

統計という「言語」の便利な使い方

日時： 3月11日（金） 10:00～17:15

場所： 東京大学（本郷キャンパス）赤門総合研究棟5階・センター会議室（549号室）

料金： 一般 5,000円、学生 2,500円

講師： 佐々木弾（東京大学）

■本コースの内容

統計学を数学として（だけ）でなく、より人間的・社会的な意思疎通手段として捉え直します。道具や機械は人間生活に不可欠のものですが、それらは単独ではなく、それらを駆使する人間の手があって初めて役に立ちます。俗に言う「鉄は使いよう」です。統計についても同じで、統計は使いようによります。

情報化社会と言われて既に久しい現今ですが、その「情報」とは具体的にどのようなものでしょうか？ 情報化以前と比較して、どのようなタイプの情報が増産されているのでしょうか？ その典型である「統計」に本講では着目し、統計の持つ「言語」としての役割、即ちコミュニケーション機能や論理的思考ツールとしての有用性とその将来性を平易にご紹介し、併せて現代および近未来における統計の意外にも便利な使い方について参加者の皆様とご一緒に考察していきます。（詳細は別添「コース内容の概要」参照。）

■次のような方におすすめです

- ・高校～大学学部～大学院で統計の授業をつまらなく感じていた方。
- ・統計の数学的・機械的な側面だけでなく、その人間的・社会的な面を知りたい方。
- ・理系から人文社会系へ研究分野を拡大・転進したい方。
- ・数学や計算に自信（または興味）が湧かず、統計学に馴染めずにいる方。
- ・統計学を勉強してから年数が経ち、時代に遅れたと感じている方。
- ・統計学をいくら勉強しても、指導教員や上司・同僚に「もっと勉強しろ」と叱られる方。
- ・統計学の教本を何冊も読んだが、そこに書いてない何かがある、と不満を感じる方。
- ・統計学を面白く教えたり書いたりしたくて、ネタを探している方。

■注意事項

- ・本講だけの受講も可能です。受講の前提として他コース（3月1～10日開催）への参加が条件ではありません。
- ・受講者へは、詳細な講義資料を事前にお渡しします。各自が資料をダウンロードして予習の上、必要に応じ講義当日ご持参ください。当日は、資料ハードコピーの配布はいたしません。
- ・資料の入手方法については、受講者へ別途お知らせします。

■コース内容の概要

統計セミナーを通して統計（学）の技術的側面を修得された皆さんの次なるステップは、統計学を数学として（だけ）でなく、より人間的・社会的な意思疎通手段として捉え直すことです。道具や機械は人間生活に不可欠のものですが、それらは単独ではなく、それらを駆使する人間の手があって初めて役に立ちます。俗に言う「鉄は使いよう」です。統計についても同じで、統計は使いようによります。

情報化社会と言われて既に久しい現今ですが、その「情報」とは具体的にどのようなものでしょうか？ 情報化以前と比較して、どのようなタイプの情報が増産されているのでしょうか？ その典型である「統計」に本講では着目し、統計の持つ「言語」としての役割、即ちコミュニケーション機能や論理的思考ツールとしての有用性とその将来性を平易にご紹介し、併せて現代および近未来における統計の意外にも便利な使い方について参加者の皆さんとご一緒に考察していきます。

情報化社会の初期段階では、狭い意味での計算機の発達により、旧来長らく「読み書き」と並んで国民的基礎教養とされた「計算」が機械に取って代られ、人の手を離れました。しかし周知の通り、このことは決して人々が数学を学ばなくてよくなったことを意味しません。単純作業としての「そろばん」から開放された人々は、より概念的・論理的に高度な数学に専念できるようになったわけです。この恩恵を最も多大に享受した分野の一つが統計であり、情報化により特に統計情報が質量ともに突出した成長を遂げている所以でもあります。

他方、いわゆる情報化により却って存在感の薄れた種類の情報というのも、無いわけではありません。俗に「出版不況」などと評される昨今の社会では、旧来的な意味での書籍類、即ち日常言語による文章情報の比重が趨勢的に縮減していることがわかります。加えて技術面の情報化、即ち情報技術の革新は、人々の使用言語の盛衰も促します。太陽の没することなき大英帝国の世界展開以来、長らく世界標準語としての地位を誇ってきた英語は、我が国を含む世界各地で産官学界全ての分野で謂わば最重要必修科目と考えられてきました。ところが機械による自動翻訳機能の飛躍的進化により、近年中にこの状況が一変する可能性が頓に現実味を帯びてきました。ちょうど「そろばん」が衰退したのと同様、単純作業としての読み書きから人の手が離れ、機械で代用できないようなより高度な論理的思考や、すぐれて社会的・非機械的な目的を有するコミュニケーションへと人的作業がシフトする兆候は、既に現れはじめています。

