

2023年度秋学期 統計学 第1回
イントロダクション — 統計的なものの見方・考え方について

浅野 晃
関西大学総合情報学部



「統計的見方」
「確率的見方」
「統計学と確率」

「統計的見方」

コロナ禍は
終わったわけではありません🦠

公衆衛生学とは

感染症を扱う医学は、「公衆衛生学」👤👤

ほかの医学が扱うのは、目の前のひとりの「人」👤

公衆衛生学が扱うのは、社会を構成する「人々」👤👤

「人々」の行動を完全にコントロールはできない👤👤

感染したかどうか、検査で完全にはわからない🦠

ワクチン／は、感染を完全に防ぐわけではない

感染症と闘う統計学

社会を全体として見たときに、
感染の拡がりを抑えなければならない

統計学は、集団を全体として見て、その姿を把握する

「密閉・密集・密接の『三密』を避けよう」

「大人数の会食をやめよう」

- 統計学によって現状を把握して得られた指針
- 感染を社会全体として減らし、医療の逼迫を防ぐため
(三密や大人数の会食を避けても、絶対に感染しないというわけではない)

感染を必ずさけられるのではなくても

「密閉・密集・密接の『三密』を避けよう」

「大人数の会食をやめよう」

感染を必ず避けられるのではないのなら、いったい何のため？

一度に大人数に感染させる「クラスター」を防ぐ

一人の感染者が一人の人にしかうつさなければ、もとの感染者は回復するので、
社会全体の感染者の数は増えない

一人の感染者が感染させる人数が「平均して」一人未満なら、
社会全体の感染者数は減っていく
(実効再生産数が1未満)

個人ではなく、社会を救う

「平均して」「社会全体の」

というのが、統計学の発想です

統計学で社会全体のようにすを把握し、感染を社会全体で減らすのが↓

「密閉・密集・密接の『三密』を避けよう」


「大人数の会食をやめよう」

あなた個人👤👤を救うのではなく、社会全体🇯🇵🇺🇸を救う

「確率的見方」

確率が小さいこととは

ワクチン接種について

「コロナワクチン接種で重篤な副反応が出るのは10万人に1人の確率だといっても、その副反応が出た人にとっては100%重篤な事態だ」

それはそのとおりで、「確率が小さいかどうか」と「事態の重篤さが小さいかどうか」は関係ありません。

くじ引きで、「当たり確率」と「賞金の額」は別の問題なのと同じ

確率とは

確率とは この講義では、後半のはじめ(第9回)で説明しますが、

「くじの当たり確率 0.3」とは、次のような意味です(どちらでも同じ)

- くじを十分多くの回数引くと、そのうち10回に3回の割合で当たる
- 十分多くの人がそれぞれ1回くじを引くと、その人たちのうち10人中3人が当たりをひく

いずれにしても、
「十分多くの回数」「十分多くの人」について言っていることを
「1回」「ひとり」に当てはめている

確率がわかってても

確率がわかってても、
次の「1回」のくじ引きの結果はわからない。

確率は、くじ引きのような「ランダム現象」を扱う

ランダム現象とは、「結果に人知の及ばない現象」

確率を云々しても、人知が及ばないことには変わりはないけれど
「どんな結果になることがどのくらい多いか」を考える

期待値とは

期待値とは

さきほど「別の話」と言った「当たり確率」と「賞金の額」を結びつけて

くじ引きで考えれば、(どちらでも同じ)

- ・くじを十分多くの回数引いたときの、1回あたりに得られる賞金の平均
- ・十分多くの人がそれぞれ1回くじを引いたとき、ひとりが得られる賞金の平均

プロのギャンブラーは

いくらプロのギャンブラーでも

次の1回の賭けに勝てるかどうかはわからない

プロのギャンブラーは

日頃から多くの回数の賭けをする→
賞金の期待値の大きい賭け方を見抜いて賭けることができれば、

1回1回の賭けでは勝ち負けがあっても、
多くの賭けの合計では勝つことができる

リスクとメリットは、考慮できるか

ワクチン／の話にもどると

リスクとメリットを考慮して、といわれても

日頃から多くの回数の賭けをするギャンブラーなら
賞金の期待値を問題にすることができるけれど

一生に1度しかしないことの確率や期待値を考えるのはむずかしい

人間の思考の限界? 😞

「統計学と確率」

統計的推測とは

もうずいぶん昔ですが、1994年に
ノルウェー🇳🇴のリレハンメルで開かれた五輪の開会式で、アナウンサーが

「ノルウェー人は背の高い人が多く、平均身長は男179cm、女170cmです」

ノルウェー人全員の身長を測ったんですか??

