

確率という数学概念における 高校までと大学以降のギャップ

島谷健一郎(統計数理研究所)

1. 統計学での確率概念の使われ方
2. 確率論的不確実性を体得するための確率過程シミュレーション

本研究集会における統数研教員の役割:コメンテータ

統計学の専門知識を有する側から見た統計教育実践に関して気づいた点などを指摘する

- ・多くの実践例に対し、共通する意見を考えた場合が多発。
- ・短い質疑の時間にそれを伝えることは難しかった(長いコメントになって質疑時間の大半を占拠してしまう)。
- ・まとめて説明する時間をもらうことで統計教育に貢献できるかもしれない ...

2020年度

科学哲学という統計学への第3の入り口: 1は数学、2は実データ

2023年度(2024年3月)

統計教育に欠かせない**科学哲学**の基礎知識

森元良太
(2024)「統計学
再入門: 科学哲
学から探る統計
思考の原点」

1. 統計学での確率**概念**の使われ方
2. 確率論的不確実性を体得するための**確率過程シミュレーション**

2022年度(2023年3月) 授業で使うデータに関する試みと提案

1. 書籍やインターネットで公開されているデータ
2. **人工的**に作った数値
3. 授業時に学生自身で調査をしてデータを作る

- ・有益・実践的な優れた講演は他の方から提供される。
- ・今はあまり注目されていないように感じられるが**今後重要視される**であろう視点を提供する、という意図です

確率という数学概念における高校までと大学以降のギャップ

島谷健一郎(統計数理研究所)

1. 統計学での確率概念の使われ方

1.1 真値と誤差論

1.2 集団的思考

1.3 予測におけるごまかし

1.4 帰納推論を補佐する演繹推論

統計教育に欠か
せない**科学哲学**
の基礎知識

2. 確率論的不確実性を体得するための確率過程シミュレーション

人工的に作っ
た数値

1. 統計学における確率概念の使われ方:

確率と統計 セットで登場 でもそもそもは数学として別の概念

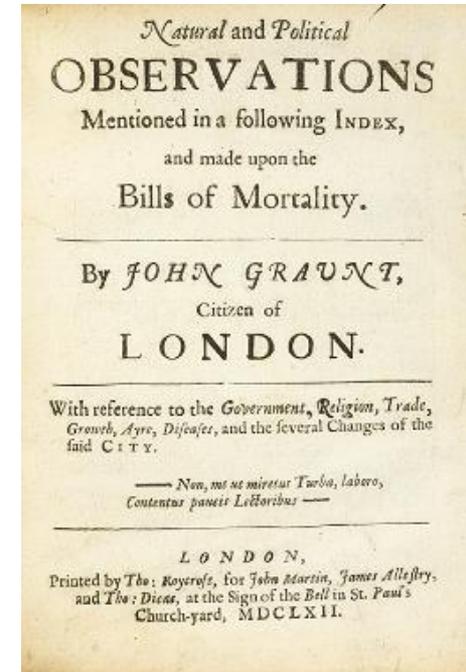
確率の起源: パスカル...???

統計の起源のひとつ: ロンドンの人口の推定

(a) ジョン・グラント

(b) 『死亡表に関する自然のおよび政治的諸観察』

確率は出
てこない



森元良太著「統計学再入門: 科学哲学から探る統計思考の原点」

図 1.6 ジョン・グラントと 1662 年出版の『死亡表に関する自然のおよび政治的諸観察』

1. 統計学における確率概念の使われ方:

確率と統計 セットで登場 でもそもそもは数学として別の概念

確率の起源: パスカル...???

統計の起源のひとつ: ロンドンの人口の推定

確率と統計がセットで登場するのは、ガウス、
ラプラスによる正規分布の導出以降?

- ・計測**誤差**: 観測値は**真値**の周辺に散らばる
- ・標本平均と真値に関する仮定 > 正規分布の導出
- ・正規分布は経験的な散らばりと合う
- ・正規分布を仮定 > 標本平均を真値に近似として使うことの妥当性と限界が標準偏差や分散で定量的に示される(一致性、不偏性、有効性、...)。

真値と誤差論 の発想

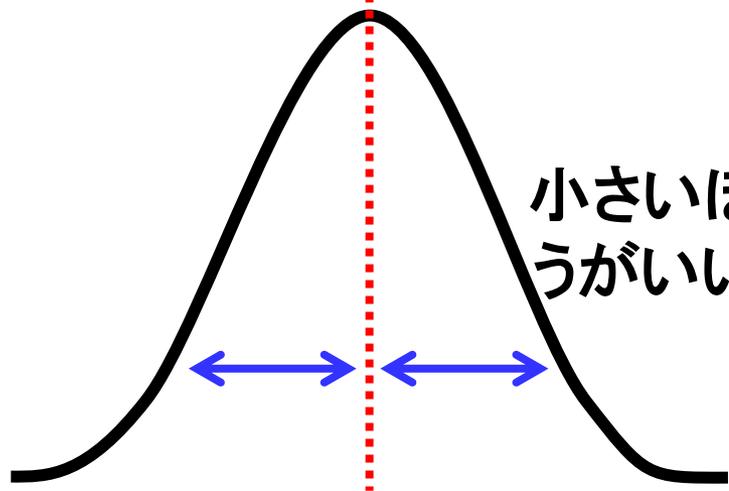
物理学や天文学などで誤差論から正規分布は導出された。

誤差論

平均 = 実在

分散

小さいほうがいい



変異 = 誤差, 非実在

物理学や天文学などで誤差論から正規分布は導出された。 高校の統計ではヒストグラムの細分で正規分布を導入...??

図 2.1 のような国内のある年齢の男子生徒の身長の数値分布（身長を x 、人数比を y としたときのヒストグラム）は、図中にあるような正規分布と呼ばれる釣鐘状の連続曲線で近似できることが知られています。

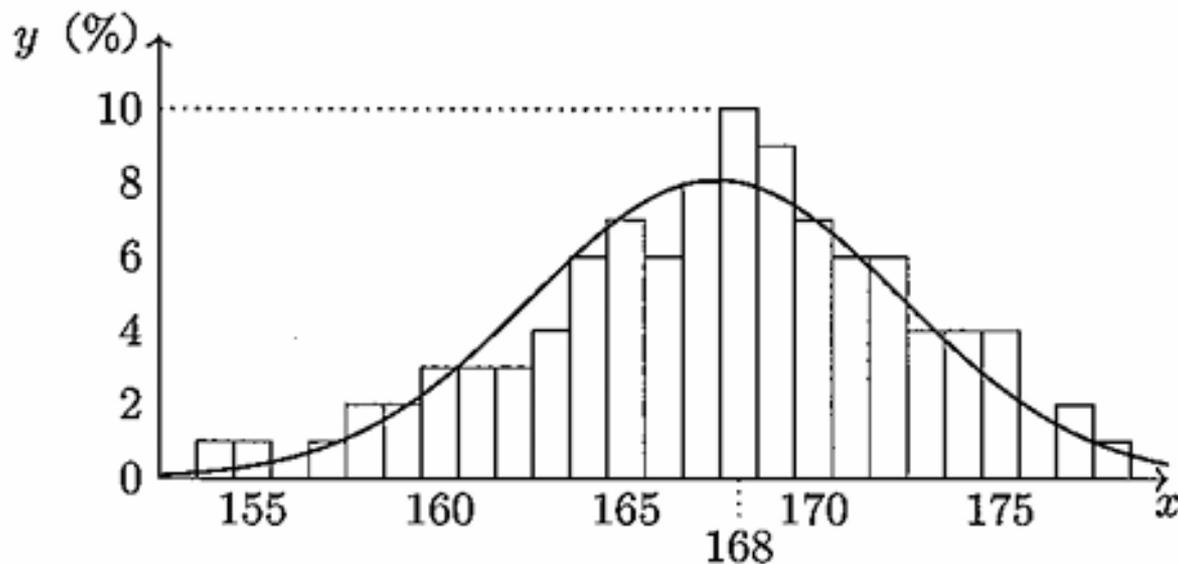


図 2.1

・度数分布がいつのまにか割合に変わり、確率になる:

「身長が165-166cmの人の数」が「身長が165-166cmである人の割合」、「ある人の身長が165-166cmである確率」に変わって連続型確率分布へ…???

