



データサイエンス基礎

第1回 担当：長谷川 成明

自己紹介



長谷川 成明

(はせがわ しげあき)

shasegaw@sonoda-u.ac.jp

経営学部ビジネス学科

- 数理生態学
- 植物生態学
- 生態系モデリング

はじめに

データ
サイエンス
って、なに？



はじめに

データサイエンスとは
社会に溢れているデータから
《価値》を引き出す学問

滋賀大学 データサイエンス学部

<https://www.ds.shiga-u.ac.jp/about/ds/introduction/>



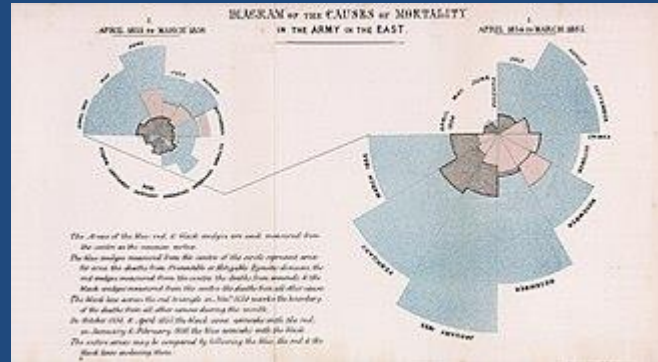
統計的手法や
データ分析
人工知能などが
用いられる

<https://www.tableau.com/ja-jp/learn/articles/what-is-Data-Science>

データをどう活用するか

フローレンス・ナイチンゲール

「近代看護教育の母」

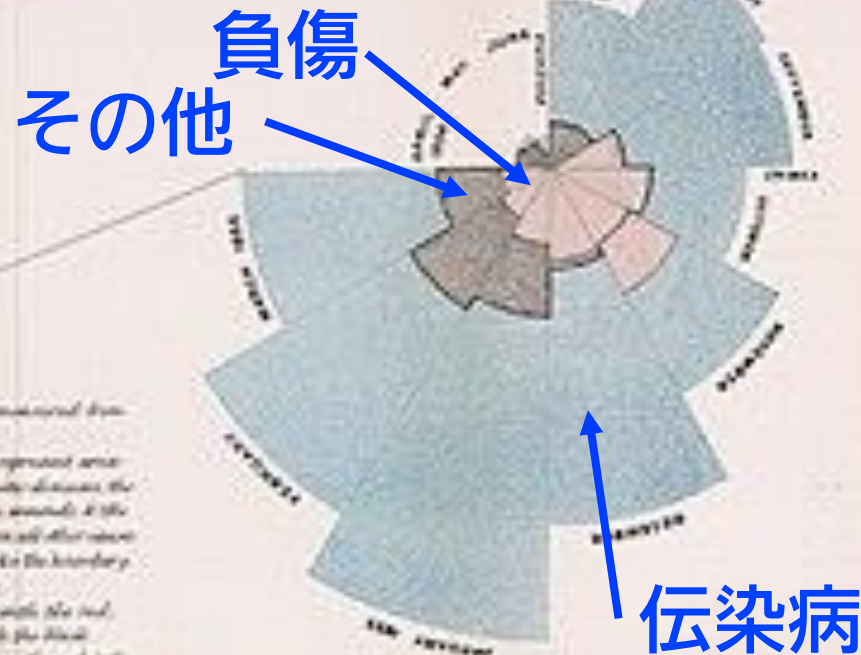


医療統計学を駆使した
データサイエンティスト

※詳しくは総務省統計局図書館ウェブページ


<https://www.stat.go.jp/library/pdf/minigougai02.pdf>

「鶏のトサカ」死亡要因の分類



<https://www.stat.go.jp/library/pdf/minigougai02.pdf>

コンビニ

- 
- 品目
 - 個数
 - 一緒に買ったもの
 - 買った人

コンビニ本部には
詳細データが集まっている

改札機

- いつ
- どこから来て
- どこを通ったか



<https://www.jreast.co.jp/suica/howtoget/>

世界はデータで溢れている



これまで
無かった
巨大データ

ビッグデータ

ビッグデータをどう扱うのか?

マネーボール



アメリカ メジャーリーグの
弱小球団
オークランド・アスレチックス
が統計的手法
「セイバーメトリクス」

アナリスト
ピーター・ブランド



<https://www.cinemacafe.net/article/2012/02/02/12087.html>

データをどう活用するか

大量のデータが手に入る時代
どのように活用するかが重要

経済産業省委託事業

平成30年度我が国におけるデータ駆動型社会に係る基盤整備
(IT人材等育成支援のための調査分析事業)

ー IT人材需給に関する調査 ー

調査報告書

2019年3月

みずほ情報総研株式会社

2030年に
データサイエンティスト
(AI関連)が
国内で12万4千人不足

データをどう活用するか

大量のデータが手に入る時代
どのように活用するかが重要

ビッグ
データ
の世界



ビッグ
データ
の世界



この授業!!



