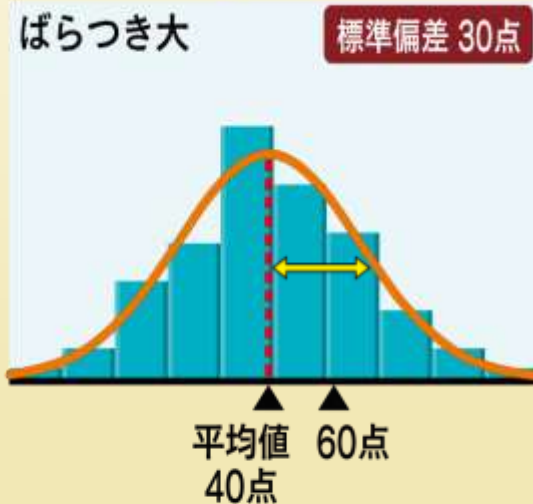
標準偏差 ばらつきの大さの単位

分散の平方根

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

平均からの差(偏差)

数学・・・60点(全体の平均 40点)
 $60\text{点} - 40\text{点} = +20\text{点}$ 大or小?



分散 variance

V , Var , s^2 , σ^2 , **VAR**

標準偏差 standard deviation

S.D., s , シグマ, σ , **STDEV**

標準(化)得点, zスコア

相対的(順位的な)位置を知る

■ 英語 70点

(平均60点 標準偏差10点)

■ 数学 50点

(平均40点 標準偏差 5点)

どちらがよくできた？

標準(化)得点, zスコア

$$z = \frac{\text{データの値} - \text{平均}}{\text{標準偏差}}$$

$$\text{英語} : \frac{70 - 60}{10} = 1 \text{ (点)}$$

$$\text{数学} : \frac{50 - 40}{5} = 2 \text{ (点)}$$

偏差値

平均 50点

標準偏差 10点

になるように変換したもの



$$\text{標準得点} \times 10 + 50$$

国語のテストの点数85点
(平均値 70点, 標準偏差 10点)

- 平均値より **+15** 点高い
 - ・ 平均より標準偏差の **+1.5** 倍離れている
- 国語の標準得点は **+1.5** 点である
- 国語の偏差値は **65** 点
(平均値50点, 標準偏差10点)

分散 variance

V , Var , s^2 , σ^2 , **VAR**

標準偏差 standard deviation

S.D., s , シグマ, σ , **STDEV**

標準(化)得点, zスコア

相対的(順位的な)位置を知る

■ 英語 70点

(平均60点 標準偏差10点)

■ 数学 50点

(平均40点 標準偏差 5点)

どちらがよくできた？

標準(化)得点, zスコア

$$z = \frac{\text{データの値} - \text{平均}}{\text{標準偏差}}$$

$$\text{英語} : \frac{70 - 60}{10} = 1 (\text{点})$$

$$\text{数学} : \frac{50 - 40}{5} = 2 (\text{点})$$

偏差値

平均 50点

標準偏差 10点

になるように変換したもの



$$\text{標準得点} \times 10 + 50$$

国語のテストの点数85点
(平均値 70点, 標準偏差 10点)

■ 平均値より **+15** 点高い

・ 平均より標準偏差の **+1.5** 倍離れている

■ 国語の標準得点は **+1.5** 点である

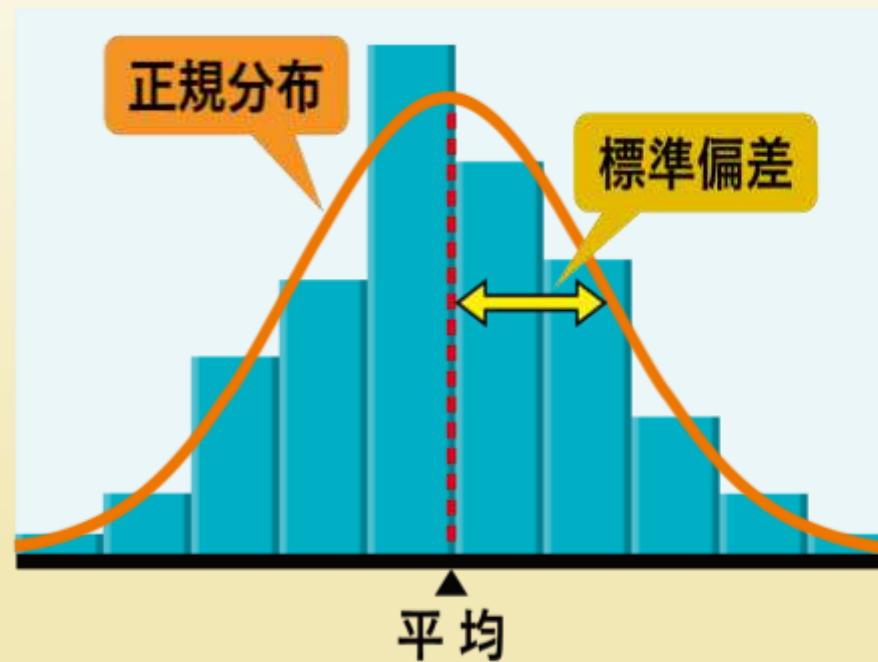
■ 国語の偏差値は **65**点

(平均値50点, 標準偏差10点)

練習問題

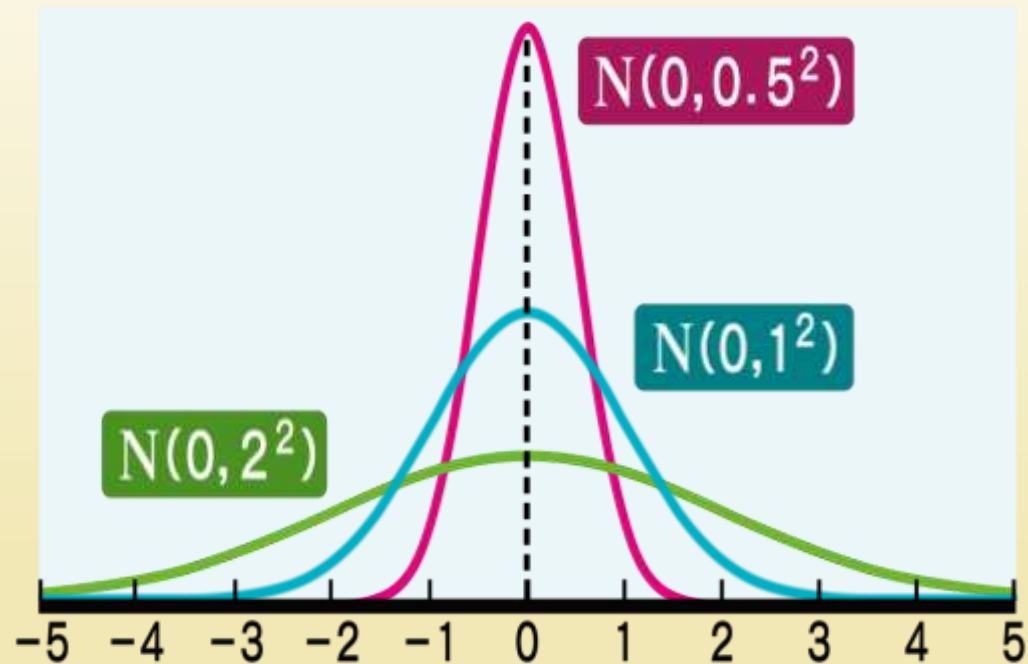
1. 英語の点数が60 平均点が65 標準偏差が20のとき
標準得点(zスコア)と偏差値を求めなさい。
2. 数学の点数が80 平均点が75 標準偏差が10のとき
標準得点(zスコア)と偏差値を求めなさい。