図 2.1 のような国内のある年齢の男子生徒の身長の度数分布（身長を x ，人数比を y としたときのヒストグラム）は，図中にあるような正規分布と呼ばれる釣鐘状の連続曲線で近似できることが知られています。

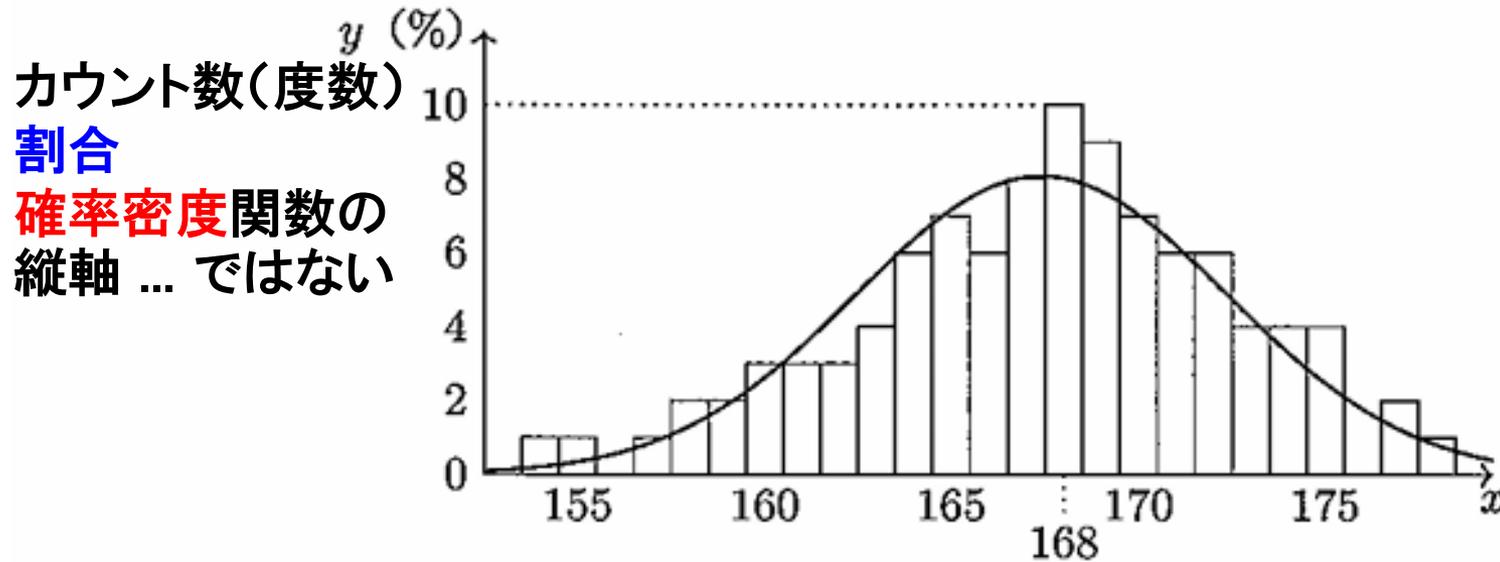


図 2.1

- ・高校の統計では**誤差論**でなく**集団的思考**で正規分布を導入する？
- ・度数分布がいつのまにか割合に変わり、確率になる：
 「身長が165-166cmの人の数」が「身長が165-166cmである人の割合」、「ある人の身長が165-166cmである確率」に変わって連続型確率分布へ…???

図 2.1 のような国内のある年齢の男子生徒の身長の度数分布（身長を x 、人数比を y としたときのヒストグラム）は、図中にあるような正規分布と呼ばれる釣鐘状の連続曲線で近似できることが知られています。

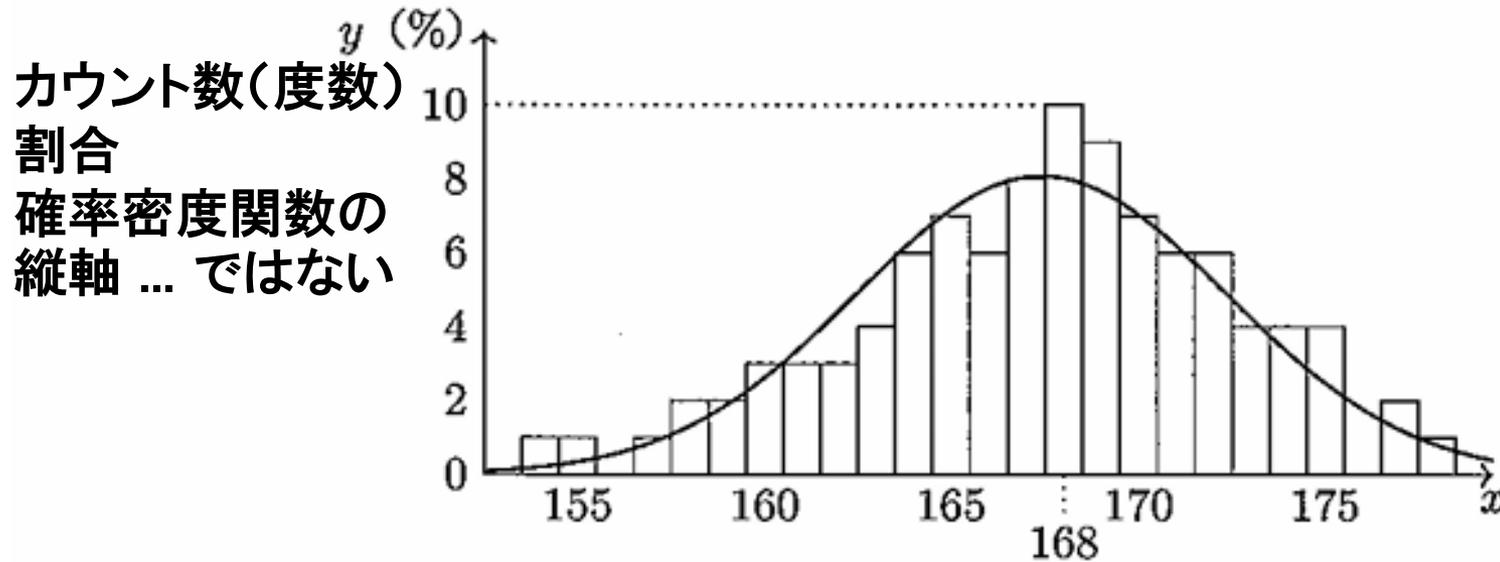
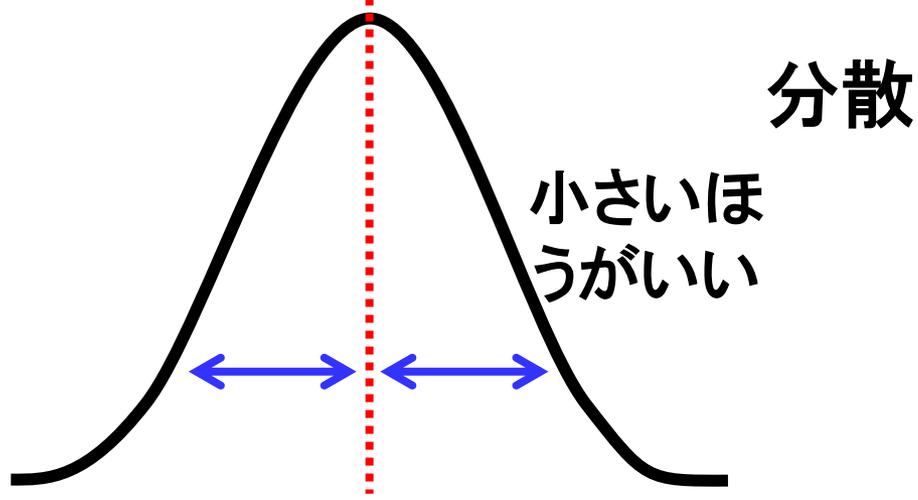


