

2020年度春学期 統計学 第1回
 イン트로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



世界はいま、
 大混乱のさなかにあります

感染症と闘う統計学

感染症と闘う医学は、「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することはできない
 感染したかどうか、検査では完全にはわからない

そんなときに、「人々」、つまり集団を
 全体として見て、その姿を把握するのが**統計学**

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

「密閉・密集・密接の『三密』を避けよう」

では、「三密」を避ければ
 絶対感染しないのか？

そうではありません。
 とんりの人の吐き出したウイルスをたまたま
 吸って感染してしまうかもしれません。

感染症と闘う統計学

「密閉・密集・密接の『三密』を避けよう」

感染を必ず避けられるのではないのなら、
「三密回避」は何のため？

一度に多人数に感染させるのを防いで
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

一人の感染者がうつす人数が「平均して」一人以下
なら、社会全体の感染者数は減っていく

感染症と闘う統計学

「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様子を把握し、
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」

統計・確率的思考とは何か🤔

現在進行形の事象を扱うのは
むずかしいので、

日本🇯🇵が2011年に遭った危機を題材に

数量的思考 
微積分的思考 
統計・確率的思考 

数量的思考 

福島で原発事故！

沖縄で
480万ベクレルの
放射性ヨウ素！！

記事をよく読むと…

1平方キロメートルあたり
480万ベクレル

1平方メートルあたり
4.8ベクレル

人体👤にはすでに
4000ベクレル

バナナ🍌1本には
20ベクレル

数量的📊に考えましょう。

1910年, ハレー彗星接近

彗星の尾には青酸がある！

地球と彗星の尾が交差するので、
人類は青酸ガスで滅亡？！

実際には、彗星のガスは地球の大気よりもはるかに薄く、
ガスは地球には届きませんでした。

百年前の人を笑えないのでは？

微積分的思考 📈

世の中の数学に関する勘違いは
たいてい

割合と合計に関するもの
↓ ↓
微分 積分

原発近くで〇ミリシーベルト**毎時**の放射線を
検出

これは**1時間**浴び続けると
レントゲン写真△枚分の被曝に相当...

えっ、レントゲン△枚分の放射線を浴びたの🤯？

原発近くで〇ミリシーベルト**毎時**の放射線を検出

これは**1時間**浴び続けると
レントゲン写真△枚分の被曝に相当...

その量の放射線は一瞬出ただけ
かもしれないし、

そんなところに1時間立っているわけでもありません。

1時間あたり100ミリ
の雨🌧️ **雨の勢い**

〇時までの1時間に100ミリ
の雨🌧️ **実際に降った
雨の量**

この2つは大きく違います。

統計・確率的思考 🎲

放射線障害とは

放射線の粒子が持つ
エネルギーによって
遺伝子に「傷」がつく

傷がつくかどうかは、
偶然🎲による

傷がついても、
直って病気にならない
こともある

偶然 =
起きるかどうかが、人にはわからない
(人知が及ばない)

確率は
「起きやすさ」を言っているだけ

ある量の放射線を浴びた時、「安全か？」と
言われても困る。

確率を推定する

確率を推定する

ある量の放射線を浴びた時、病気になる**確率**が
どれだけ大きくなるのか？

ある量以下の放射線は安全、って
おかしいんじゃないの？😞

確率を推定する

データを集めて確率を推定するのは、簡単にいえば

くじびき🎰の結果から
当たり確率を推定すること

そんなこと、できる？

くじのあたり確率

「夏祭り、夜店のくじに当たりなし
露天商の男を逮捕」

(朝日新聞大阪版2013年7月29日)

記事によると

「1万円以上をつぎ込んだ男性が**不審に**
思い、府警に相談。28日に露店を家宅捜
索し、当たりがないことを確認した」

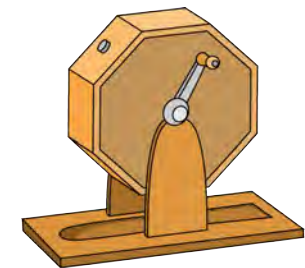
もう少し単純な問題で

半分当たるというくじ

「確率 $1/2$ で当たる」というくじを
10回ひいても、1回も当たらなかった

運が悪い🙵のか？

それとも
「確率 $1/2$ で当たる」とい
うのがウソ😬か？



こう考える

警察みたいに
全部のくじを調べられないなら、

仮に、本当に「確率1/2で当たる」とする

そのとき、10回ひいて1回も当たらない
確率は、 $(1/2)^{10}=1/1024$

こう考える

本当に「確率1/2で当たる」なら
10回ひいて1回も当たらない確率は
1/1024(約0.001)

それでも「確率1/2で当たる」を
信じるのは、

確率0.001でしか起きないことが、
いま目の前で起きていると信じるのと同じ

こう考える

確率0.001でしか起きないことが、
いま目の前で起きていると信じる

そりゃちょっと無理がありません
か？🤔

というわけで、
「確率1/2で当たる」はウソ、と
考えるほうが自然 と答える。

「仮説検定」という統計手法です

確率の計算

ところで、なぜ $(1/2)^{10}$ という
計算になるんですか？

「くじびき」というランダム現象を表す
確率分布モデルを考えているからです

ランダム現象とは、どういう結果になるかが偶然によって決
まる現象

確率分布モデルとは、どの結果になる確率がいくらになるか
を、各々の現象にあわせて数式で表現するもの

ところで放射線については

ある量以下の放射線は安全, っておかしいんじゃないの?