標本調査と統計的推測

当然ながら、身長は人によって違う(分布している)

ノルウェー人全員ではなく、一部の^が人だけ(標本)を調べて、
分布全体のような気がするのか?

わかります。かなりの程度わかります。

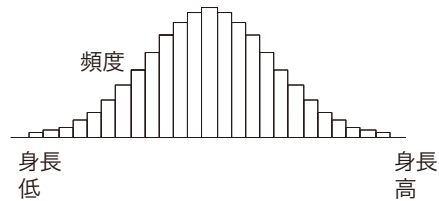
「一部の^が人」を選ぶのに、くじびきで選ぶ(無作為抽出)

くじびきで選べば、たいていはいろいろな人がまんべんなく選ばれる

無作為抽出すると

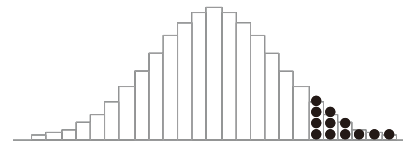
分布がこんなようすのとき

データ全体
(実際には不明)



偶然こんな標本(●)が選ばれて
しまう確率は小さい

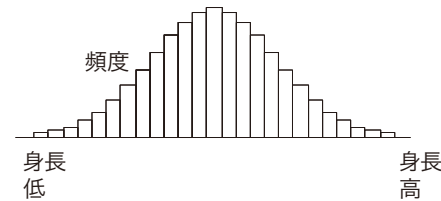
こんな標本が選ばれたら
→大きく偏った推測



無作為抽出すると

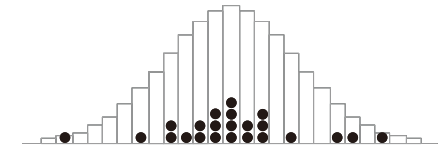
分布がこんなようすのとき

データ全体
(実際には不明)



たいていは、
こんなふう^にに選ばれる

こんなふう^にに
標本が選ばれれば
→ほぼ間違っていない推測



「たいてい」と「ほぼ」

くじびきで選べば、**たいてい**はいろいろな人がまんべんなく選ばれる
→選ばれた人の平均は、**ほぼ**全体の平均に近い

本当？ 😊

たまにはバレーボール🏐の選手みたいな人ばかり選ばれることもあるのでは。

そのとおりです。「**たまには**」そういう失敗をします。
でも、**失敗をする確率を計算**できます。

区間推定

「区間推定」という統計学の方法では、

「ほぼ」
「ノルウェー人男性全体の平均身長は、179cm~182cmの間と推測する。
この推測が当たっている確率は95%」

「たいてい」(失敗の確率5%)

と答える

リスクを(再び)考える

「ほぼ」
「ノルウェー人男性全体の平均身長は、179cm~182cmの間と推測する。
この推測が当たっている確率は95%」

「たいてい」(失敗の確率5%)

失敗の確率は「このような統計的推測を何度も行うとき、どのくらいの割合の推測が失敗するか」を表す

→1回だけ推測するとき、それが成功するか失敗するかはわからない

このような統計的推測を何度も行うのなら、そのうちどのくらいの割合で失敗するかも想定できるので、それに対する備えをしておく、すなわち「**リスクを考える**」ことができる

人間の統計学と
機械学習の統計学

機械のための新しい統計学

統計学は、人間が集団の姿を把握するためのものだった

統計学(statistics)は、国家(state)と同語源

最近急速に進歩してきた機械学習は、コンピュータが集団の姿を把握する統計学

人間にわかるかどうかは別問題

コンピュータ棋士は、なぜその手を指すのか、人間にわかるようには教えてくれない

この講義では、人間のための、「伝統的な」統計学を扱います。

今日の最後に

思い込みにとらわれないための統計学



なぜベンチが

「線路に向かって座る」から

「列車の進む向きに座る」に変わったのだろうか？

転落事故56件を調査すると

うち33件(6割弱)は

こうではなく線路に向かって歩いて落ちていた

読売新聞2015. 3. 31



思い込みにとらわれず、

きちんとデータを調べよう