図 2.1

物理学や天文学などで**誤差論**から正規分布は導出された。
生物学、社会学、心理学、... で統計学は**集団自体を基礎的**
なものとする。

誤差論

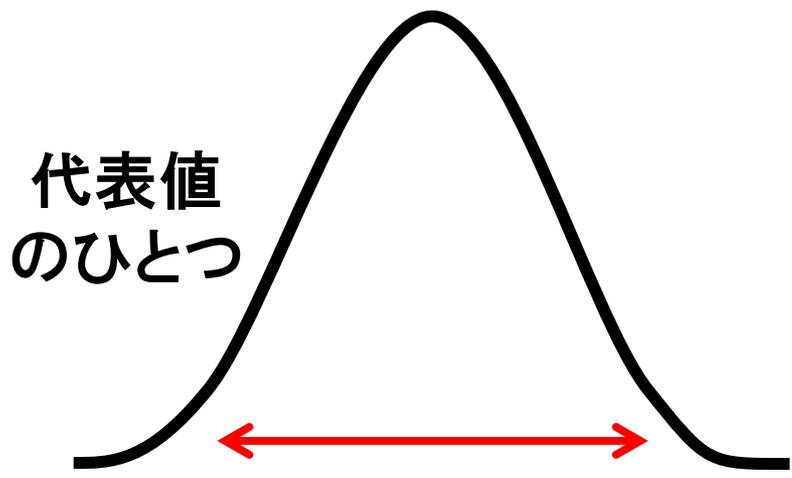
平均 = 実在



変異 = 誤差, 非実在

集団的思考

平均 = 代表値の一つ



変異 = 実在

データの分布を観る

レポート例: 大谷投手やダルビッシュ投手の 球速の分布

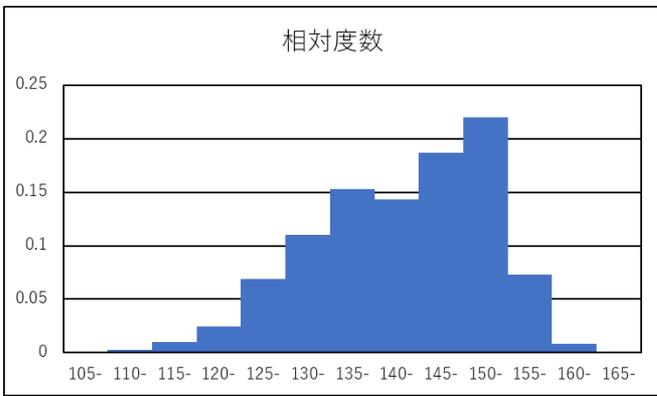


図 1 2022MLB 球速分布

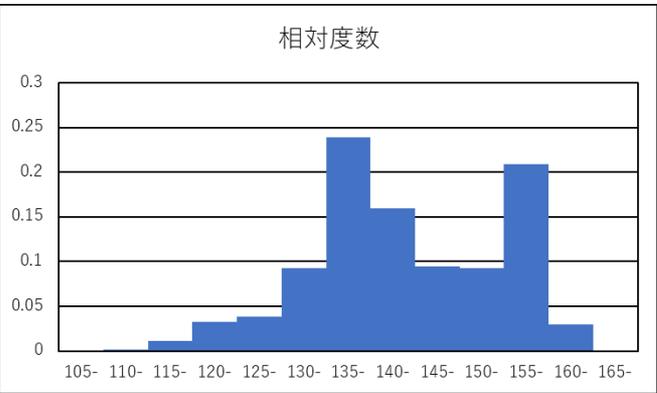


図 3 2022大谷翔平選手 球速分布

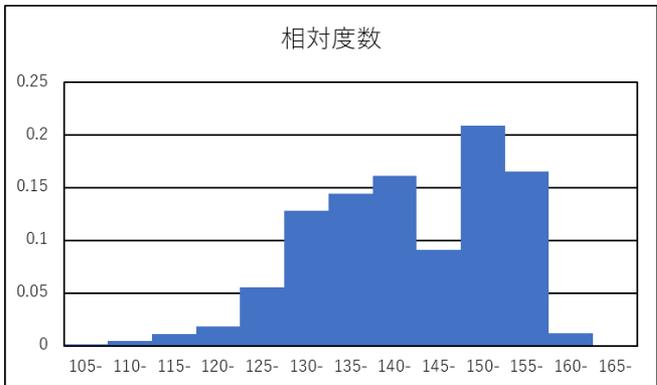


図 5 2021大谷翔平選手 球速分布

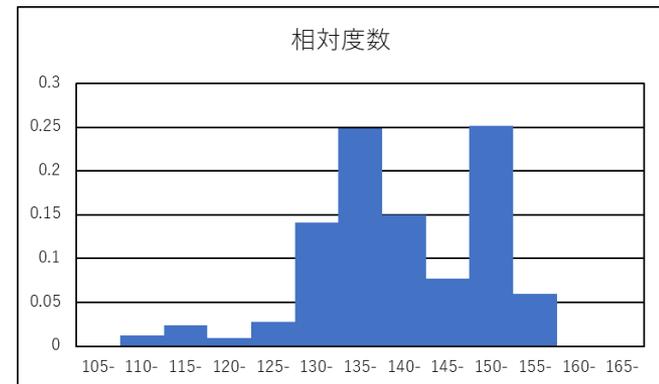
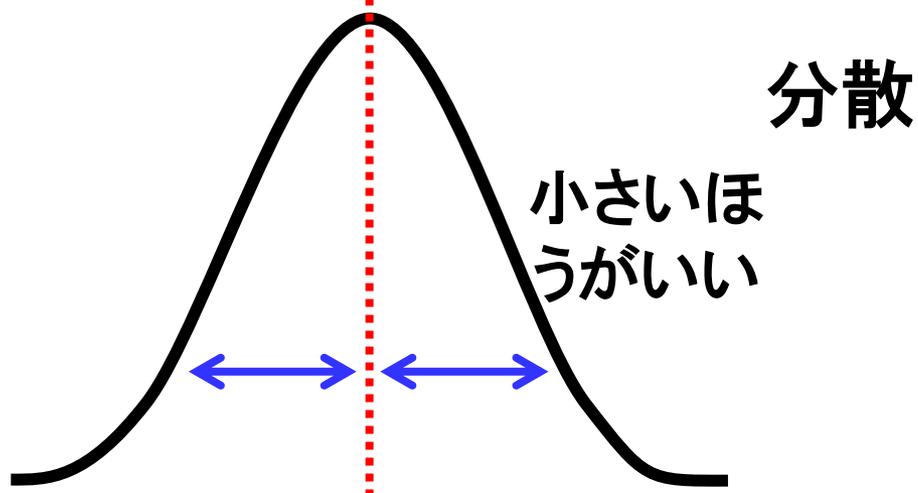