確率分布モデル



正規分布 (Normal Distribution)

いろいろな正規分布



正規分布(*Normal Distribution*)

代表的な確率分布

正規分布(ガウス分布とも呼ぶ)

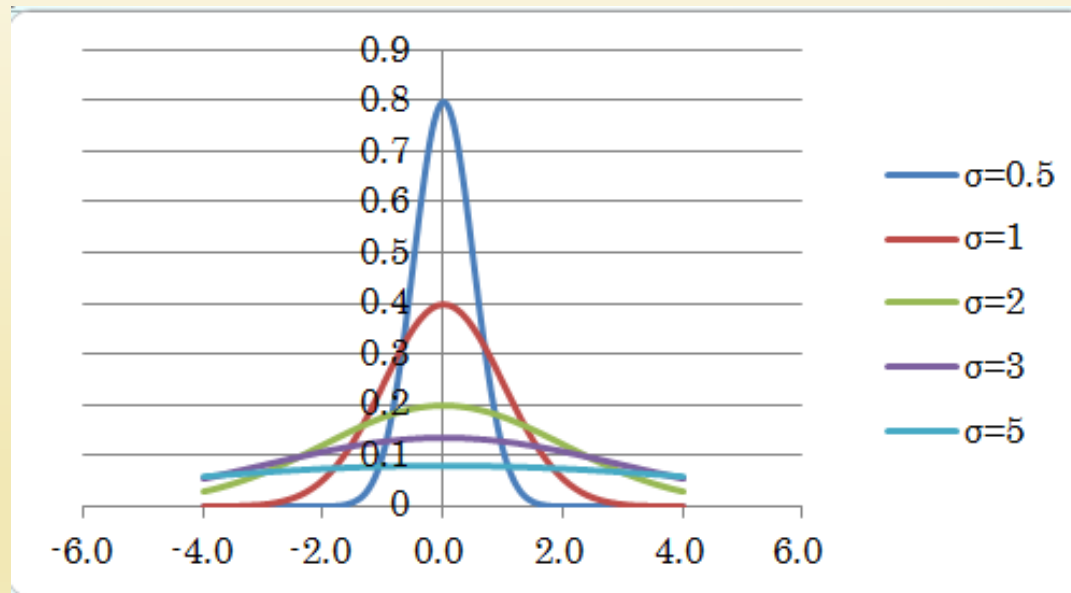
正規分布は、非常によく使われる分布で、平均が μ ,
分散が σ^2 の正規分布を $N(\mu, \sigma^2)$ と書く。

平均: μ 分散: σ^2

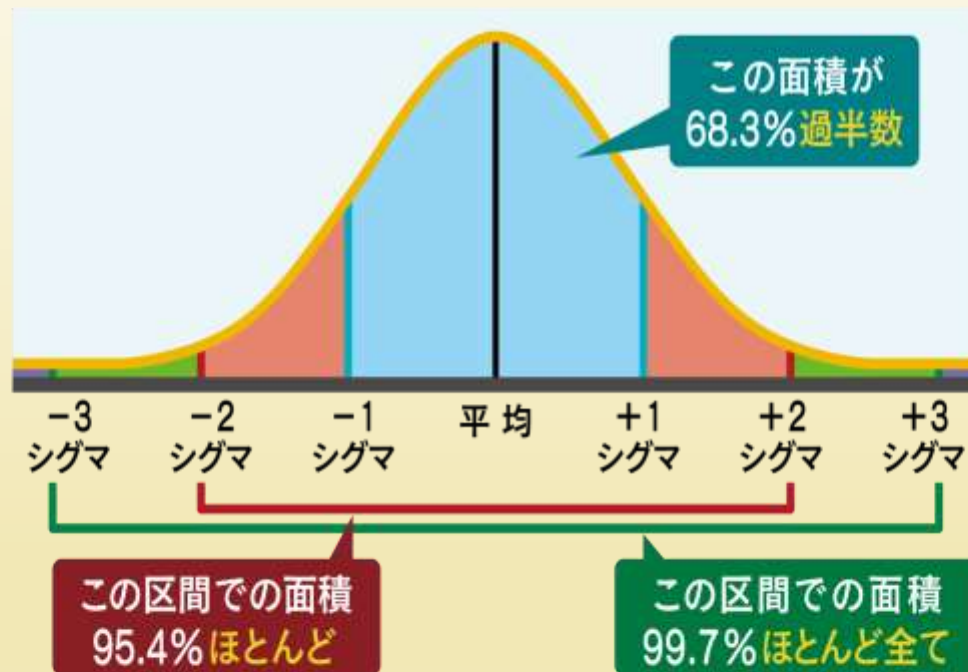
正規分布は, x (グラフ横軸の値) の関数として次式で表される。

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

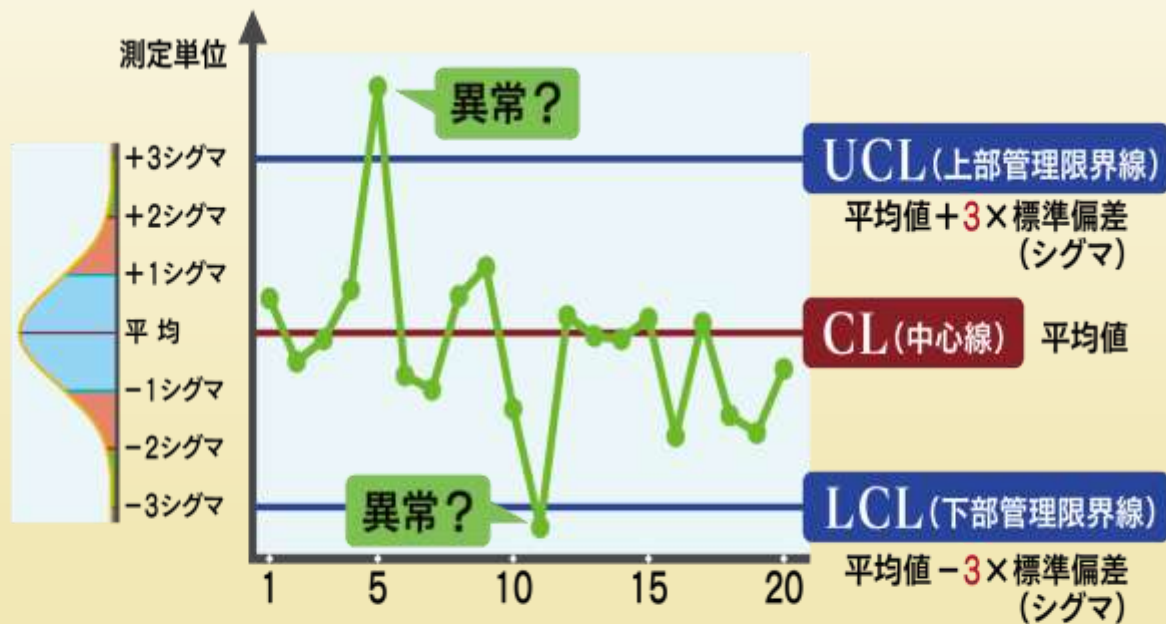
平均が 0, 標準偏差が 1 である
正規分布 $N(0, 1^2)$ を標準正規分布と呼ぶ。



