統計学におけるモデルの役割について

大提访倒凝提集明眼

直感・説明・検証

晰役

三中 信宏

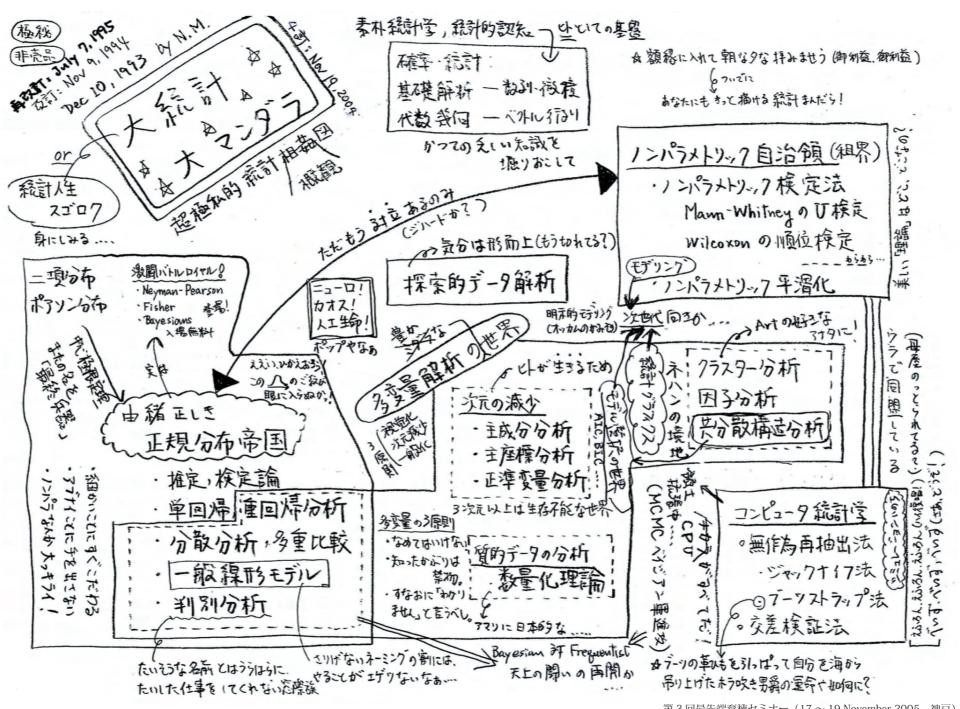
MINAKA Nobuhiro

国立研究開発法人 農研機構 農業環境変動研究センター [系統体系学] 東京大学大学院 農学生命科学研究科 生物・環境工学専攻 [生態系計測学] 東京農業大学大学院 農学研究科 [応用昆虫学]