図 7 2022ダルビッシュ有選手 球速分布

物理学や天文学などで**誤差論**から正規分布は導出された。
生物学、社会学、心理学、... で統計学は**集団自体を基礎的**
なものとする。

誤差論

平均 = 実在

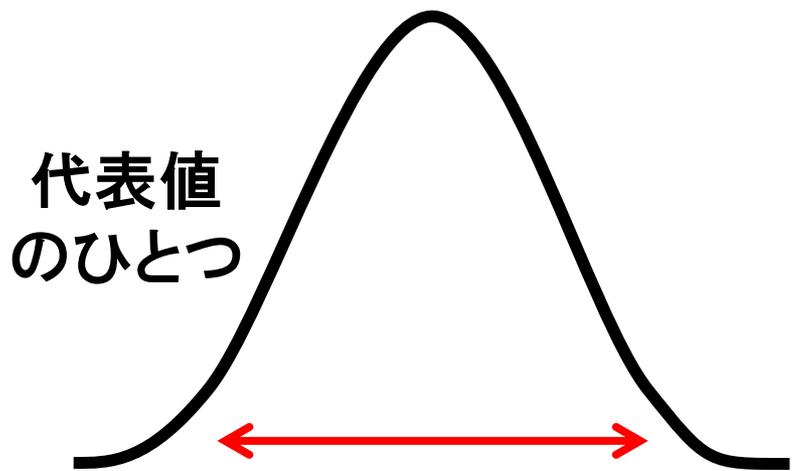


変異 = 誤差, 非実在

オレンジジュースのビタミン、
化学実験、測定誤差は減らしたい

集団的思考

平均 = 代表値の一つ



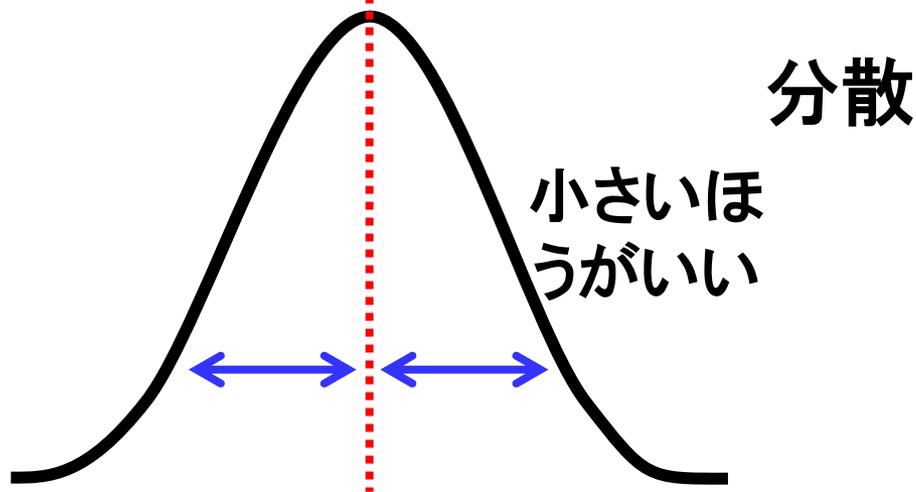
変異 = 実在

キウイなどの果実: 個体差
(変異)に本質あり

物理学や天文学などで**誤差論**から正規分布は導出された。
生物学、社会学、心理学、... で統計学は**集団自体を基礎的**
なものとする。

誤差論

平均 = 実在

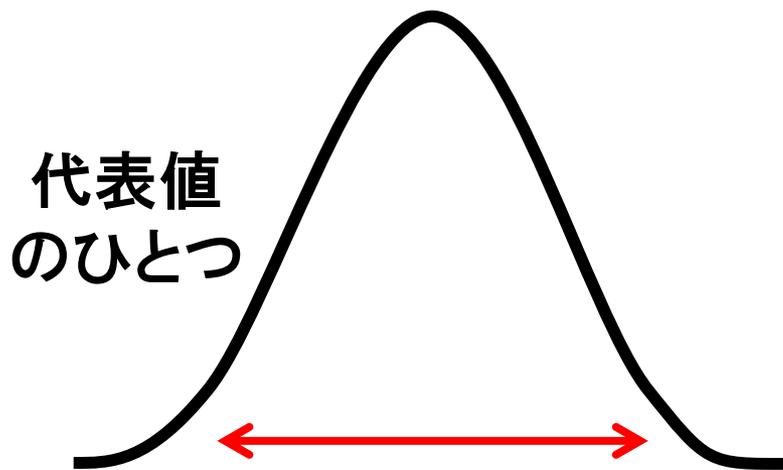


変異 = 誤差, 非実在

今日の発表: 心拍数と
計測機器の使い方

集団的思考

平均 = 代表値の一つ



変異 = 実在

個人差: 今日とは言及せず

1. 統計学での確率概念の使われ方

1.1 真値と誤差論

1.2 集団的思考

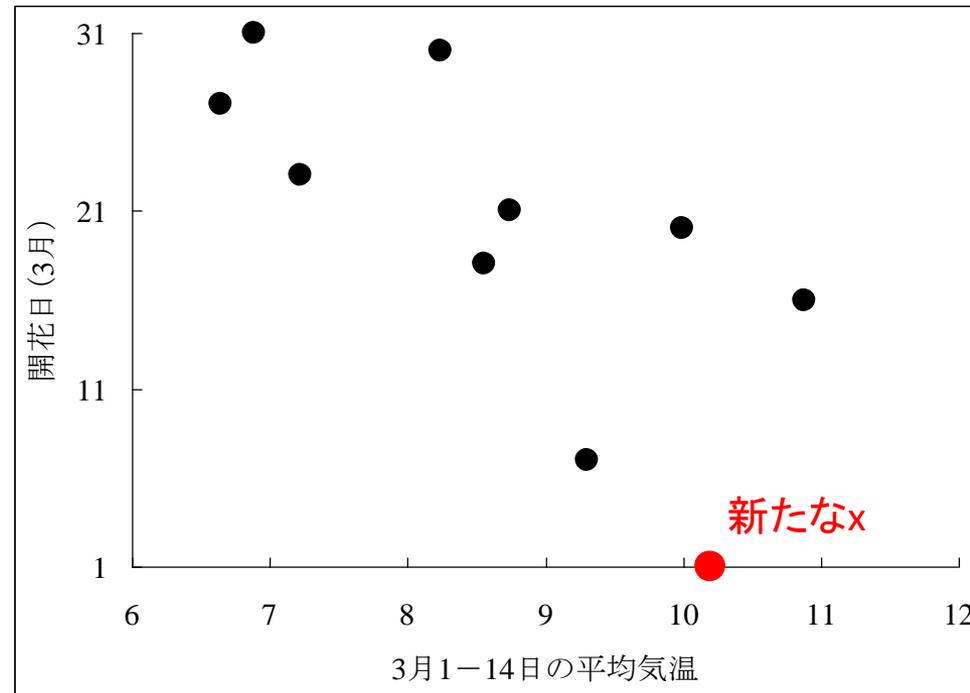
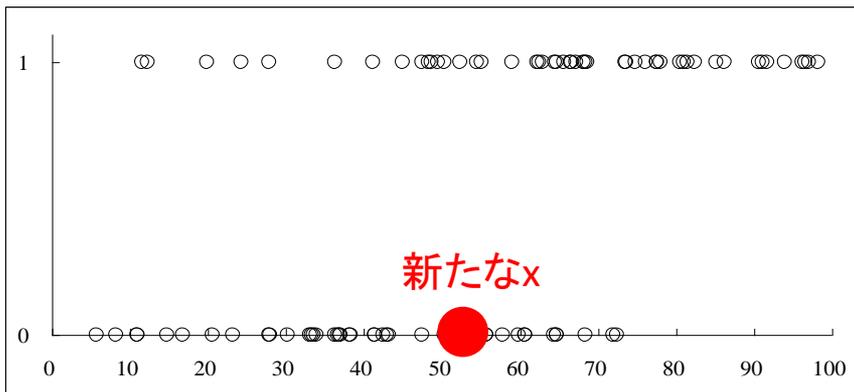
1.3 予測におけるごまかし

1.4 帰納推論を補佐する演繹推論

2. 確率論的不確実性を体得するための確率過程シミュレーション

森元良太 (2024)「統計学再入門: 科学哲学から探る統計思考の原点」

過去のデータと新たな説明変数(●)から目的変数を予測



1. 統計学での確率概念の使われ方

1.1 真値と誤差論

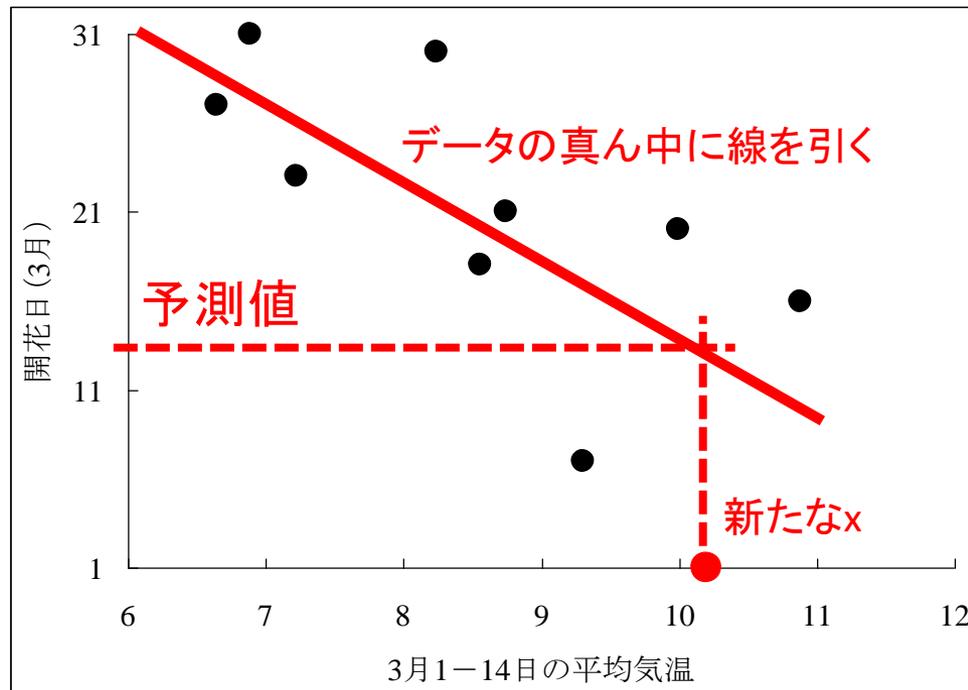
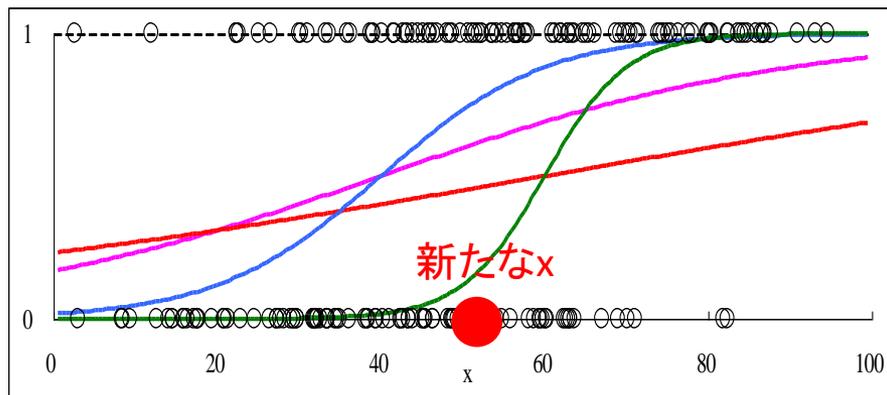
1.2 集団的思考