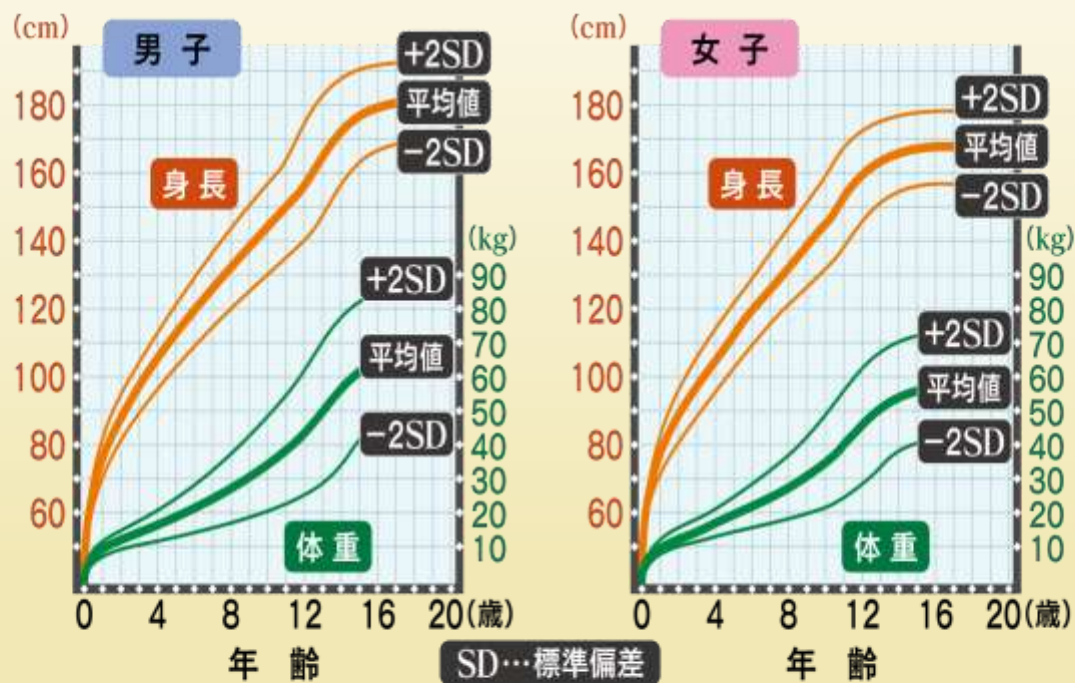
1シグマ 2シグマ 3シグマの法則



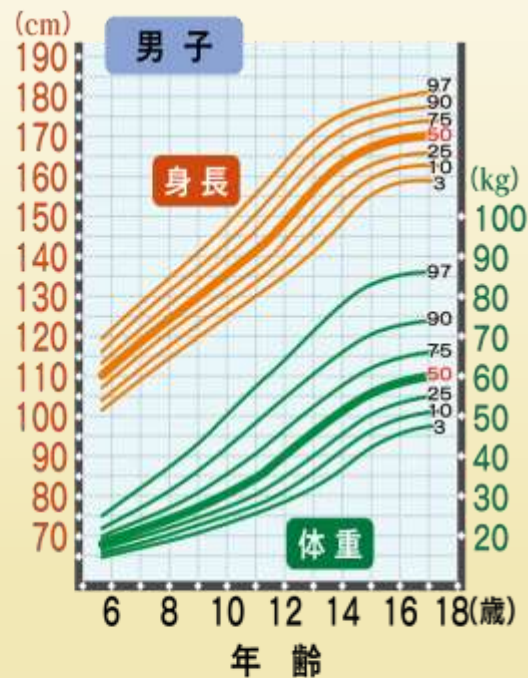
管理図(control chart)



成長曲線 (growth chart)



成長曲線 (growth chart)