基礎的な
統計学



教科書



統計学の
やさしい授業
みみたとサブローの学習ノート

三井 正 著,
原山 みりん 作画,
オーム社

授業計画

- 第1回 統計とは何か
- 第2回 代表値
- 第3回 データのばらつき
- 第4回 確率分布
- 第5回 二項分布
- 第6回 正規分布
- 第7回 中心極限定理



授業計画

- 第 8回 母集団の推定(1)
- 第 9回 母集団の推定(2)
- 第10回 母集団の推定(3)
- 第11回 仮説検定(1)
- 第12回 仮説検定(2)
- 第13回 母平均の検定
- 第14回 独立性の検定
- 第15回 授業まとめと確認



授業の進め方

成績評価

平常点	20%
授業内演習	20%
授業課題	30%
試験(筆記)	30%



平常点

リアクションペーパーにより評価

講義を受講して
理解した内容をまとめて
manabaを通じて提出する



平常点	20%
授業内演習	20%
授業課題	30%
試験(筆記)	30%

平常点

リアクションペーパーの評価ポイント

- 授業内容の理解度
- 内容の正確性

※感想は一切不要

文字数: 200字以上
授業当日 17:00×切



平常点	20%
授業内演習	20%
授業課題	30%
試験(筆記)	30%

授業内演習

授業中に実際に解く演習を行う
manabaを通じて提出すること

演習ではMicrosoft Excelを
使うことがある



平常点	20%
授業内演習	20%
授業課題	30%
試験(筆記)	30%

授業課題

授業で学んだことを中心に
授業課題を出します
manabaを通じて提出すること

×切：原則として授業週の
金曜日 17:00



平常点	20%
授業内演習	20%
授業課題	30%
試験(筆記)	30%

授業課題

課題の提出は

- manabaから電子的に
- 紙媒体で長谷川研究室ポストに
(5号館5階508室)

紙媒体の場合のみ
採点コメントを
付けて返却します

(電子提出は得点のみ返します)



授業課題

課題は基本的にWordで出します

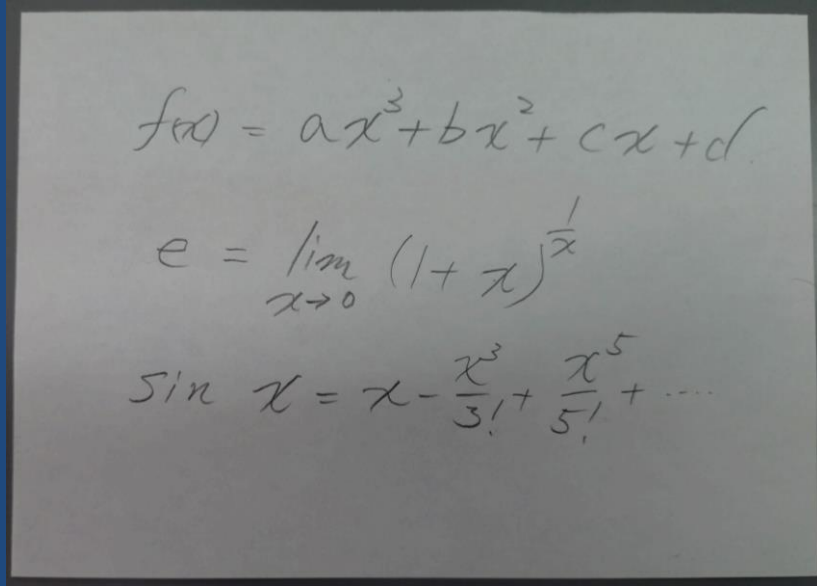
電子的に提出する場合は**写真不可**

Adobe Scanなど
文書スキャンアプリ
で取り込んだものを
提出のこと



<https://www.adobe.com/jp/acrobat/mobile/scanner-app.html>

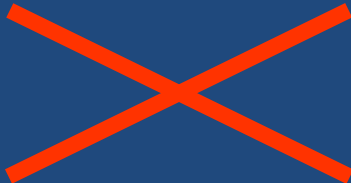
授業課題



Handwritten mathematical formulas on a piece of paper:

$$f(x) = ax^3 + bx^2 + cx + d$$
$$e = \lim_{x \rightarrow 0} (1+x)^{\frac{1}{x}}$$
$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$$

写メ

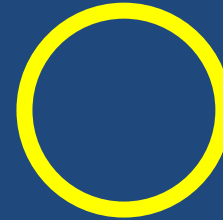


$$f(x) = ax^3 + bx^2 + cx + d$$

$$e = \lim_{x \rightarrow 0} (1+x)^{\frac{1}{x}}$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \dots$$

Adobe Scan



試験(筆記)