1.3 予測におけるごまかし

1.4 帰納推論を補佐する演繹推論

2. 確率論的不確実性を体得するための確率過程シミュレーション

過去のデータと新たな説明変数(●)から目的変数を予測



説明変数が与えられたとき、目的変数をピッタリあてては不可能

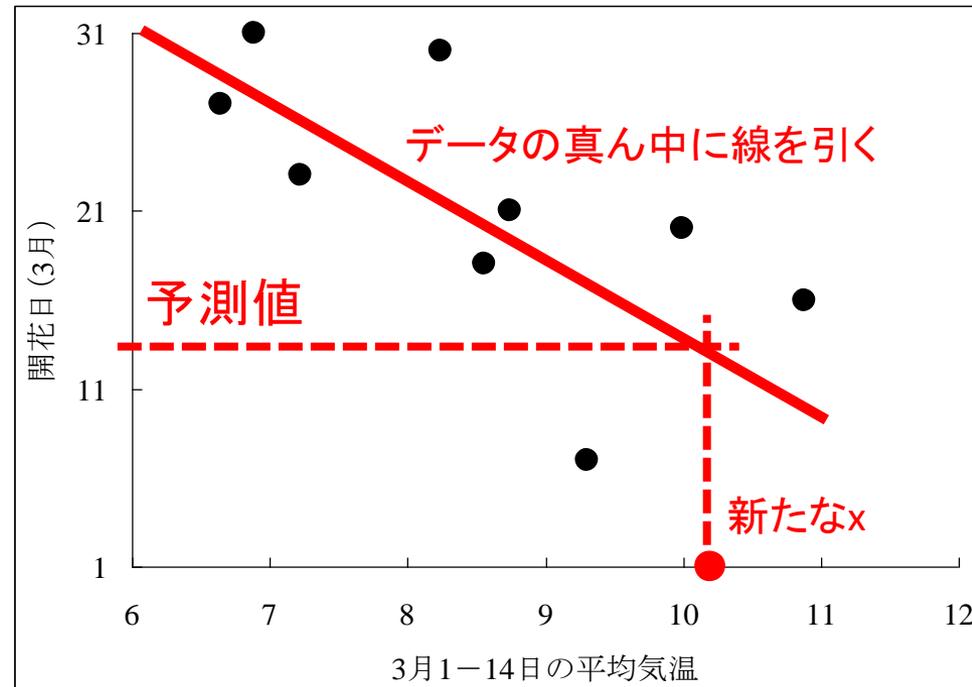
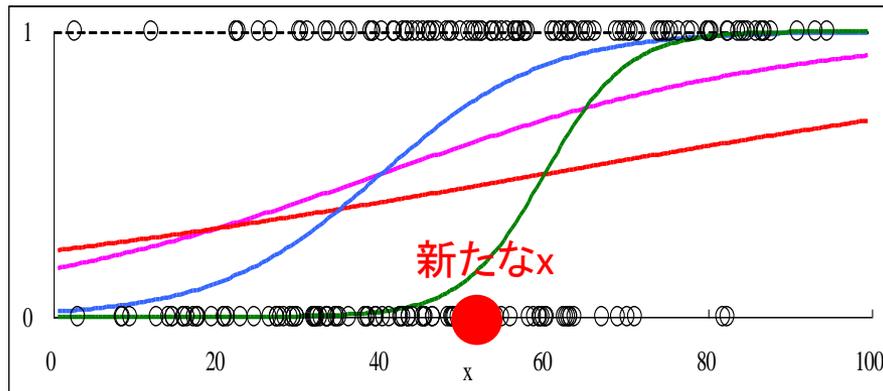
統計モデルによる予測

0. まずデータを全部観る

- 1.ピッタリ当てるのは無理そう
- 2.でもベストを尽くしたい
- 3.どのくらい外すかも「予測」

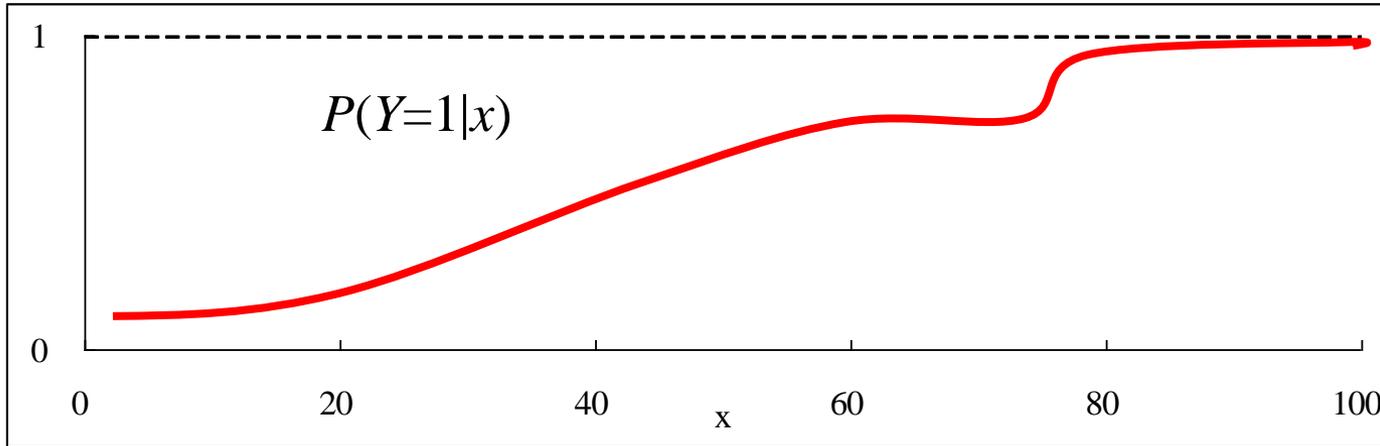
1. 確率(分布)で誤魔化す(?!)
2. 最適解(例:最尤法 さいゆうほう)
3. 不確実さも明示(例:乱数生成・シミュレーション)