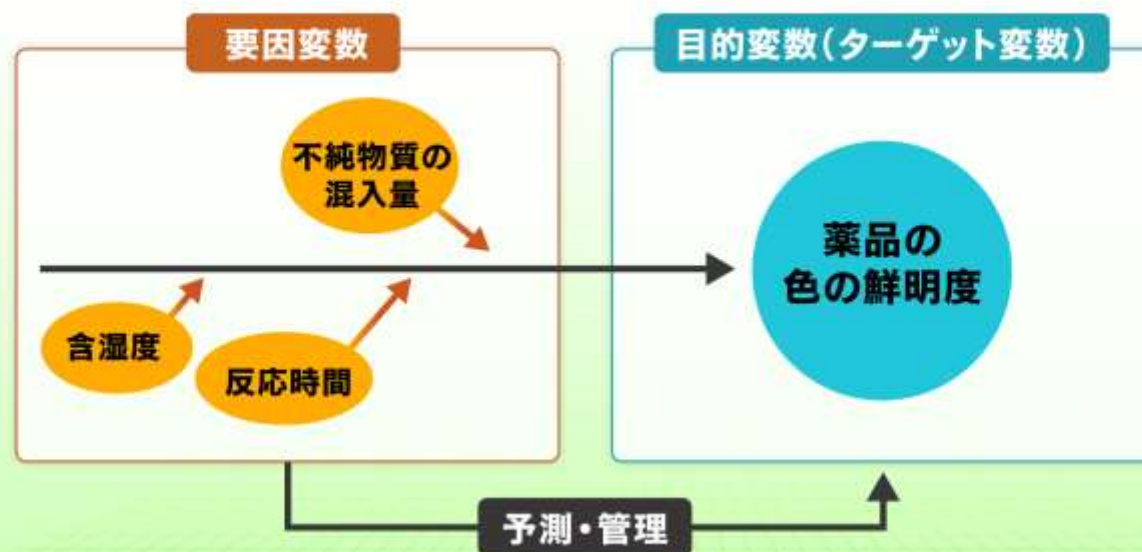


パーセント点

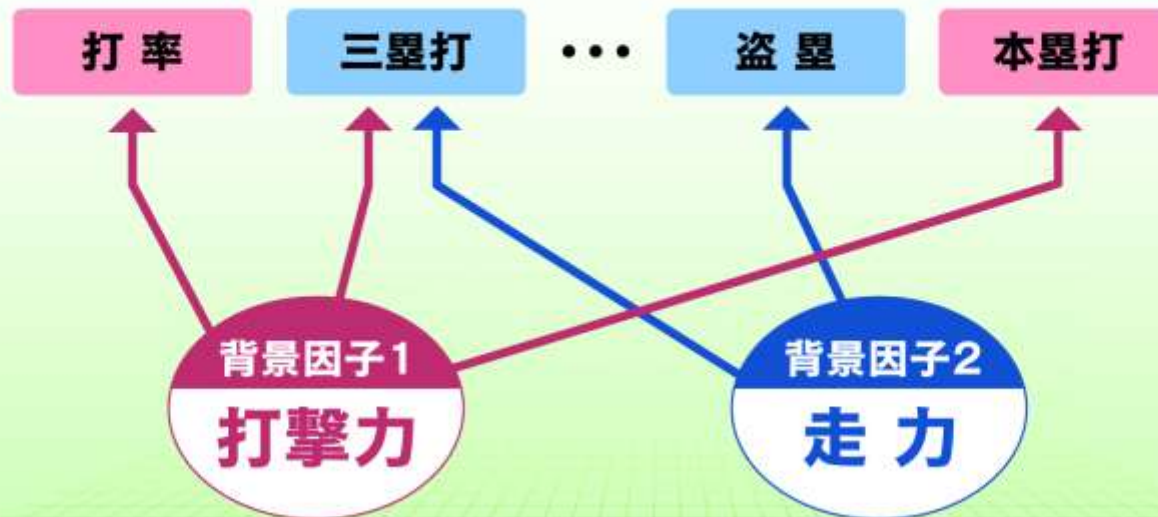
データを大きさの順に
並べたときに、上から(下から)
数えてちょうどある割合($\alpha\%$)に
データが分かれる値

上側 $\alpha\%$ 点
(下側 $\alpha\%$ 点)

変数間の関係を探る(因果関係)



変数間の関係を探る(潜在構造)



散布図 (scatterplot)

三振 VS 本塁打

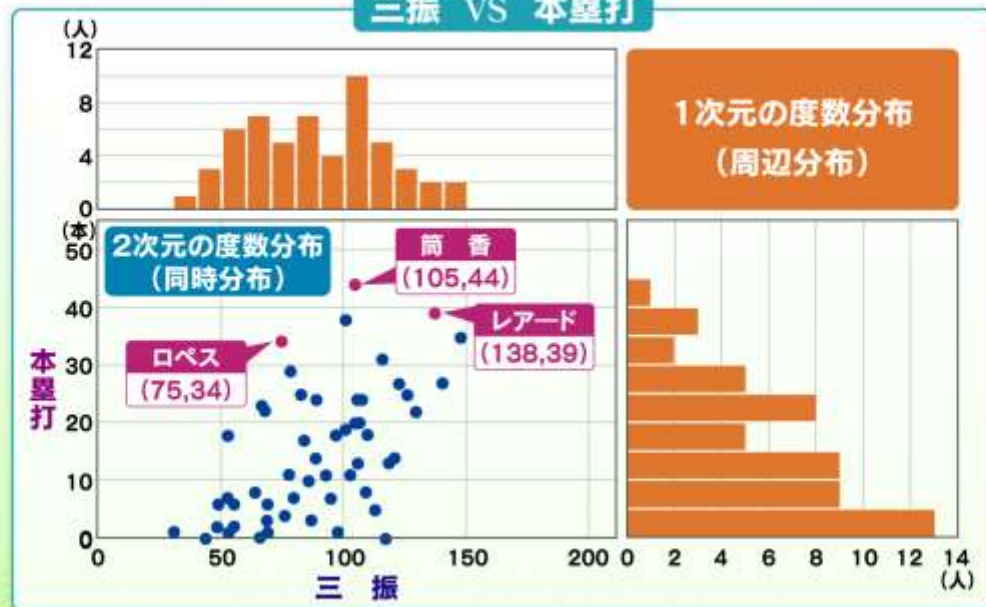


資料:「日本野球機構」公式サイト

2つの変数のばらつきを同時に見るグラフ
観測対象(ケース)のポジションと全体の関係の傾向

散布図 (scatterplot)

三振 VS 本塁打



資料:「日本野球機構」公式サイト

散布図 (scatterplot)

三振 VS 本塁打

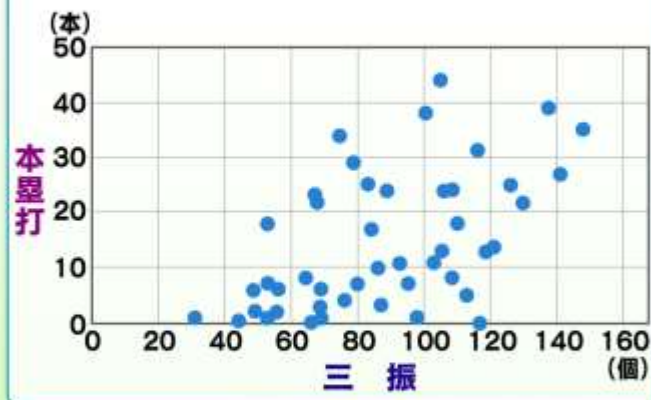


資料:「日本野球機構」公式サイト

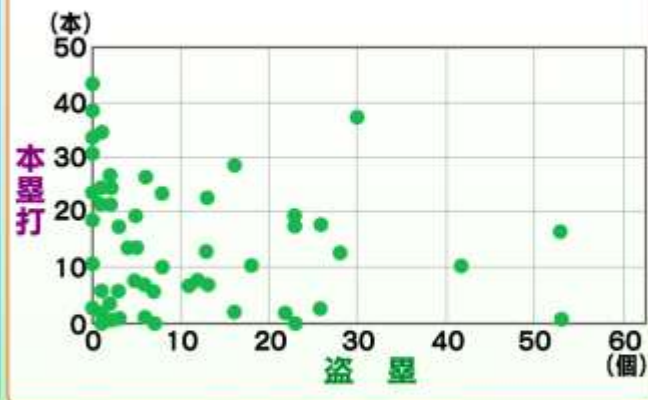
2つの変数間の関係の傾向と
傾向から外れた対象(ケース)の特定

相関関係(correlation)

三振 VS 本塁打



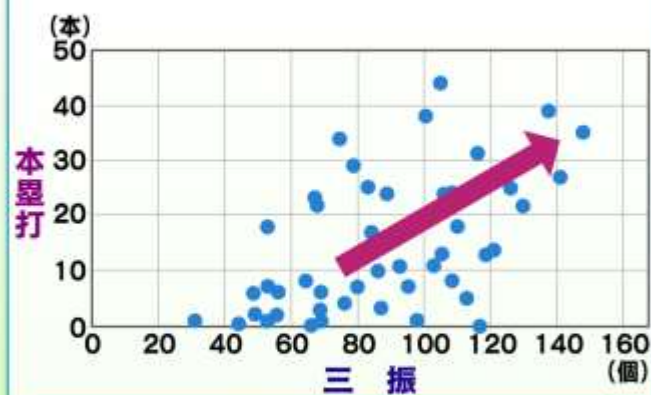
盗塁 VS 本塁打



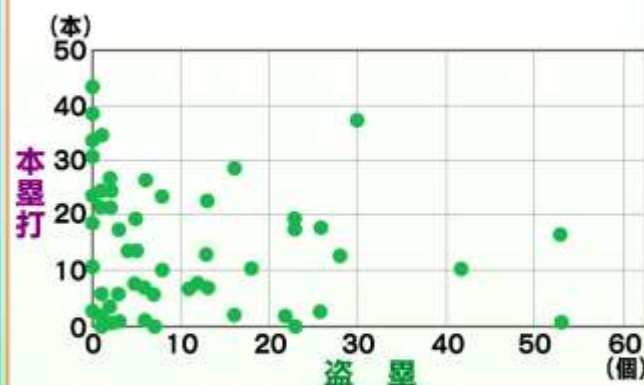
資料:「日本野球機構」公式サイト

相関関係(correlation)