第15回の授業冒頭に実施します



平常点	20%
授業内演習	20%
授業課題	30%
試験(筆記)	30%

授業の進め方

全ての提出物について

締切厳守

締切を超過した提出物はゼロ点評価



授業の進め方

提出物に盗用、コピーペストが
見られた場合は
総合評価をゼロ点(単位不可)とします

書き写しが疑われる場合
書き写した方も、書き写させた方も
総合評価をゼロ点(単位不可)とします

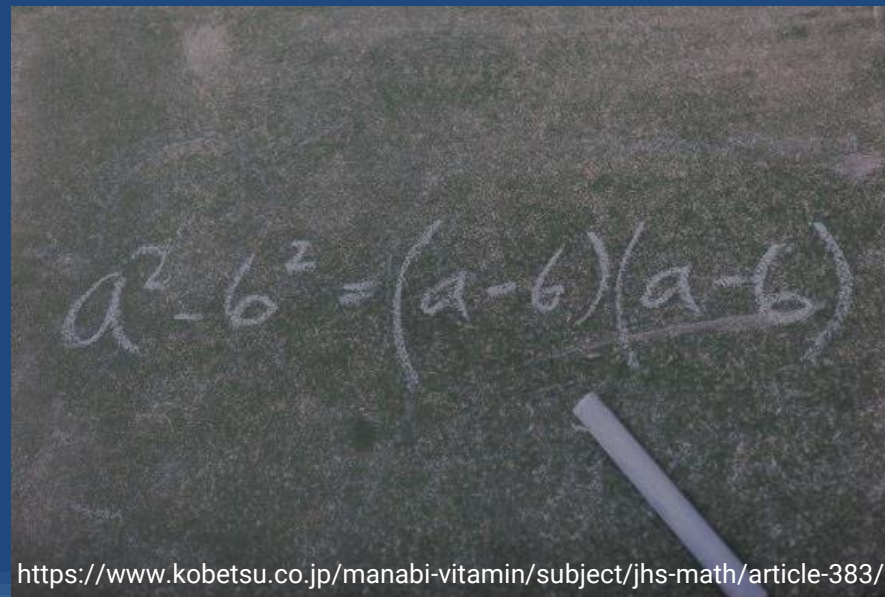
Take home notes

- 授業の進め方について確認しました

おしながき

統計とは何か

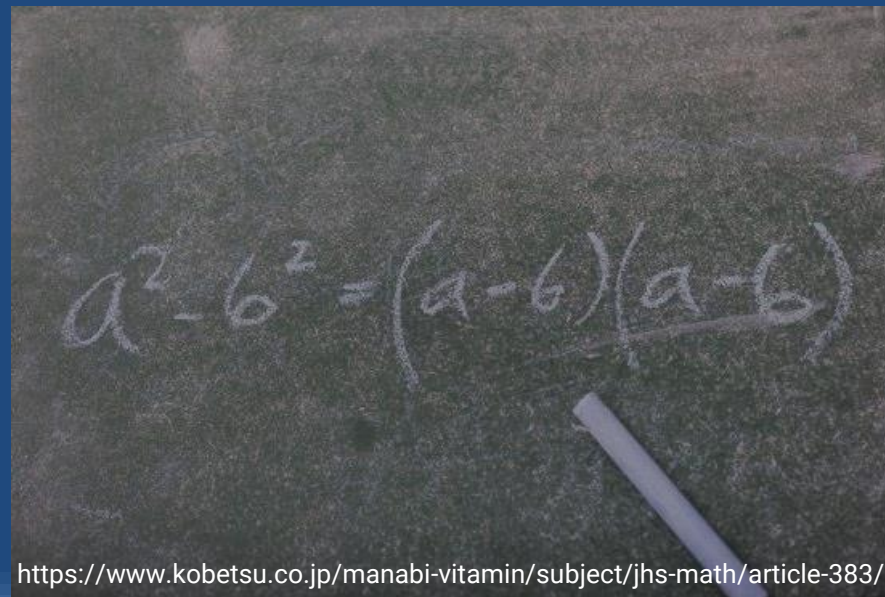
- 統計のしくみ
- 統計と確率
- 統計の役目



おしながき

統計とは何か

- 統計のしくみ
- 統計と確率
- 統計の役目



統計のしくみ



統計のしくみ

社長のいつもの気まぐれか
突然統計係に任命されたんだ

なんスか
その今書いた
ような辞令！

我が社も
AIを導入しようと
思ってるね

まんじゅうも
AIの時代だよキミ

和菓子屋に
AIっスか？

抵抗したけど…

無理っス！
僕デザイナー
っスよ！

フッフ…
君の机に
『イカでもわかるAI』
って本があるの
知ってるんだよ

あれは
画像ソフト
っス！

※イラレっスよ!!