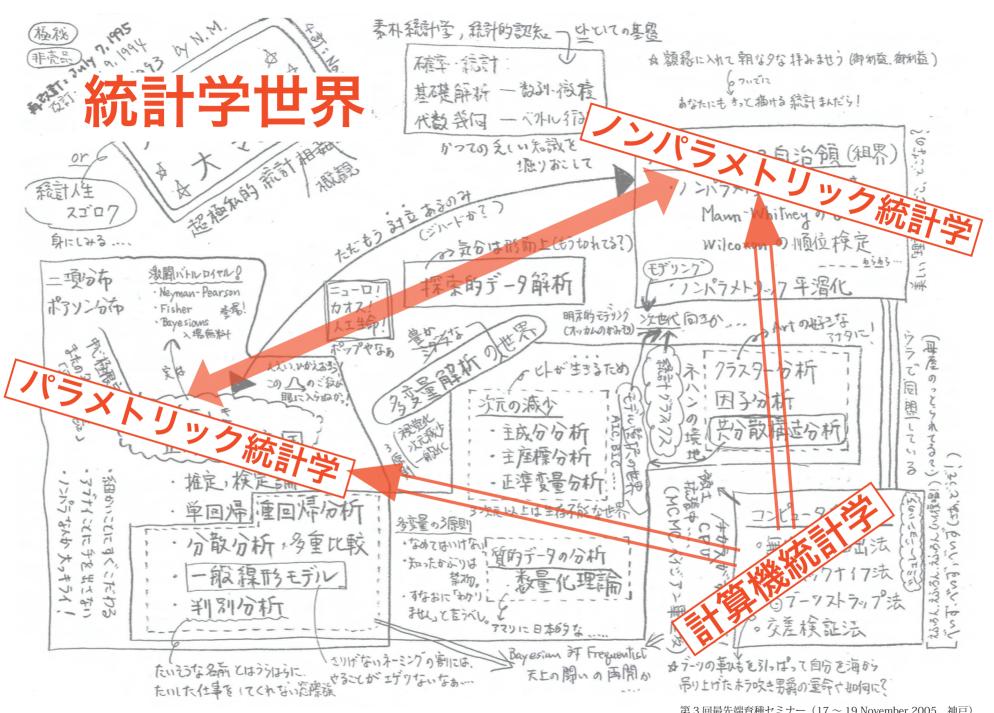
mail: minaka@affrc.go.jp

website: http://leeswijzer.org/

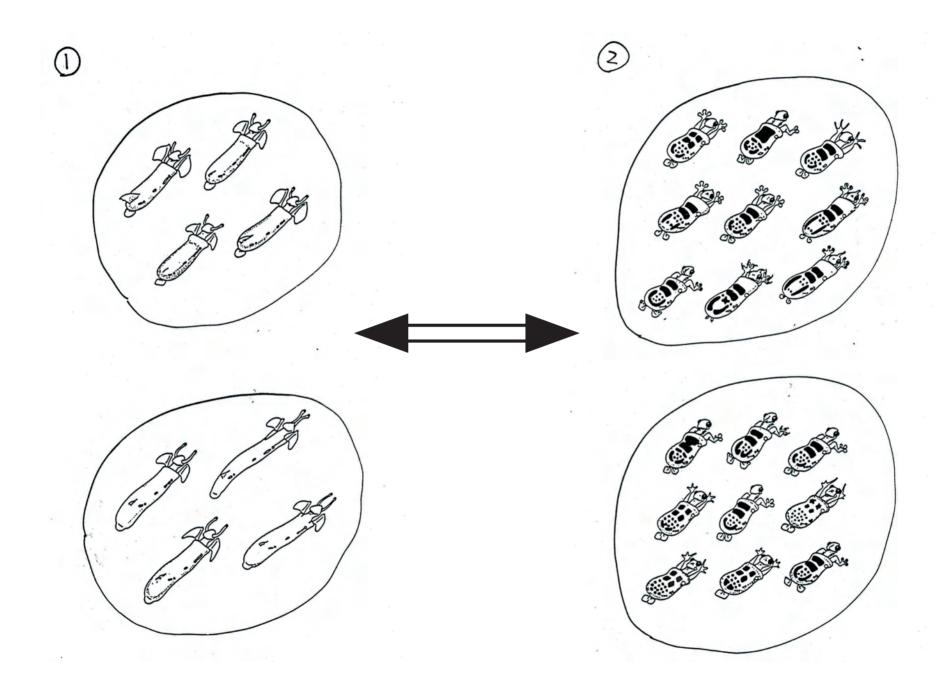
twitter: @leeswijzer



第3回最先端育種セミナー(17~19 November 2005,神戸) 〈家畜・水産動物における遺伝的多様性研究―その目的と新しい方法論―〉 三中信宏:生物進化の歴史を推定する―分子系統樹を構築する理論と方法



第3回最先端育種セミナー(17~19 November 2005,神戸) 〈家畜・水産動物における遺伝的多様性研究―その目的と新しい方法論―〉 三中信宏:生物進化の歴史を推定する―分子系統樹を構築する理論と方法



ヒトの素朴統計学的「直感」

いつ「差がある」と感じるか?