三振 VS 本塁打



盗塁 VS 本塁打

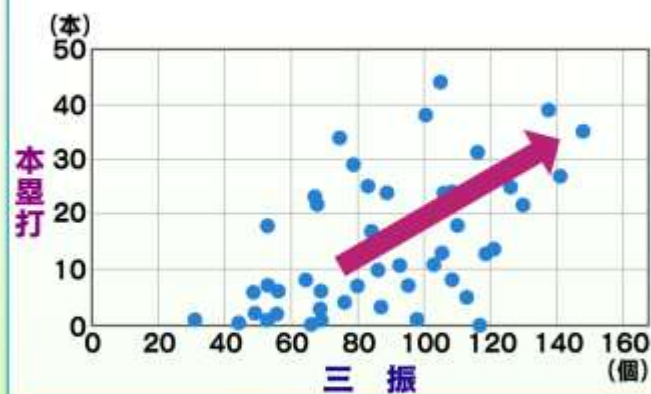


右肩上がりの傾向
正の相関関係

資料:「日本野球機構」公式サイト

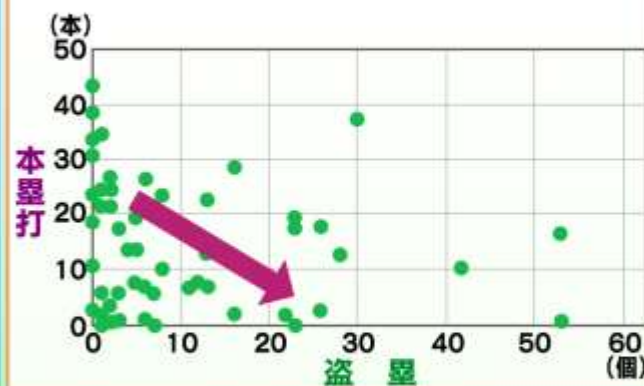
相関関係(correlation)