あまり私を
ナメない
ことだ…

社長、
華麗に
スルー！

あっ
会議の時間だ
それじゃ
頼んだからね！

えっ
あっ
待って
ください！

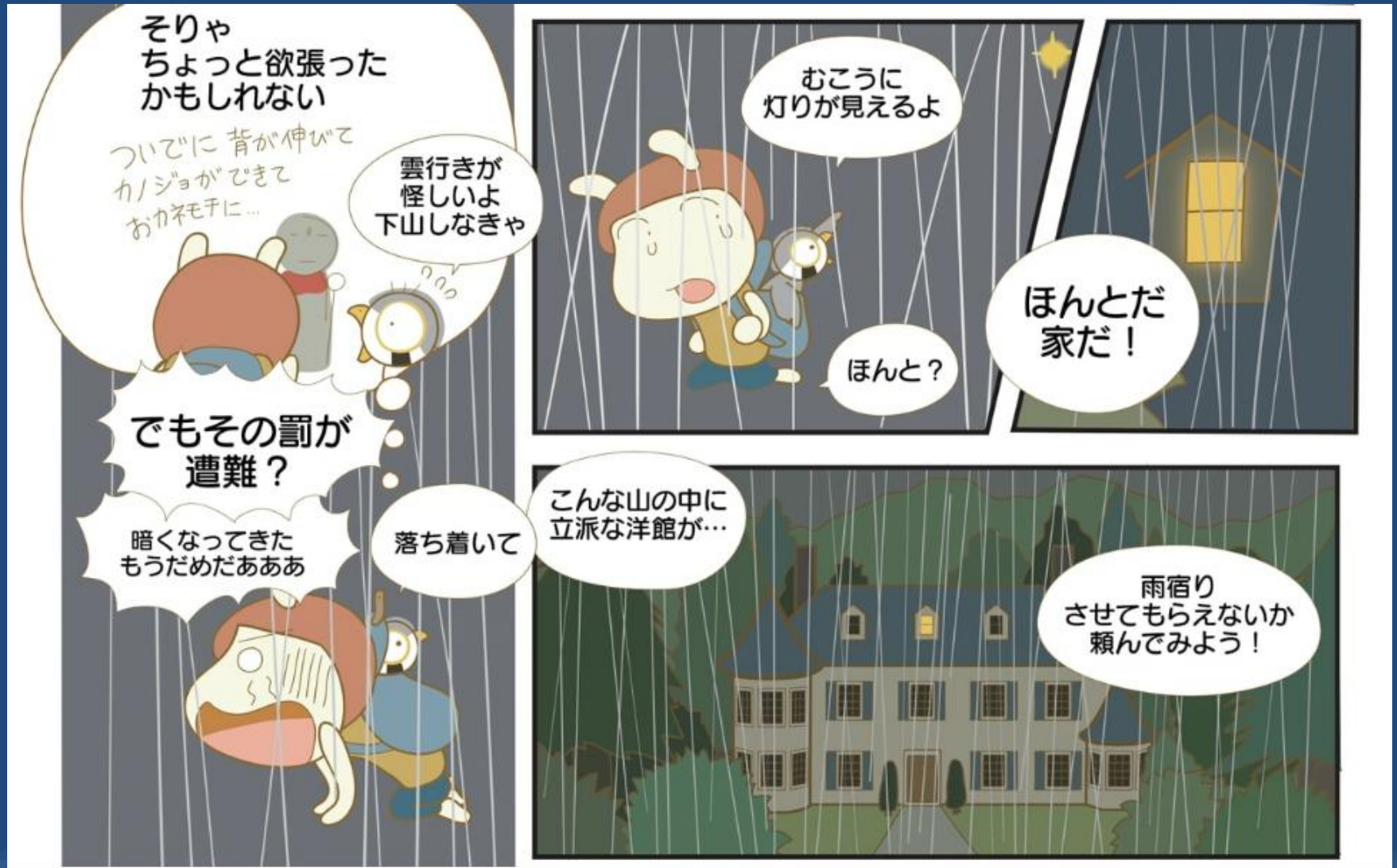
辞令は
下った！

※ Adobe
Illustrator
II
AI

統計のしくみ



統計のしくみ



統計のしくみ



統計のしくみ



統計のしくみ

登場人物紹介



みみた
老舗の和菓子屋・山田屋に
勤めるデザイナー。
趣味: ゲーム
好物: 甘いもの



サブロー
ハヤブサの苦学生。
みみたの家に居候している。
趣味: 彫刻・木工
好物: 焼き鳥