病気になった例(=当たりくじ)が数えるほどもない

「確率は1/2よりずっと小さい」とは言えても、「ではいくら?」かわからない

他の原因でも同じ病気にはなるから, 他の原因と区別がつかない

標本調査と統計的推測

統計的推測とは

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い」という話で

「ノルウェー人の平均身長は, 男179cm, 女170cmです」

ノルウェー人全員の身長を測ったの?

標本調査

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, 一部の人のみ(標本)を調べて, 分布全体のようなすがわかるか?

くじびきで調べる

一部の人だけ(標本)を調べて、
分布全体のようなすがわかるか？

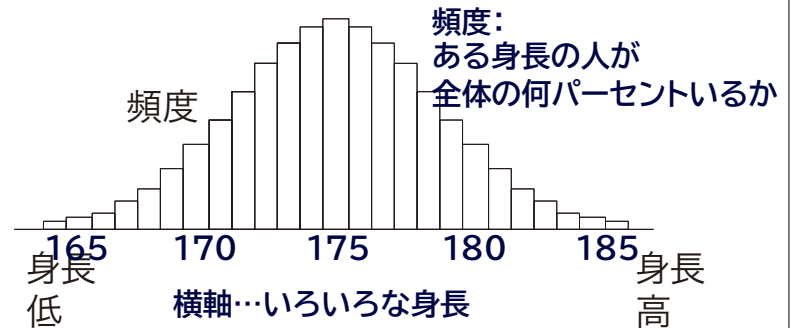
わかります。かなりの程度わかります。

「一部の人」を選ぶのに、
くじびきで選ぶ(無作為抽出)

無作為抽出すると

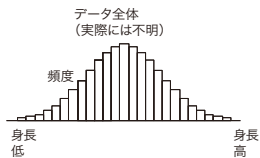
分布がこんなようすのとき

データ全体
(実際には不明)

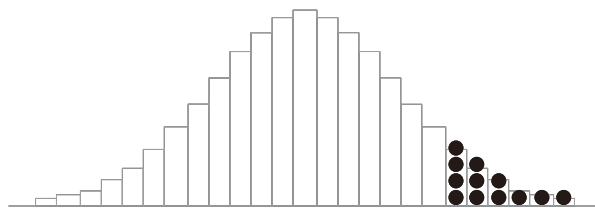


無作為抽出すると

偶然こんな標本(●)が
選ばれてしまう確率は
小さい

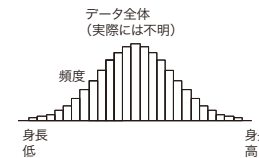


こんな標本が選ばれたら
→大きく偏った推測

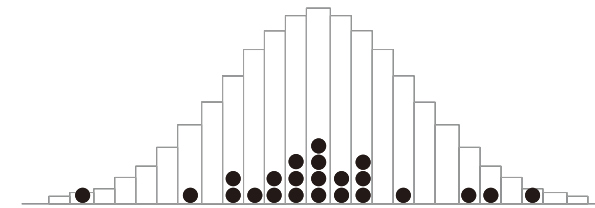


無作為抽出すると

たいていこんなふう
に選ばれる



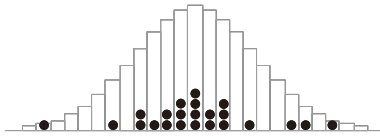
こんなふう
に標本が選ばれれば
→ほぼ間違っていない推測



無作為抽出すると

こんなふうに
標本が選ばれれば
→ほぼ間違っていない推測

たいていこんなふう
に選ばれる

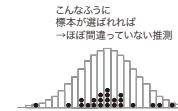


→標本(●)を平均すれば
データ全体の平均に
ほぼ近い

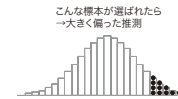
推測できた💡

「たいてい、ほぼ」

標本(●)を平均すれば
データ全体の平均にほぼ近い **のはいいとして...**



たいていこんなふう
に選ばれるだけであって、



こうなる確率はゼロで
はない**(大外し)**

いま取り出した標本が
ほぼ当たってるか、大外しかはわからない
(なぜなら、ノルウェー人全体は調べてないから。)

区間推定

いま取り出した標本が
ほぼ当たってるか、大外しかはわからない

わからないが、
「ある程度以上はずす確率」は計算できる

「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%」

このような答え方を、**区間推定**という

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

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

このクラスでは、人間のための、
「伝統的な」統計学を扱います。

今日のまとめ

統計学は 

集団の姿を把握する「俯瞰的視点」

確率的思考 

偶然に起きるできごとの「起きやすさ」を考える

統計学と確率

データ全体を調べきれない場合、

くじびきで選んで調べる

→調べた結果が当たっているかどうかも

偶然によるが、当たる確率を述べる

今日の最後に



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件は
こうではなく
線路に向かって歩いて
落ちていた

思い込みにとらわれず
きちんと調べよう