三振 VS 本塁打



右肩上がりの傾向
正の相関関係

盗塁 VS 本塁打



右肩下がり傾向
負の相関関係

資料:「日本野球機構」公式サイト

関係のいろいろなパターン

直線関係 相関関係

正の相関



負の相関



曲線関係



曲線関係

2次曲線



3次曲線

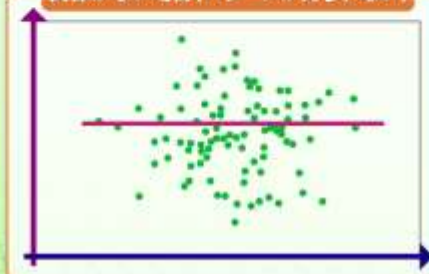


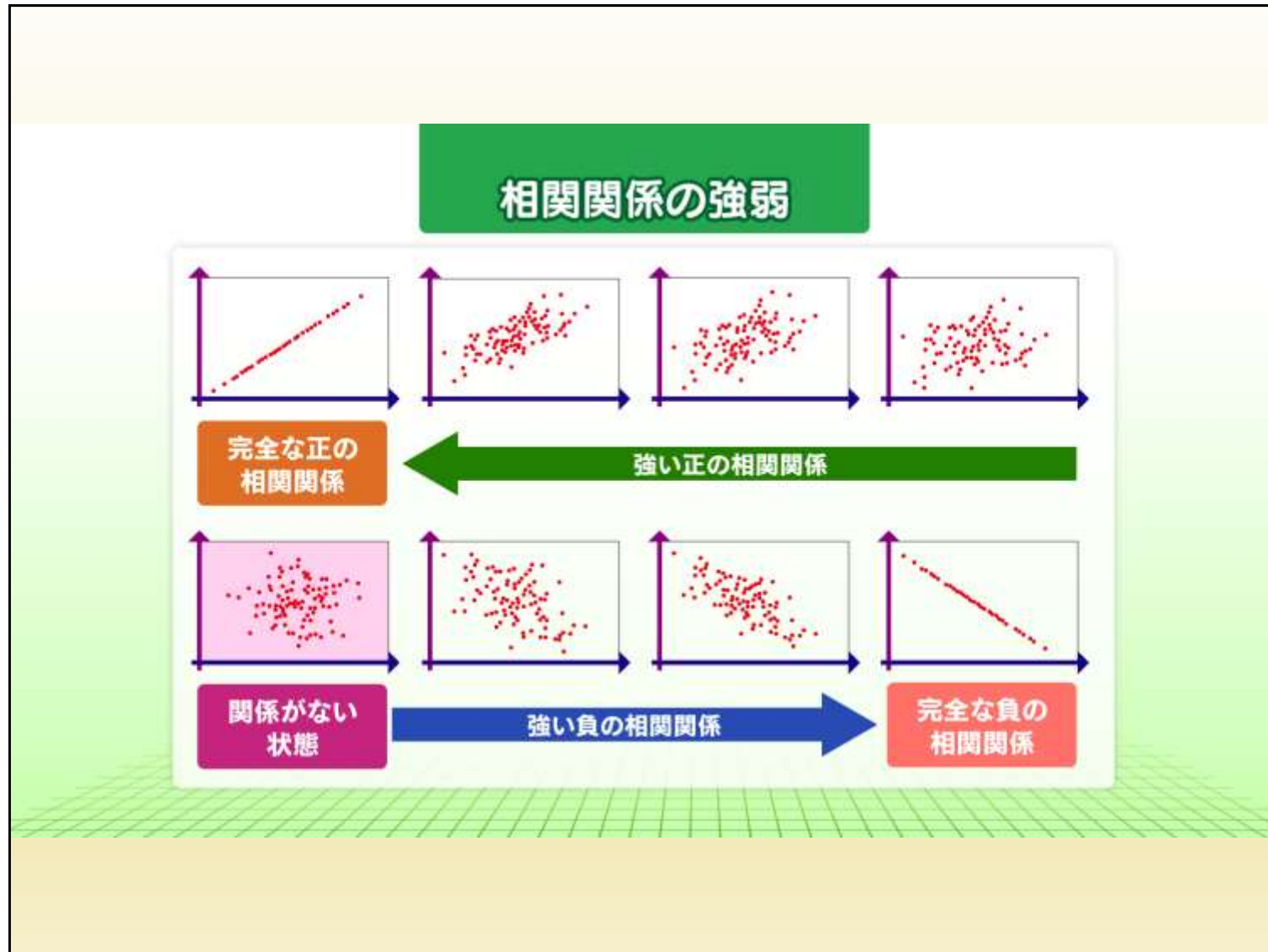
指数曲線



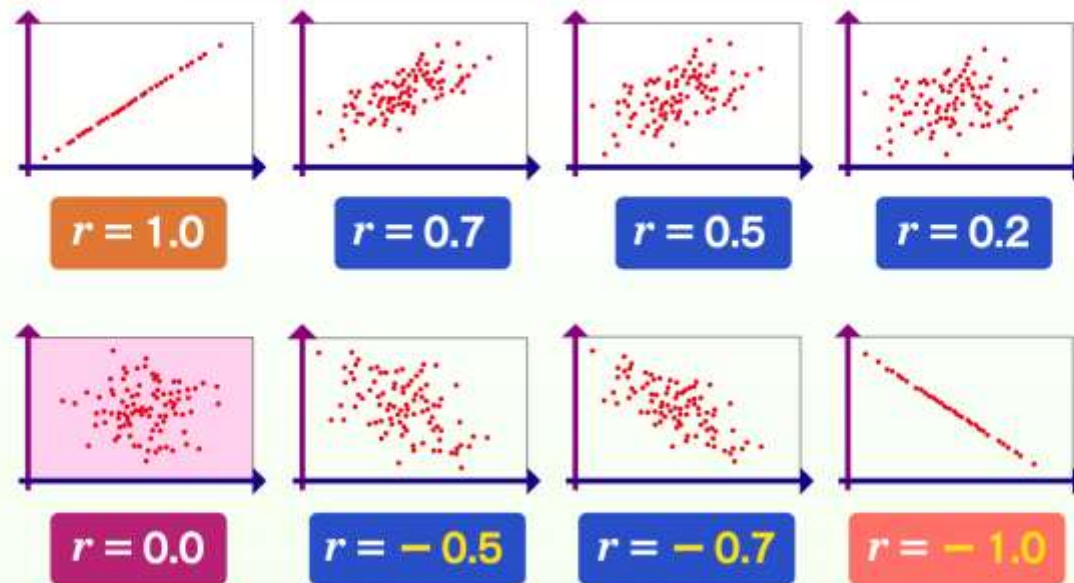
⋮

関係がない場合 (パターンが見られない)



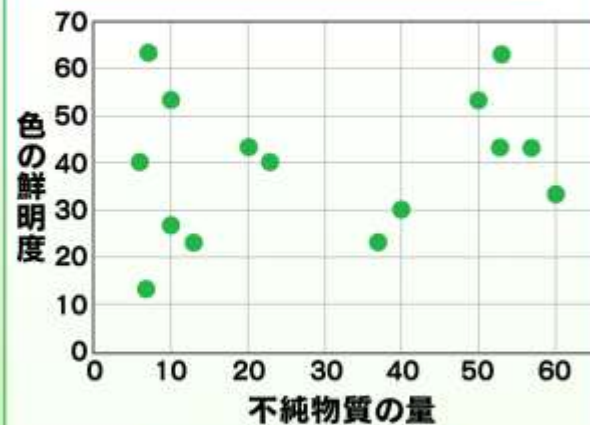


相関係数(相関関係を測る指標)



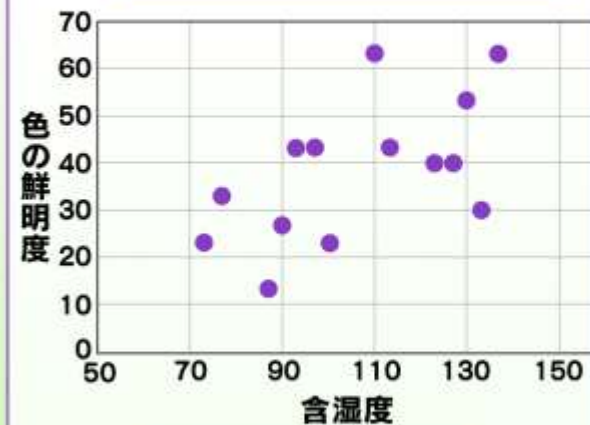
相関係数(相関関係を測る指標)

不純物質の量と鮮明度の散布図



$$r = 0.199$$

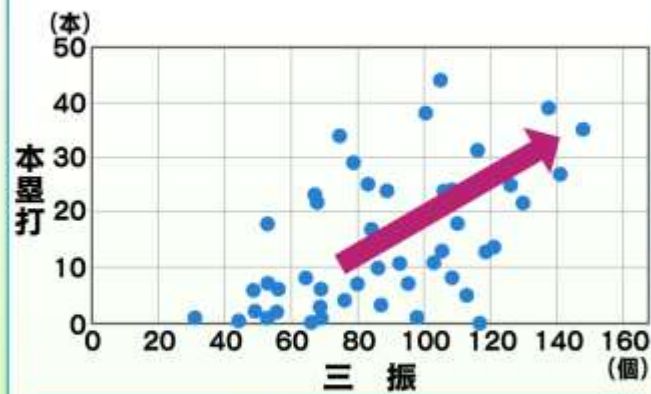
含湿度と色の鮮明度の散布図



$$r = 0.616$$

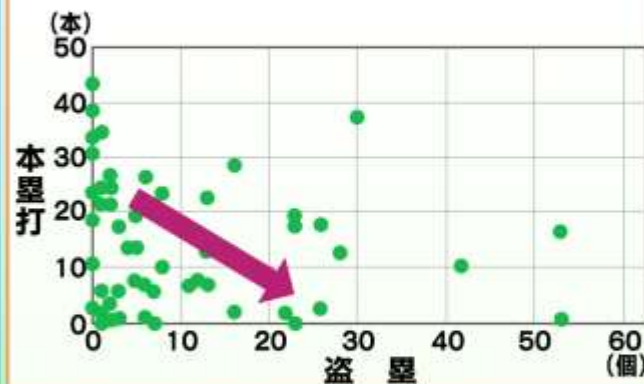
相関係数(相関関係を測る指標)

三振 VS 本塁打



$$r = 0.566$$

盗塁 VS 本塁打



$$r = -0.197$$

資料:「日本野球機構」公式サイト

相関係数の計算

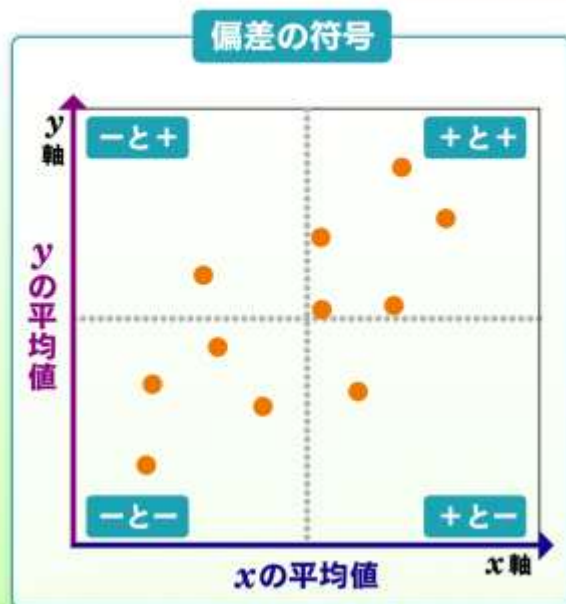
ケース番号	データの値 (x)	データの値 (y)	偏差積
1	x_1	y_1	$(x_1 - \bar{x})(y_1 - \bar{y})$
2	x_2	y_2	$(x_2 - \bar{x})(y_2 - \bar{y})$
3	x_3	y_3	$(x_3 - \bar{x})(y_3 - \bar{y})$
4	x_4	y_4	$(x_4 - \bar{x})(y_4 - \bar{y})$
5	x_5	y_5	$(x_5 - \bar{x})(y_5 - \bar{y})$
合 計	x	y	$\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y})$ 偏差積和
平 均	\bar{x}	\bar{y}	$\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y}) / 5$ 共分散 V_{xy}