過去のデータと新たな説明変数(●)から目的変数を予測



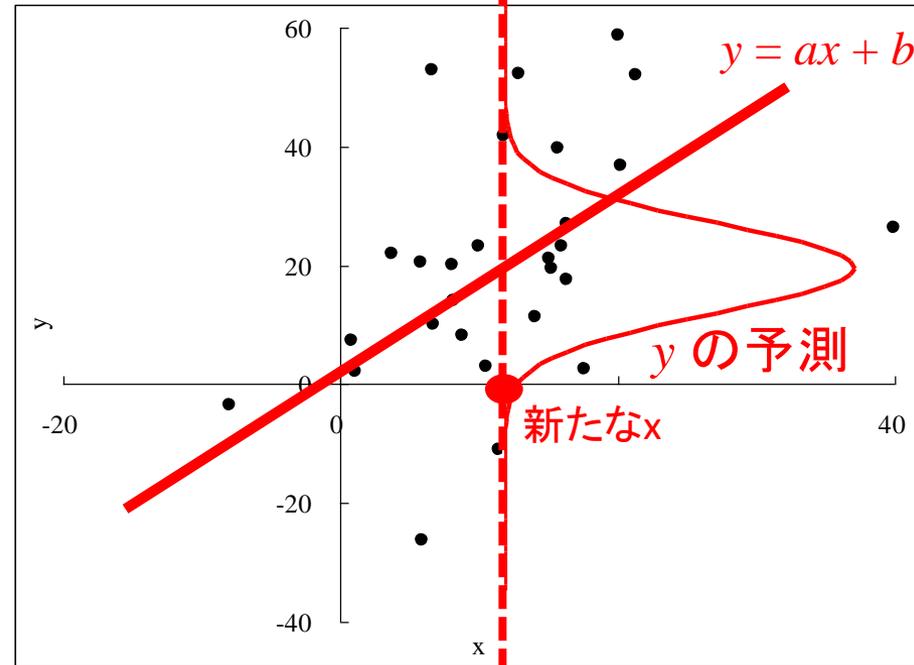
ピッタリ当てられないので確率で誤魔化す(?!)という発想

ある x が与えられたとき Y が1である確率をデータから予測する

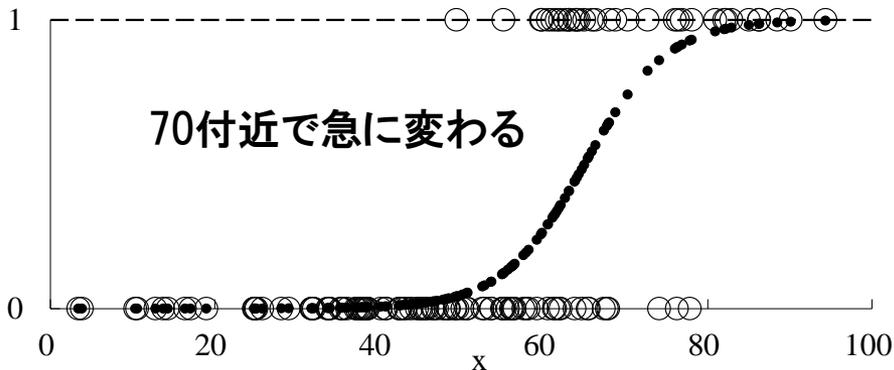


統計学の役目は0と1の確率を出すところまで。1と予測するか0と予測するかは、自分で判断せよ。

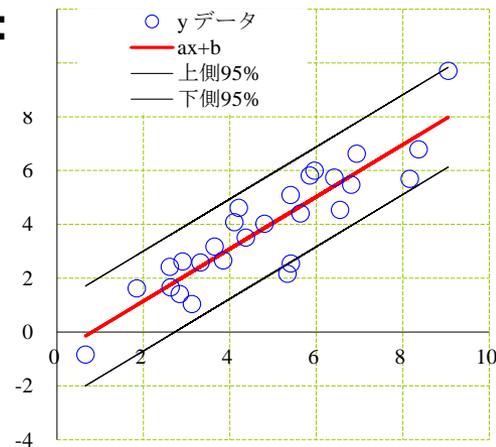
新たな x に対して、 y は、直線 $y = ax + b$ を平均パラメータ (m), ある s を標準偏差パラメータとする正規分布に従う、と「予測」する。



0や1に近い確率が多い:予測できる

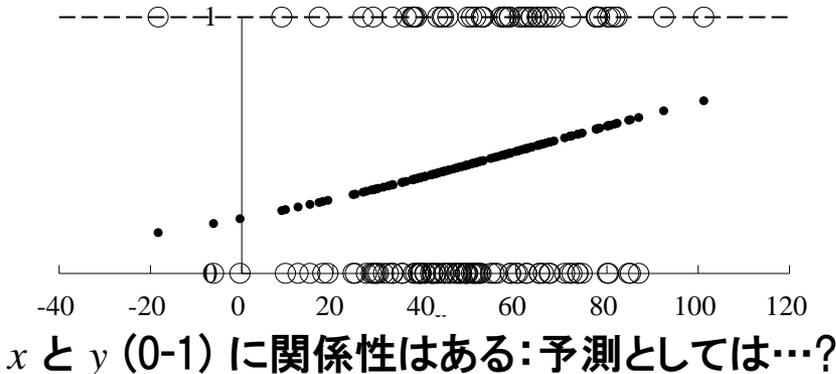
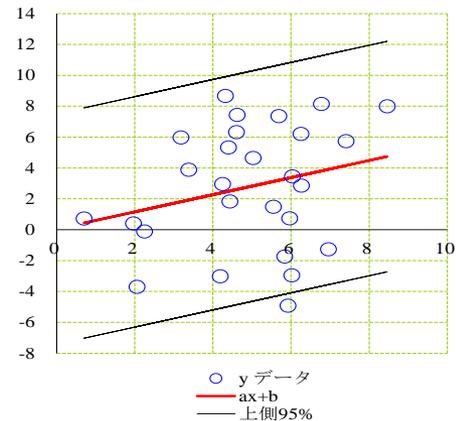


標準偏差小:
予測できる



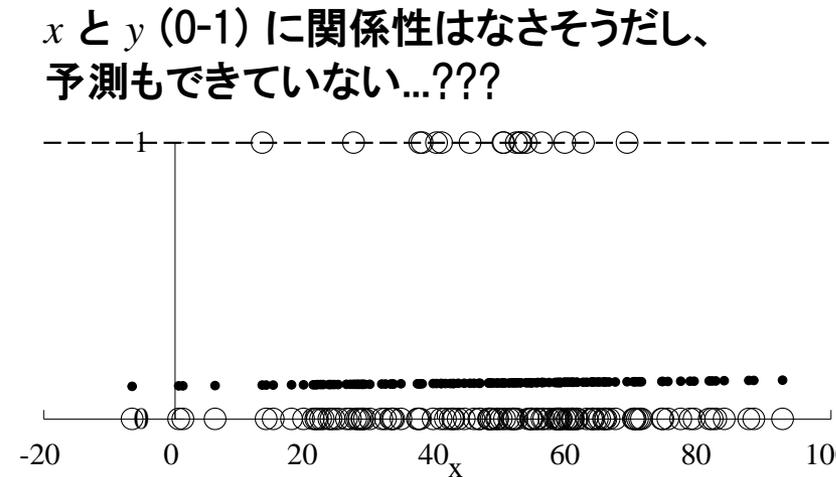
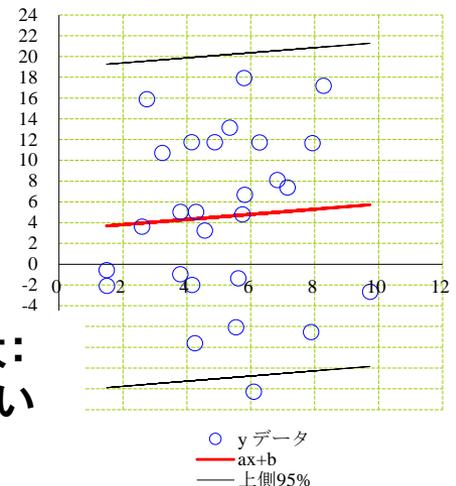
一般化線形
モデル(GLM)
による予測の
イメージ

標準偏差大:
予測になってない?



x と y (0-1) に関係性はある:予測としては…?

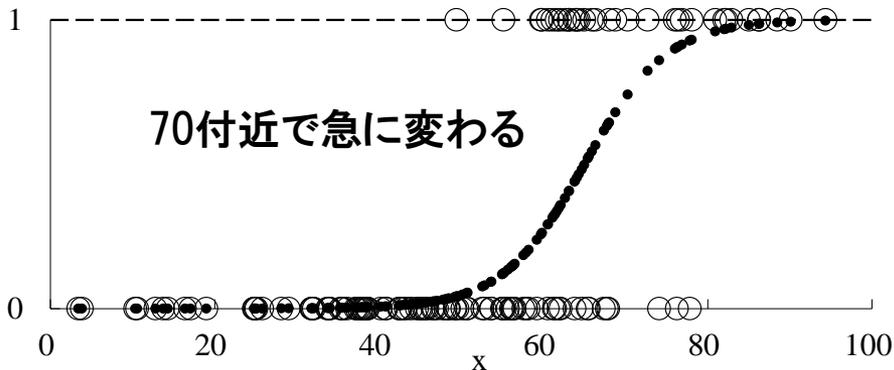
有意性と予
測は別概念



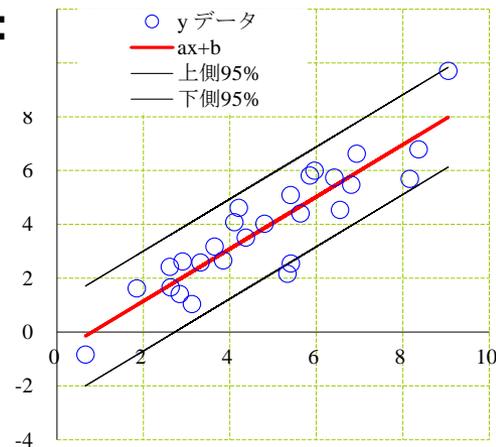
x と y (0-1) に関係性はなさそうだし、
予測もできていない…???

