

2020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



世界はいまだ、
大混乱のさなかにあります

感染症と闘う統計学

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、完全にはわからない🦠

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

「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。
となりの人の吐き出した🦠をたまたま吸って、感染してしまうかもしれません。

「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「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（実効再生産数が1未満）

「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」

あなた👤👤を救うのではなく、社会全体🇯🇵🇯🇵を救う

統計・確率的思考とは何か🤔

現在進行形で
まだ決着のついていない事象🌱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 📊

2011年3月の原発事故のとき

福島で原発事故



遠く離れた沖縄で、480万ベクレルの放射性ヨウ素が！ 🤦

という新聞記事がありました。

記事をよく読むと…

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

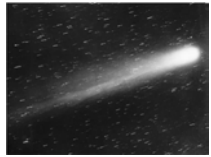
1平方メートルあたり4.8ベクレル…

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!

http://upload.wikimedia.org/wikipedia/commons/2/b29/Halley%27s_Comet_-_May_29_1910.jpg



<http://www.poptar.co.jp/shop/shosai.php?shosekicode=30200240>

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, たいいてい



やっぱり, 2011年3月の原発事故のとき

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

という記事を読んで

えっ, レントゲン△枚分の放射線を浴びたの? 🤔

と言った人が多数…

落ちついてください

原発近くで0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。

天気予報で考えると

1時間あたり100ミリの雨☔

雨の勢い

○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、そもそも何か

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**



http://upload.wikimedia.org/wikipedia/commons/8/81/ADN_animation.gif

偶然とはなにか

偶然 =
起きるかどうかわからない(人知が及ばない)

確率は、「起きやすさ」を言っているだけ

だから、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」

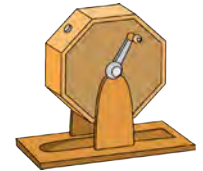
もう少し単純な問題で

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, $(1/2)^{10} = 1/1024$ (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

ところで、放射線の話は

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

「確率はとても小さい」とは言えても、
確率の推定はきわめて不確か

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
他の原因を心配するほうがよい

標本調査と統計的推測

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、一部の人のみ(標本)を調べて、
分布全体のようなすがわかるのか？

わかります。かなりの程度わかります。

「一部の人のみ」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、たいていはいろんな人がまんべんなく選ばれる

区間推定

くじびきで選べば、**たいてい**はいろんな人がまんべんなく選ばれる

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。

「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

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？



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

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