$$r_{xy} = \frac{V_{xy}}{s_x s_y} = \frac{\sum_{i=1}^5 (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^5 (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^5 (y_i - \bar{y})^2}}$$

x と y の z スコアの積の平均

s_x, s_y は標準偏差

相関係数の計算



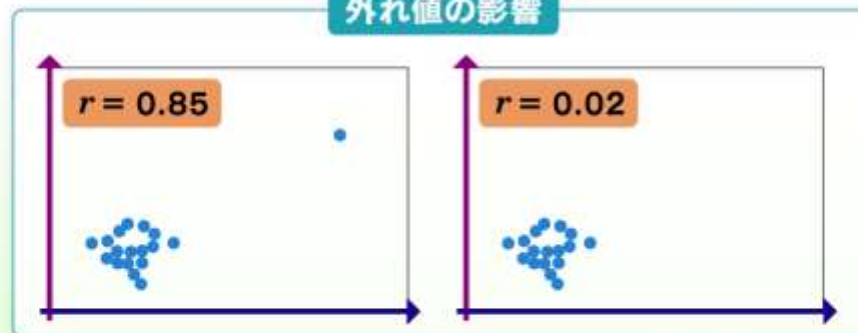
$$r_{xy} = \frac{V_{xy}}{s_x s_y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

xとyのzスコアの積の平均

データの単位やばらつきの大きさに
依存しない指標

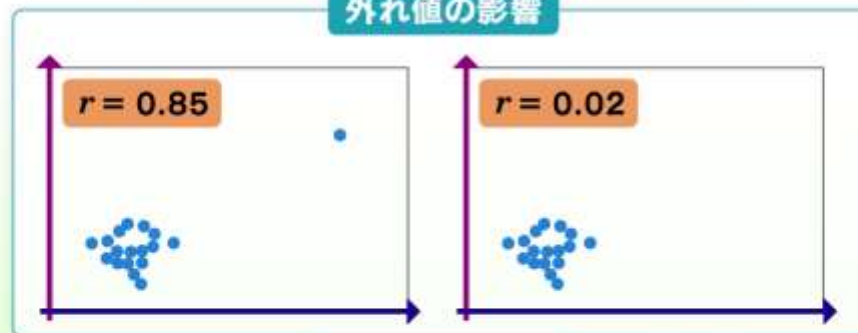
相関係数の注意点

外れ値の影響



相関係数の注意点

外れ値の影響

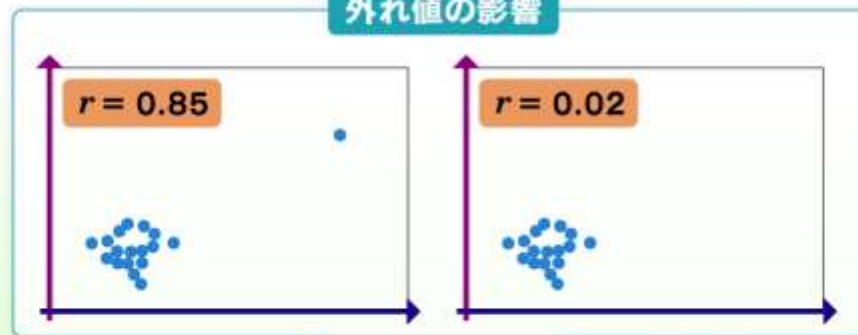


グループの影響



相関係数の注意点

外れ値の影響

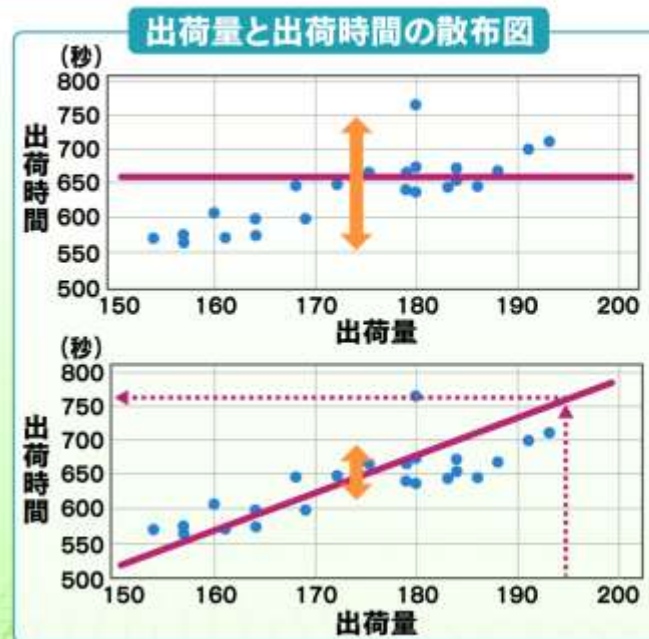


グループの影響

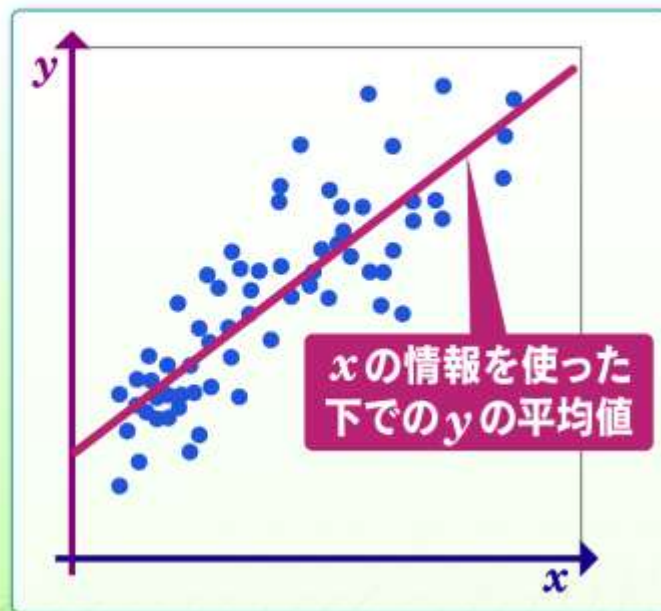


相関関係を利用した予測

出荷量	出荷時間(秒)
157	564
154	569
161	570
157	572
164	573
164	599
169	599
160	605
180	637
179	640
168	645
183	645
186	645