社長
山田屋の社長。
割とお気楽な性格。
趣味: 温泉
好物: 和菓子



先生
山の中の洋館に住んでいる。
不思議な縁によりみみた達に
統計の手ほどきをする。ことに。
趣味: お菓子作り、家庭菜園
好物: 紅茶

統計は何をするものか



統計の概念図

統計は何をするものか

データ: なんらかの対象を観察して
得られた数値や文字といった
観測値の集まり



データから情報を引き出し
それをもとに
正しく判断するための
道具が統計

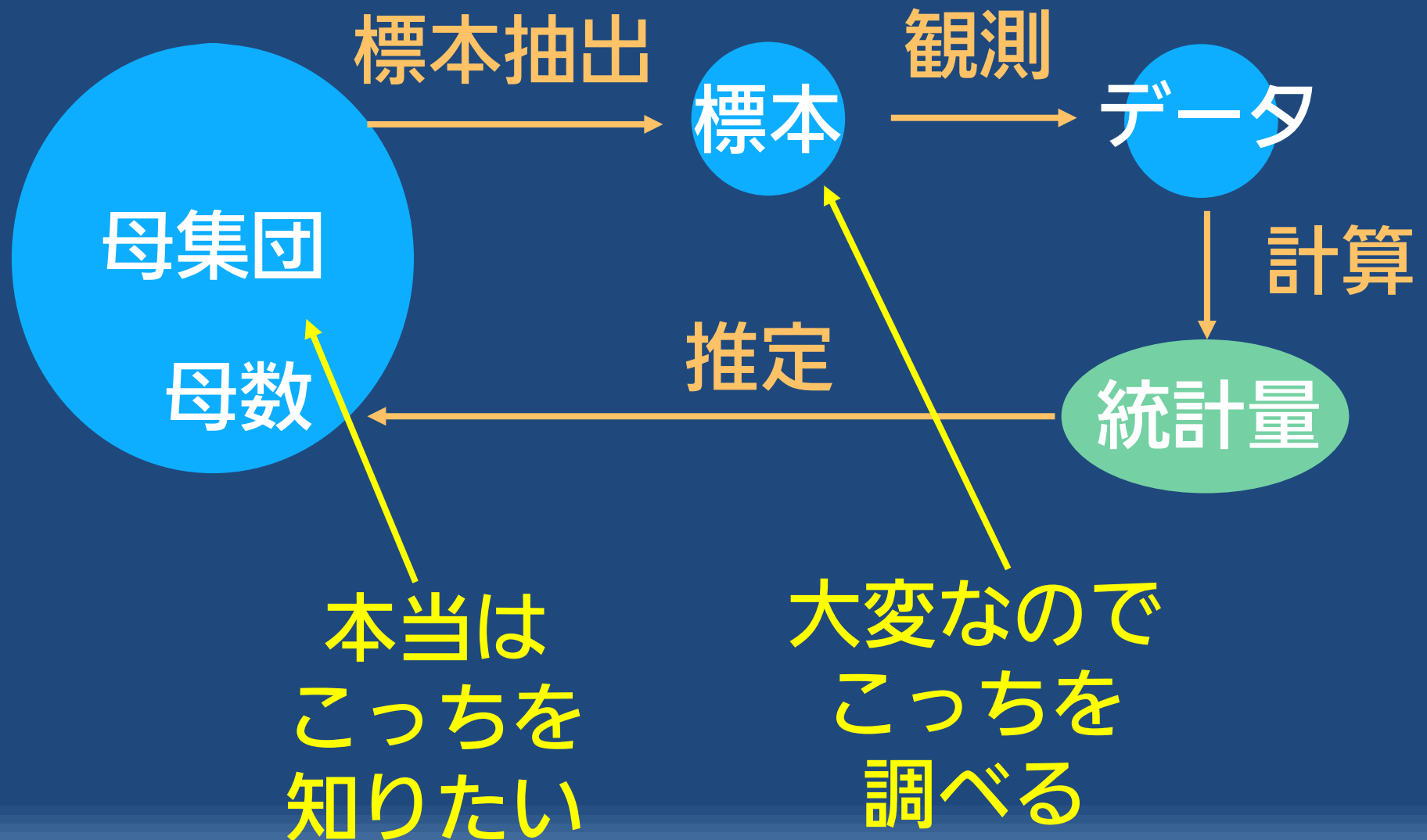
データと統計量

統計量: データから導き出すことのできる情報