優れた共通言語を持つ社会が繁栄することは、人類史上繰返し証明されてきた一般則です。19世紀、文明開化した日本は「国語元年」で近代化に成功しました。20世紀の世界は、英語で栄えました。21世紀の共通言語は、統計です。言語にはそれぞれ、その言語に特有な論理があり、哲学があり、世界観があります。統計も決して例外ではありません。それを駆使できるインテリジェンスを具えた「選民」の皆さんが、今世紀の世界を支配し、繁栄へと導くことでしょう。

では一体、統計とはどういう「言語」なのでしょう？ 以下に、本講の概容をご紹介します。

(0) 統計事始め [今更聞けない！編]

降水確率 0%って言った日でも、実際には正の確率で雨が降る。天気予報は嘘つき？
期待年利が $\pm 0\%$ の資産に長期投資すると、最終的に得も損もしない？
平均年収やら平均給与やらの数字、何となくウチより高め。ウチ、そんなに薄給？
子供たちの数学の成績と英語の成績との間には強い正相関。どっちがどっちの役に立ってるの？
確率・統計には「ベイジアン」と「頻度論」という 2 大学派。どう違うの？
...等々、高校や大学の講義で「うっかり聞き逃した」「昔聞いたけど忘れた」間違えやすい基礎概念を復習します。

(1) 統計を読む・聞く [リテラシー編]

データの出所は？ 誰がどういう目的で発表したもの？
データの性質は？ 収集・編集の方法は？
分析の方法は？ データの性質や研究目的に適ったものですか？
説明（独立、外生）変数と被説明（従属、内生）変数との区別は妥当？ 同時性や逆因果は？
相関と因果とを混同・同一視してませんか？
得られた結論は（ここが大事！）直観的に説明でき、常識的に納得できる？
...単純に計算が合ってるかだけでなく、統計を社会的メッセージとして読解する練習です。

(2) 統計を書く・話す [コミュニケーション編]

いよいよ自分から統計を発信する練習です。本来であれば、
〔甲〕或る決まった研究目的でデータを探し、利用可能な範囲のデータで出来るだけの分析を行う、
〔乙〕或る与えられたデータに基づき、可能な範囲で分析を行い結果を導出する、
の何れかを行うのが、真っ正直な統計分析の正攻法に他なりません。しかし世には
〔似非甲〕或る決まった結論を導くべく片っ端からデータを探し、都合の悪いデータは全部捨てる、
〔似非乙〕偶々データが手に入ったというだけの理由で分析を始め、研究目的や分析結果の意味は不明、
というような「偽物」が後を絶ちません。どのような統計分析が「望ましい」か、を皆さんと一緒に科学的に論ずる思考実験です。

以上、本講の前半は謂わば「統計分析とは斯くあるべし」といった総論を扱います。しかしトルストイが「幸福な家庭は一通りだが、不幸な家庭は幾通りもある」と言ったように、望ましい統計分析が一つある毎に、その四隣には望ましくないものが散在しています。本講後半ではそうした謂わば「反面教師」に焦点を当てます。統計も言語である以上、「綺麗な言葉」と「汚い言葉」、「正しい言葉」と「誤用された言葉」があります。そのような「間違い探し」の術まで修得して初めて、真の統計通と言えるでしょう。

(3) 統計で騙す [詐欺・偽装編]

うっかりか、わざとか、は別として、誤解を招く統計が世に溢れていることは周知の事実です。しかしながら、この誤解を招くという性質が統計というものの本質と実は表裏一体であることは、そこまで周知ではありません。周知の進まない理由の一端は、統計と言うと何かしら自然現象そのもの、或いはその機械的な描写であるかのように錯覚される場合が多いことにあります。統計とはあくまで、人が編集し分析したものです。人の手が加わっていますから、それは最早、単純な自然現象や物理現象ではなく、作品であり製品であるわけです。そこには何らかの「意図」が必ず存在します。統計を発信する者のそうした意図が、それを受信する者の理解とぴったり一致しない限り、多少とも誤解が発生するわけです。そのような意図（しばしばその存在自体も）を隠し、統計を恰も客観的事実そのものであるかのように「偽装」という手口は、統計の最もポピュラーな悪用法(?)の一つです。よく見ると本論や議題とはあまり直接的に関係が無いにも拘らず何らかの統計(らしきもの)を敢えて呈示することで見せ掛けの説得力を膨らませる、という手法がその典型例で、皆さんも幾度となく目にされてきたことでしょう。

人為的な操作の可能性は、データの作成・編集に始まり、分析手法の選定やモデルの作り方、変数の取捨選択、結果の解釈に到るまで、統計分析手順の全ての段階に潜んでいます。本講では時間の許す限り、統計分析の発信者・受信者双方の立場に立ちつつ、これら「人災」の諸々について解説を試みます。

(4) 統計で盗む [窃盗・横領編]