回帰直線



$$y = a + bx + \text{残差}$$

残差平方和(S_E)を
最小にする a と b を求める

$$S_E = \sum_{i=1}^n \{y_i - (a + bx_i)\}^2$$

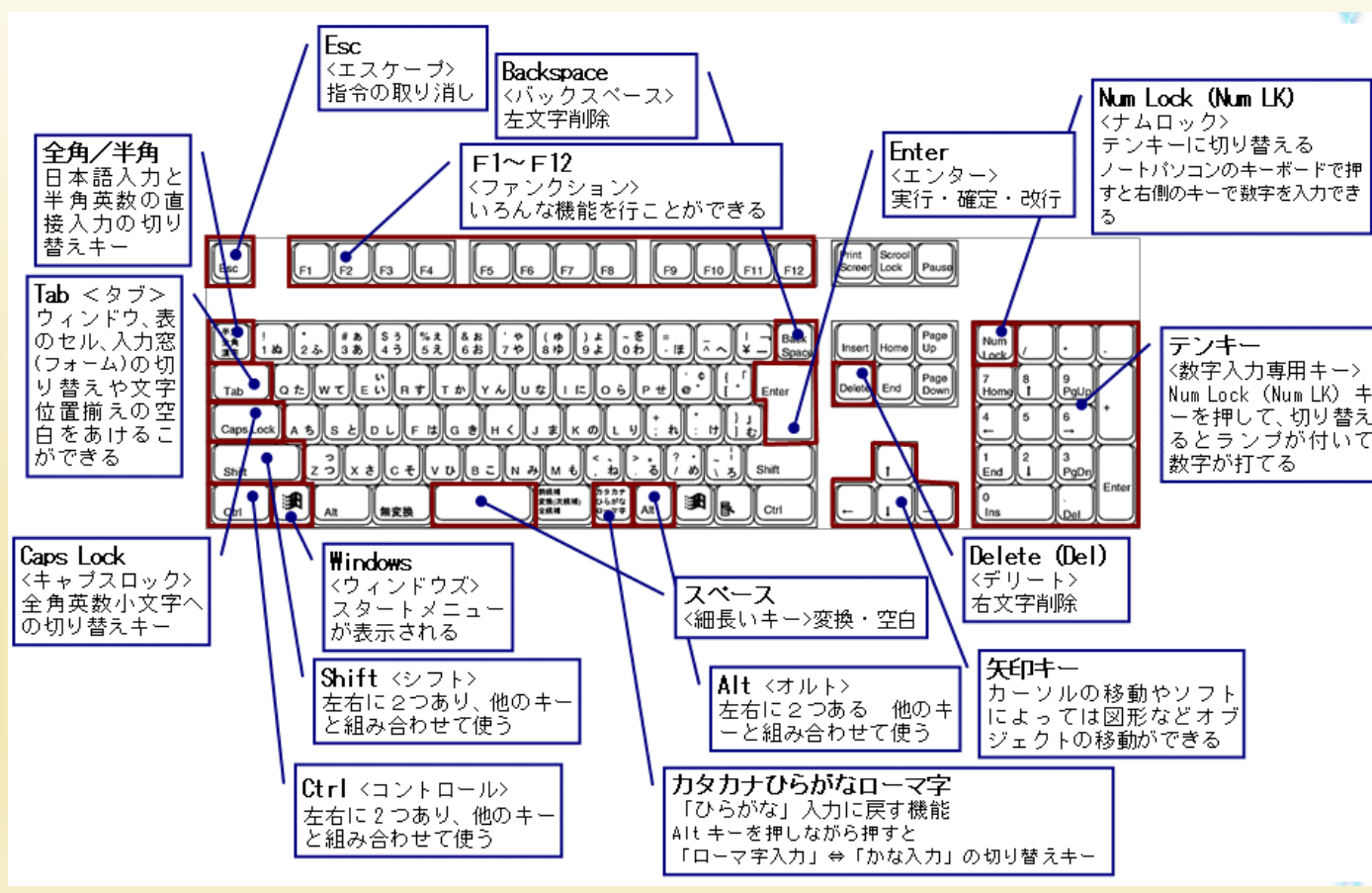


最小2乗法

～PCのキーボード操作資料～

キーボード

1. 記号の読み方と入力方法



キーボード

1. 記号の読み方と入力方法

キー記号		主な読み方	全角入力キー	半角入力キー
!	!	イクスクラメーション・感嘆符	Shift + 1	Shift + 1
”	”	ダブルクォーテーション	Shift + 2	Shift + 2
#	#	いげた・ナンバー・ハッシュマーク	Shift + 3	Shift + 3
\$	\$	ダラス、ドルマーク	Shift + 4	Shift + 4
%	%	パーセント	Shift + 5	Shift + 5
&	&	アンパサンド	Shift + 6	Shift + 6
'	'	アポストロフィー	Shift + 7	Shift + 7
((丸かっこ(開く)	Shift + 8	Shift + 8
))	丸かっこ(閉じる)	Shift + 9	Shift + 9
=	=	イコール	Shift + ほ	Shift + ほ
—	—	マイナス・ハイフン	ほ	ほ
	^	キャロット・カレット・キャレット・ハット		へ
	~	チルダ		Shift + へ
~		波ダッシュ	へ	
¥	¥	円マーク	¥	¥
@	@	単価記号・アットマーク	@	@

キーボード

1. 記号の読み方と入力方法

キー記号		主な読み方	全角入力キー	半角入力キー
	[角かっこ（左）（開く）		{ [「
]	角かっこ（右）（閉じる）		}] 」
+	+	プラス	Shift + れ	Shift + れ
*	*	アスタリスク	Shift + け	Shift + け
;	;	セミコロン	れ	れ
:	:	コロン	け	け
	,	カンマ		ね
	.	ピリオド（ドットとも言います）		る
,		読点	ね	
。		句点	る	
<	<	小なり・左アングルブラケット	Shift + ね	Shift + ね
>	>	大なり・右アングルブラケット	Shift + る	Shift + る
?	?	疑問符・クエスチョンマーク	Shift + め	Shift + め
/	/	スラッシュ		め
\		バックスラッシュ ※ ¹		
_	-	アンダーバー・アンダースコア	Shift + ろ	Shift + ろ

※¹ バックスラッシュの半角記号はWindowsでは通常入力できない。

キーボードから直接入力できない記号は記号名を入力して変換することもできる。

例えば、\ は「ななめ」と入力して変換する。

キーボード

2. ファンクションキー（F1～F12）

Fキー	機 能
F1	起動しているソフトの「ヘルプ機能」を呼び出すキー
	ワード2003 & 2007の場合 右画面に「Wordのヘルプ」の「アシスタント」画面が表示される
F2	ファイルやフォルダを選択時には、「名前の変更」することができる
F4	ワードやエクセルなどでは、直前の操作を繰り返す
F5	ウィンドウの更新を行なう
	ホームページ閲覧ソフトなどでは、そのページの最新の情報に更新される
F6	入力モードで、ひらがなに変換
F7	入力モードで、カタカナに変換
F8	入力モードで、半角カタカナに変換
F9	入力モードで、全角英数に変換
F10	入力モードで、半角英数に変換
F11	ホームページ閲覧ソフトでは、全画面表示になる
F12	ワードやエクセルなどで保存操作をする

キーボード

3. ショートカットキーによる入力方法

キーの組み合わせ	機 能
Ctrl + A	全選択
Ctrl + C	コピー
Ctrl + X	切り取り
Ctrl + V	貼り付け
Ctrl + Z	元に戻す
Ctrl + B	太字
Ctrl + U	下線
Ctrl + I	斜体