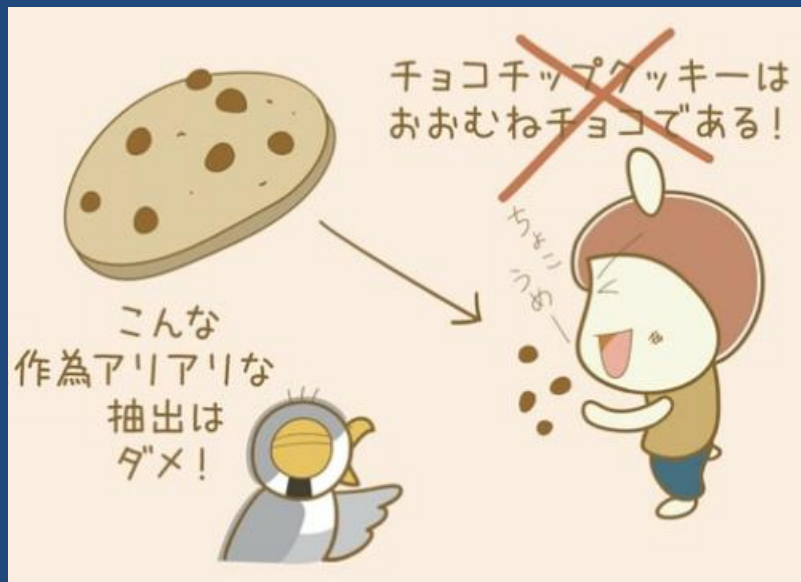
「何を知りたいか」
によって
使う
統計量は
異なる

推定



正しい推定的前提

標本が母集団から
偏りなく抽出されていること



無作為抽出
(ランダム
サンプリング)が
よく用いられる

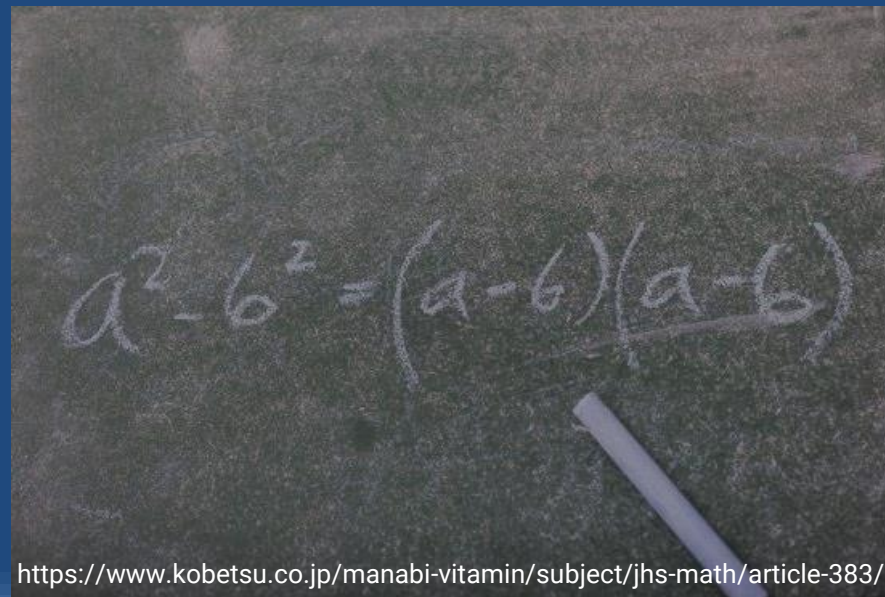
Take home notes

- 統計はデータから情報を引き出し正しく判断するための道具です
- 対象全体を調査できないときに母集団から標本を抽出して得られたデータから統計量を得ます
- 統計量をもとに母集団の母数を推定します