ただ単に誤解を招く、或いは自ら曲解する、というだけでは飽き足らず、初めから手前味噌な結論を導く目的で統計分析を始める、という事例は、世間一般に考えられているよりも多いと言えます。こう言うと、どう鼻屑目に見ても自己中心的であり、不法で邪道な行為のように聞こえるのですが、現行の法規制や倫理規程には引っ掛からないものが多いこともまた事実です。何故ご法度に触れないか、と言うと、積極的に「嘘」をつかない限り、罪には問われないからです。

ではその「消極的な嘘」とは何か？ 最も典型的なのは、都合の悪い事は言わない、書かない、という手です。統計学関係者の間ではよく、(有意な結果が出ないと論文として売れないので) 偶々t値が1.64以下だった研究は発表されず、1.65を少しでも超えたものだけめでたく発表、という笑えない冗談が囁かれます。何を発表し、何をボツにするかは、発表者の恣意に委ねられている現状です。

しかし定義によって、 $t=1.65$ で5%有意というのは、およそ何の関係も無いデタラメな変数を持ってきた場合に、ほぼ100回中5回の割合で $t=1.65$ 以上の値が偶々出る、ということに他なりません。ですからヒマとカネさえ贅沢に持っていれば、好き勝手な「統計分析」を性懲りも無く100回繰返し、そのうち偶々有意だった5回分だけを発表すれば、好き勝手な「結果」を統計的に証明できてしまうわけです。残りの95回は捨てて、口を拭えばいいわけです。自己利益になりそうな「結論」を勝手に導いてしまうわけですから、これこそ「嘘つきは泥棒の始まり」です。

それでは、こういう消極的な嘘はどうやって見破れば宜しいのか？ 自分で研究発表を行う

場合、こういう嘘を疑われないためにはどうすればよいのか？ 本講では皆さんと一緒に考えます。

(5) 統計で迫害する [中傷・虐待編]

扇動的な政治家や評論家たちは、決まって統計が大好きです。統計を悪用して政敵を中傷し、偏見や差別を煽ります。戦争中の各国政府が、それぞれ敵国民の知能の低さや健康・衛生状態の貧困さなどに関する統計的な証明合戦を展開したことは、周知の通りです。現代では、右翼や国粋政治家たちによる次のような言説を耳にされたことのある皆さんは少なくないでしょう。

「少年による凶悪犯罪が増加している。少年法改正による厳罰化が必要。」

「外国人犯罪が増加している。入国管理の厳格化が必要。」

「離婚が増加している。夫婦別姓に反対。」(←最高裁)

このような謂わば統計的ヘイトスピーチの数々は、いずれも特定集団への明確な悪意を以て発信されるものであり、この点で既に価値自由な科学とは根本的に相容れません。逆に、科学者であるべき研究者の皆さんの側から言えば、

「自らこのようなヘイトスピーチを発信しない」

「ヘイトスピーチに騙されたり、うっかり相槌を打ったりしない」

「ヘイトスピーチを見つけたら、科学的に正しい方法で反駁する」

「ヘイトスピーチのどこがどう非科学的か、一般人に解り易く説明する」

という努力が大切です。本講では、不偏不党で価値自由な統計とはどのようなもの（であるべき）か、を皆さんと一緒に考察します。誤った統計は、時として思わぬ社会対立や戦争などの惨禍の因となり得る危険極まりないものであることを、我々一人一人が再認識する必要があります。

(6) 統計で殺す [殺人・環境破壊編]

どう考えても常識的にあり得ない言説を、何故か統計により「証明」しようとする試みは、後を絶ちません。これらの愚行の中には、時として非常に多数の人命に関わるような甚大な害悪を及ぼすものもあります。

「二酸化炭素は地球温暖化等の環境変化を惹起しない」

「肉食中心の食事は不健康的ではない」

「原発の事故確率はゼロ」

「タバコは肺癌の原因にならない」

「喫煙者が非喫煙者より不健康ということはない」

明らかに嘘だと判っていて、信じてもらえないことも判りきつていながら何故、このような嘘を「統計的に証明」しようとしたり、したフリをしたりするのでしょうか？ 可能性としては

「発信者の精神的健康状態や認知能力に問題がある」

「特定利害に基づく情報戦」

「通説にわざと異を唱えてみせるキワモノ、いわゆる炎上商法」

等々が考えられます。しかしこれは逆に見れば、統計がいかに完成度の高い、汎用性の広い言

語であるかを端的に示しています。よく「統計は正直だ」などとナイーブに思い込んでいるお人よしがいますが、実は日常言語に負けず劣らず、いっばしの大嘘をまことしやかに語れるだけの語彙と文法の豊かさを具えているのが統計なのです。

人の安全に関わる有害な嘘は論外としても、本講では統計のもつこうした言語としての豊かさに着眼し、場合によっては「嘘も方便」的な使い方も可能であることなど、通常の教科書や講義では扱い難い側面にも、時間の許す範囲で言及します。