標準偏差とても大:
予測になってない

0や1に近い確率が多い:予測できる



標準偏差小:
予測できる



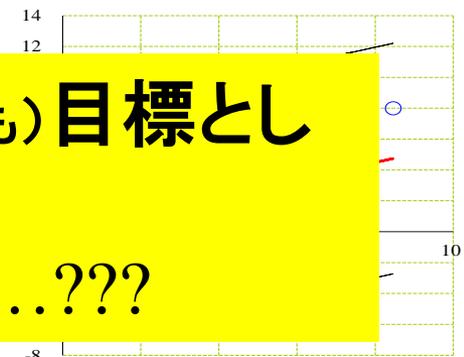
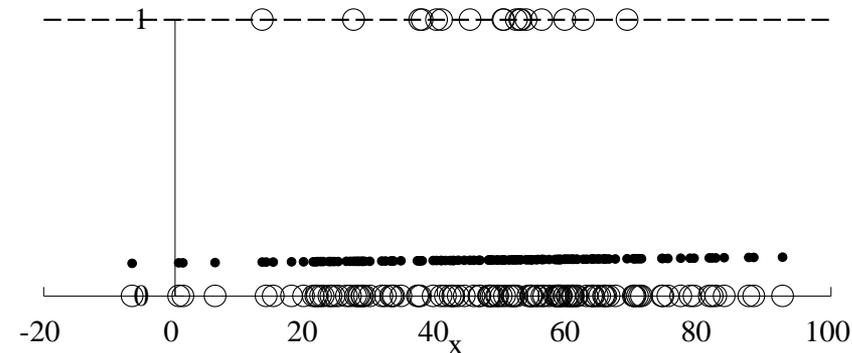
一般化線形
モデル(GLM)
による予測の
イメージ

・予測といってもズバリの中は難、(そもそも)目標としていない。確率分布で誤魔化す...

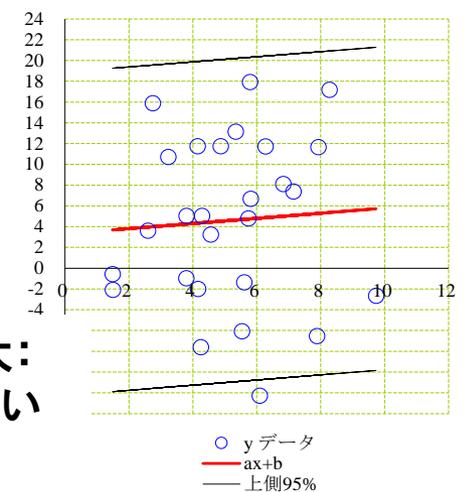
・「予測できない」という予測: 統計思考...???

x と y (0-1) に関係性はある:予測としては...?

x と y (0-1) に関係性はなさそうだし、
予測もできていない...???