おしながき

統計とは何か

- 統計のしくみ
- 統計と確率
- 統計の役目



確率

私たちの身の回りには
結果を予見することができず
確定的ではない
ものごとが存在する



確率

ある事象がどの程度
「起きやすいか」を示す数値を
確率とよぶ

確率は通常
0から1の間の
値をとる



2種類の確率

確率は2種類に分けることができる

- 数学的確率
- 経験的確率



試行



https://web.gekisaka.jp/news/photo?news_id=275565

コイントスのように
一定の条件のもと
複数回繰り返して
行うことができ
結果が確定的でない
実験や観察を
試行とよぶ





試行

さまざまな試行：



コイン
トス



トランプの
カードを
引く



じゃんけん

事象

試行の結果、生じる現象を**事象**とよぶ



コイン
トス



表が
出る

全事象

試行の事象をすべて集めた集合を
全事象とよぶ



コイン
トス



表が
出る



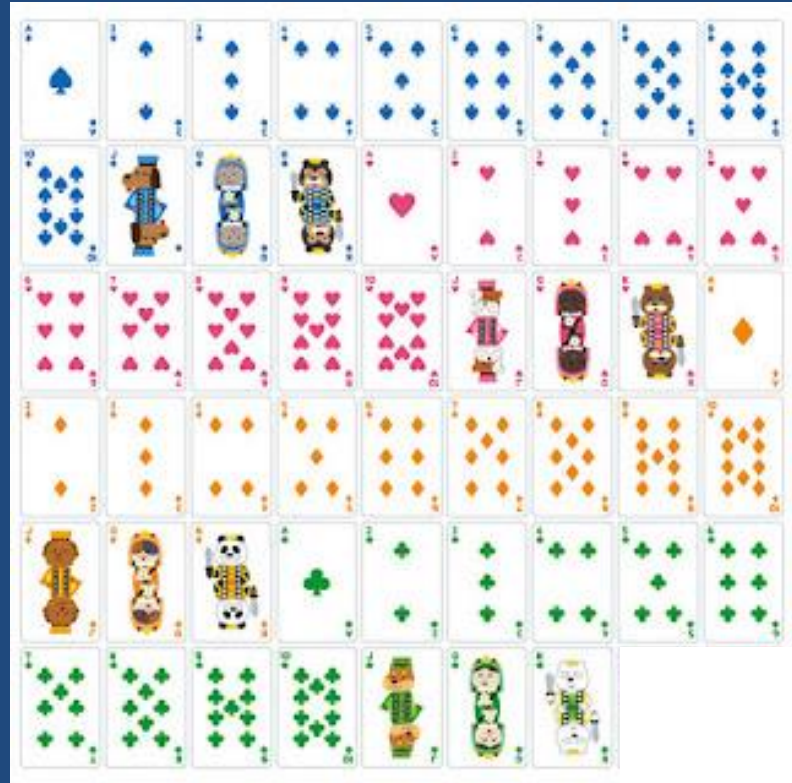
裏が
出る

全事象

試行の事象をすべて集めた集合を
全事象とよぶ



トランプの
カードを
1枚引く



52通り

数学的確率

すべての事象が同程度に確からしく
起きると仮定できるとき

$$\frac{\text{事象の数}}{\text{全事象の数}}$$

を**数学的確率**とよぶ



数学的確率

例: トランプのカードを引き
それが奇数である数学的確率は?



$$\frac{28}{52} = \frac{7}{13}$$

2種類の確率

確率は2種類に分けることができる

- 数学的確率
- 経験的確率



仮定

数学的確率は事象が
「同程度に確からしく」
起きることを前提にしている

例えば、じゃんけんの
グー、チョキ、パーは
同程度に確からしく出るだろうか？



同程度に確からしいか？

じゃんけんは
グーから始めると

- グーが
出しやすい
- パーは動作が
単純で次に
出しやすい
- チョキは出しにくい



<https://www.nansuka.jp/nansuka-editors/p000965/>

同程度に確からしいと言えない

同程度に確からしいか？

数学的確率は事象が
「同程度に確からしく」
起きることを前提にしている

前提が成り立たないと
数学的に解くことは困難

➡ 確率は分らないか？

経験的確率

ある試行を n 回行ったとき
事象 A が r 回起き
試行回数を増やすと相対度数 $\frac{r}{n}$ が
一定の値となるとき

$$P(A) = \frac{r}{n}$$

を事象 A の経験的確率とよぶ

経験的確率



https://twitter.com/au_official/status/1343376481557233665

au 2017年のCMソング「やってみよう」, WANIMA

経験的確率

725人から集めた11567回の
じゃんけんの結果

<https://toyokeizai.net/articles/-/295436?page=2>

- グー: 4054回 $\frac{4054}{11567} = 0.3505$
- チョキ: 3664回 $\frac{3664}{11567} = 0.3168$
- パー: 3849回 $\frac{3849}{11567} = 0.3328$