標準偏差とても大:
予測になってない



1. 統計学での確率概念の使い方

1.1 真値と誤差論

1.2 集団的思考

1.3 予測におけるごまかし

1.4 帰納推論を補佐する演繹推論

森元良太 (2024)「統計学再入門:
科学哲学から探る統計思考の原点」

2. 確率論的不確実性を体得するための確率過程シミュレーション

帰納推論

過去のデータから予測

例
一昨日昼の学食は混んでいた
昨日の昼の学食も混んでいた
⋮
今日の昼も混むだろう

未来は過去に似ている
自然の一様性原理

この前提がないと帰納は正しくない

Humeの問題: 帰納は原理的に正当化できない

演繹(えんえき):前提が正しければ結論は必ず正しい

AならばB
Aである

Bである

AならばB
Bでない

Aでない

検証は演繹
ではない

~~AならばB
Bである

Aである~~

科学哲学者 Karl Popper

科学とは、**反証**可能な仮説や理論の中で、これまで反証されていないものを一時的に受け入れる営み。帰納は科学の方法ではない。

帰納推論

過去のデータから予測

例
一昨日昼の学食は混んでいた
昨日の昼の学食も混んでいた
⋮

今日の昼も混むだろう

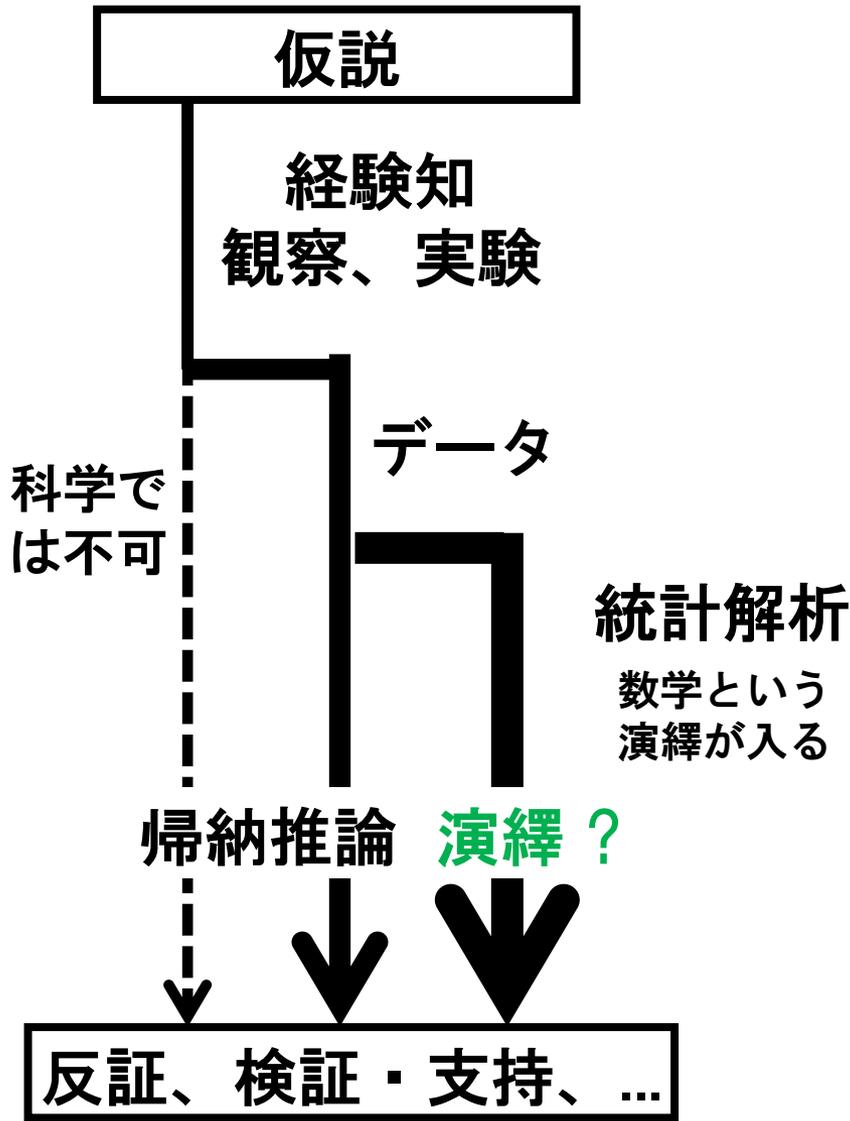
未来は過去に似ている
自然の一様性原理

この前提がないと帰納は正しくない

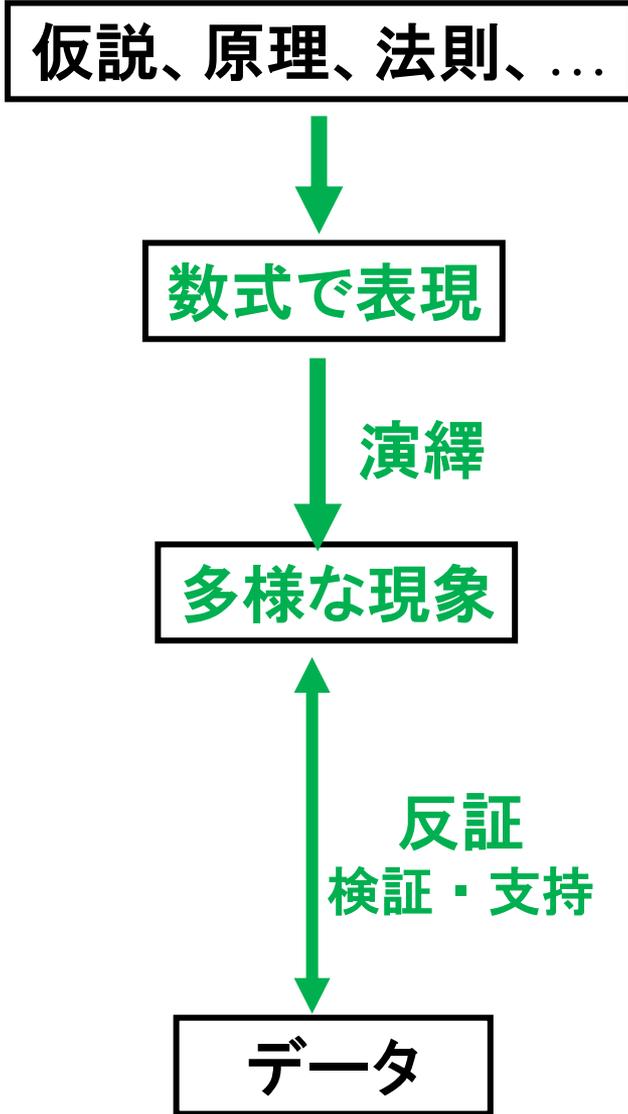
Humeの問題: 帰納は原理的に正当化できない

1.4 帰納推論を補佐する演繹推論

(生物学や社会科学などでよく見られる) 帰納推論

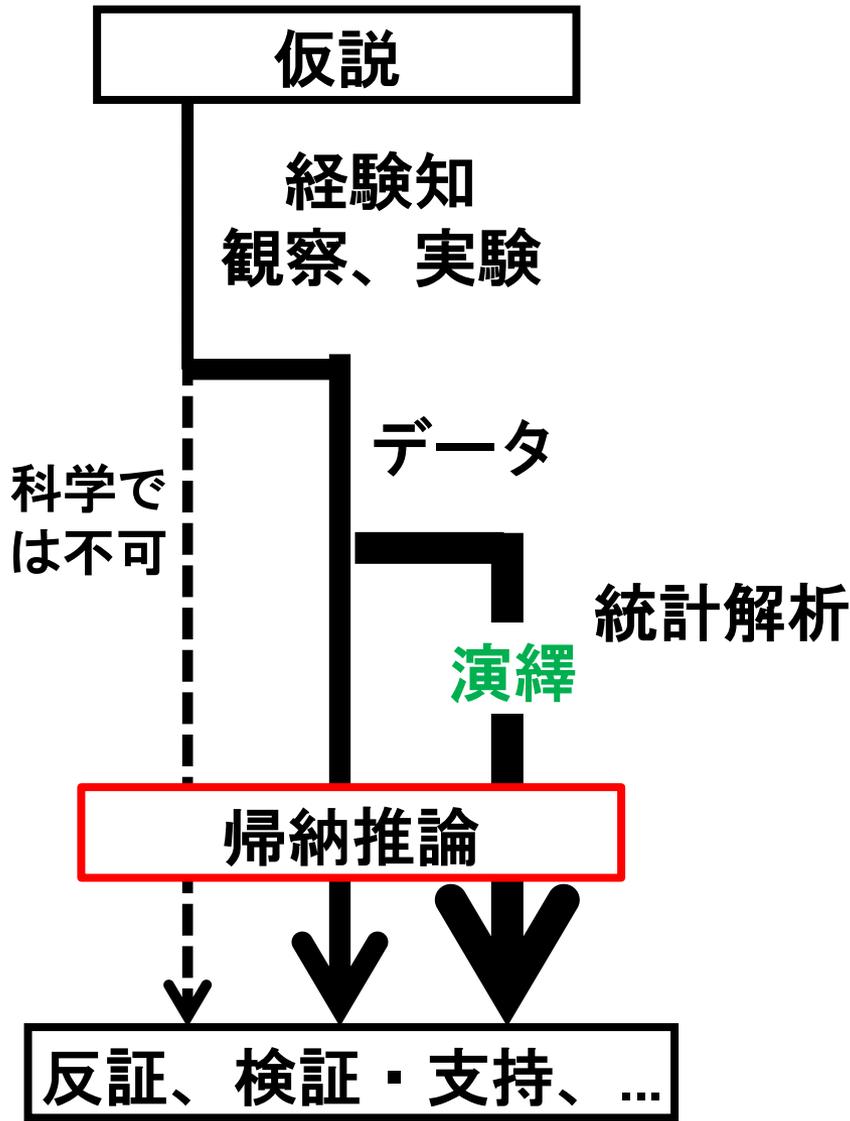


(古典物理に代表される) 科学における演繹推論

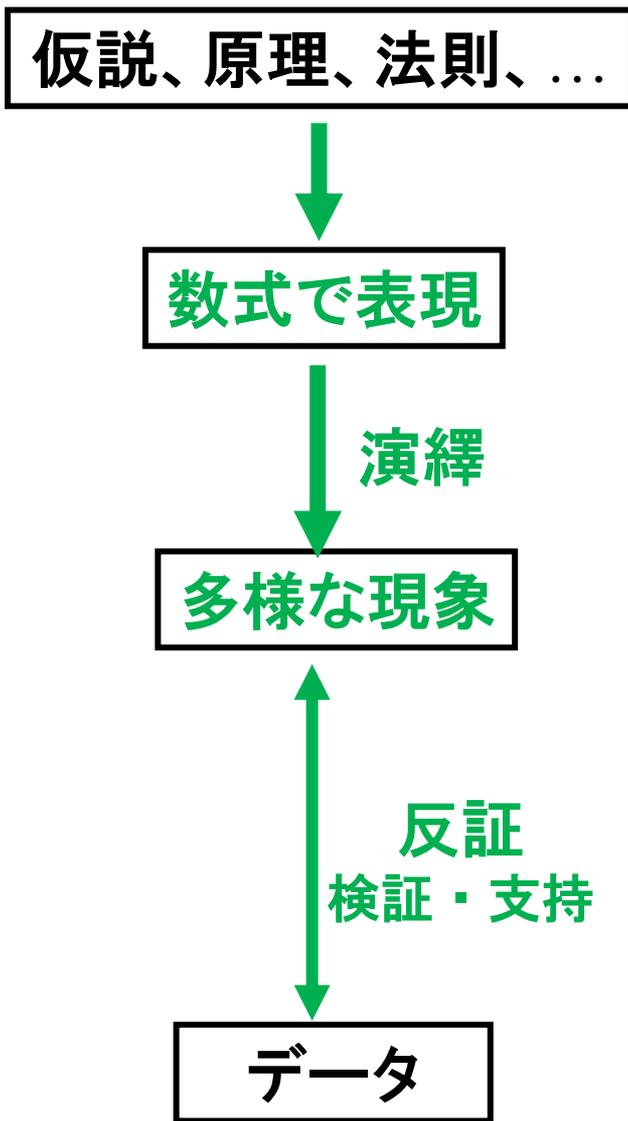


1.4 帰納推論を補佐する演繹推論

(生物学や社会科学などでよく見られる) 帰納推論



(古典物理に代表される) 科学における演繹推論



1. 統計学での確率概念の使い方

1.1 真値と誤差論

1.2 集団的思考

1.3 予測におけるごまかし

1.4 帰納推論を補佐する演繹推論

森元良太 (2024)「統計学再入門：
科学哲学から探る統計思考の原点」

・サイコロなどの「確率」概念とずいぶん違う...

・予測目的統計モデルの中で使われる確率概念：

データは不確実に変動する

2024年8月2日(金)-3日(土)

「数学を用いる生物学：理念・概念と実践・方法論」

10:00-10:40 島谷健一郎 (これからの)生物学で使われる数学をまとめ直す

(質疑)「不確実な変動」を体得するには、別な学習が必要ではないか

10:40 - 12:10 大槻久(総研大・統合進化セ)確率過程を用いる数理生物学

2. 確率論的不確実性を体得するための確率過程シミュレーション

- ・総個体数:一定 N
 - ・ N 個体(親)のすべてが死に,新しい N 個体(子供)が誕生する
 - ・子供の親は, N 個の親の中からランダムに割り当てる
 - ・この死亡と生誕を繰り返していく.
 - ・最初(第0世代)、 N 個体すべてが異なる種だったとする.
-
- ・この仕組みの元で何種くらいが永遠に共存できるか
 - ・総個体数 N が10のとき,100世代での種数を予想する。
 - ・総個体数 N が100のとき,100世代での種数を予想する。
 - ・総個体数 N が100のとき,1000世代での種数を予想する。
-
- ・1000世代で10種なら,10種くらいが永遠に共存するという予想を立てられる。3種なら3種、5種なら5種。
 - ・100世代で10種なら10種が永遠に共存と予想していいか? 10000世代まで見てから予想するべきか?

確率という数学概念における高校までと大学以降のギャップ

島谷健一郎(統計数理研究所)

1. 統計学での確率概念の使われ方

1.1 真値と誤差論

1.2 集団的思考

1.3 予測におけるごまかし

1.4 帰納推論を補佐する演繹推論

科学哲学や推論の基礎知識と絡む:
統計のわかりにくさ感、教えにくさ感
の原因のひとつ

統計学の中で確率は様々な使われ方をする:
統計のわかりにくさ感、教えにくさ感の原因のふたつめ

2. 確率論的不確実性を体得するための確率過程シミュレーション

確率論的不確実性の体得の方がいろいろな場面で重要でないか
乱数生成・シミュレーションを優先 ... ?

- ・データに基づく統計思考教育:急速に進歩し手本になる実践例多数
- ・科学的推論と統計思考・哲学的思考の教育:次の課題
- ・確率論的不確実性を体得するシミュレーション:次の課題