2種類の確率の関係

確率は2種類に分けることができる

- 数学的確率
- 経験的確率

数学的確率を計算できる
事象について
経験的確率を求めることも
可能



数学的確率と経験的確率



表が出る



表が出る



裏が出る

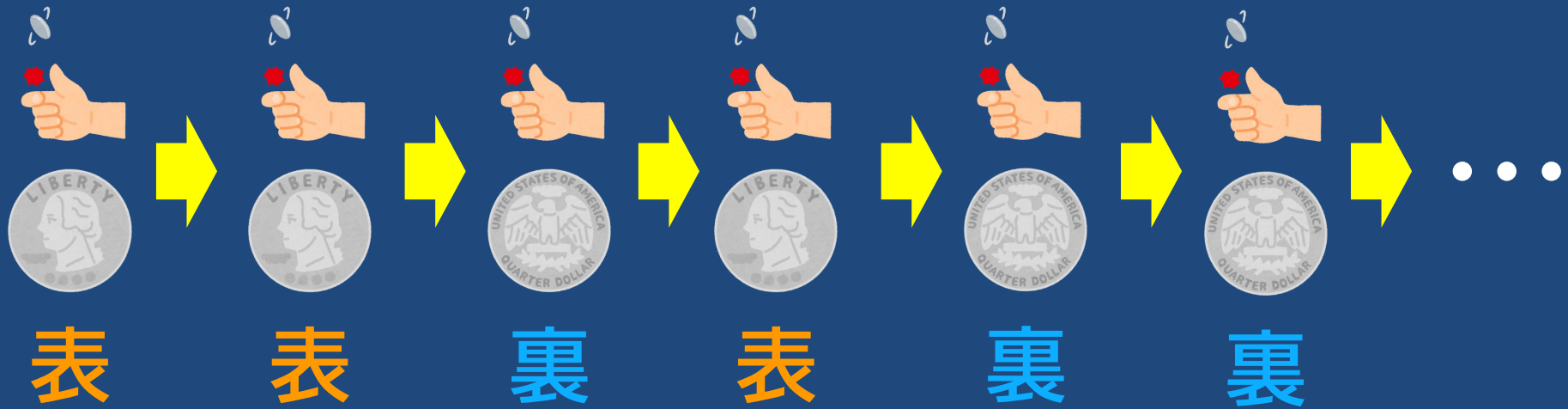
コイントスで
表が出る事象の

経験的確率は $\frac{2}{3}$

数学的確率は $\frac{1}{2}$



数学的確率と経験的確率



多数回繰り返すと
経験的確率は数学的確率と等しくなる

大数の法則

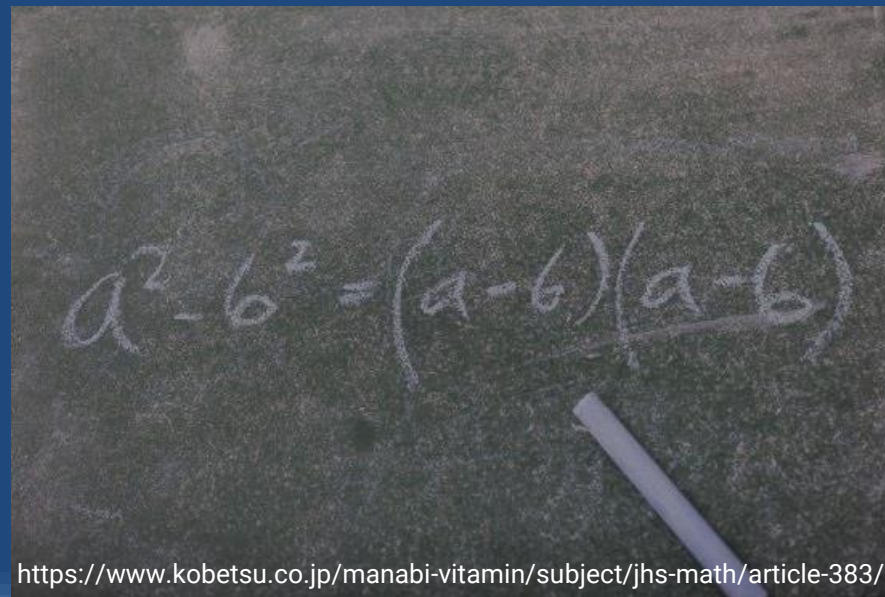
Take home notes

- 確率は数学的確率と経験的確率に分けることができます
- 数学的確率は全事象の数で対象となる事象の数を割ったものです
- 経験的確率は実際に試行を行い試行の回数で対象となる事象が起きた回数を割ったものです

おしながき

統計とは何か

- 統計のしくみ
- 統計と確率
- 統計の役目



2つの統計学

統計学は大きく

- 記述統計学
- 推測統計学

の2つに分けられる



記述統計学

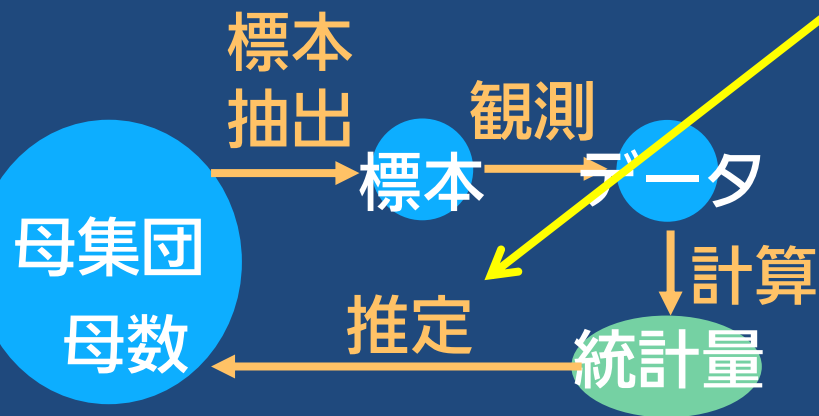
データに対して統計量を用いて
データの特徴を記述する



推測統計学

データをもとに推定をおこなう

- 統計的推測
- 仮説検定



今年は豊作になる ← 自然科学者
から
搾油機の権利を
買い占めてみたら
カンタンに ← 実業家
儲かった....
↓
よって、
金儲けはくだらん!
↑
結論が哲学者



A cartoon illustration of a philosopher with a beard and a white cloth draped over his shoulder, standing with his arms crossed.

推測統計学

データをもとに推定をおこなう

- 統計的推測
- 仮説検定

仮説を立て、それが正しいかを
客観的に判断する

仮説：兵庫県民の過半数は
阪神タイガースを
応援している



Take home notes

- 統計学は
記述統計学と推測統計学に分けることができます
- 記述統計学は統計量を用いて
データの特徴を記述します
- 推測統計学は
統計的推測と仮説検定を行います

リアクションペーパーの提出

今回の授業を通じて学んだことを
200字以上にまとめて提出してく
ださい



<https://forms.office.com/r/eMKmCyRJzb>

×切: 本日17:00

授業課題

今回の授業は
授業課題はありません