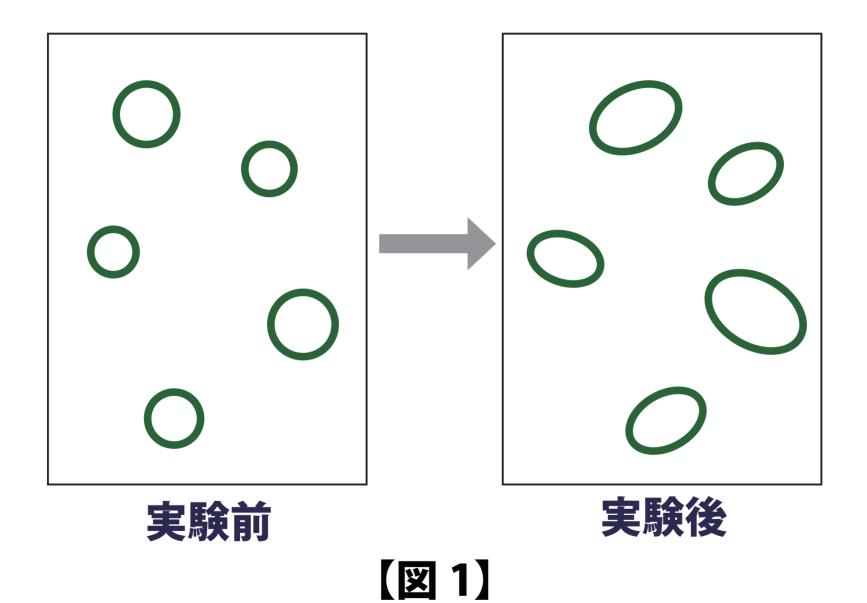
1

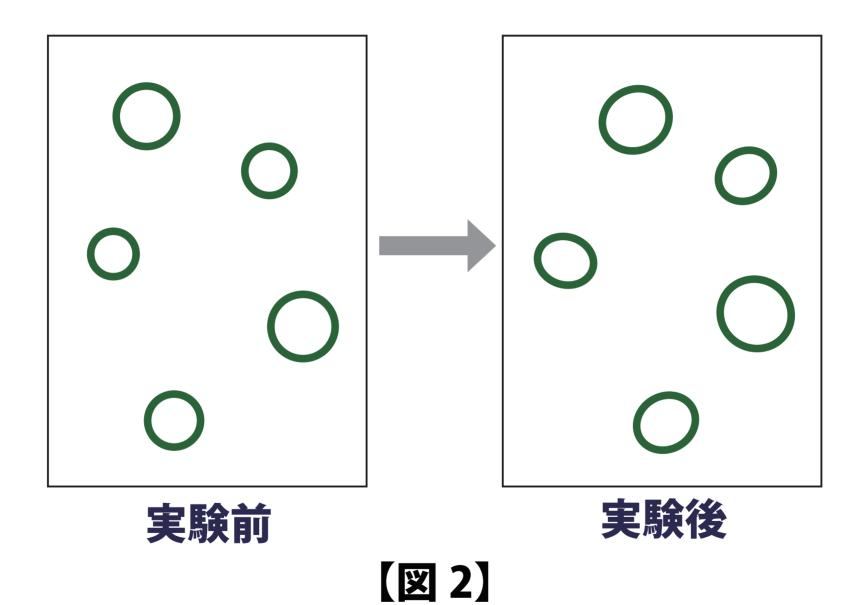
「群間のちがい/群内のちがい」が大きいとき → 「差がある」と感じる

○ケース2)

「群間のちがい/群内のちがい」が小さいとき →「差がない」と感じる

※判定基準となる比率=「群間のちがい/群内のちがい」 を定量化すると、統計学でいう「F値」になる.





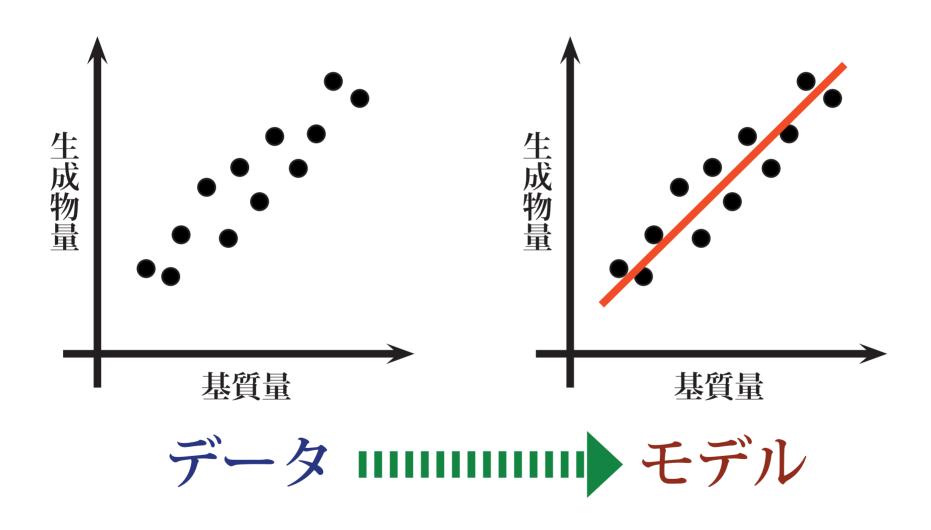
データ解析の"心"とは?

- 1
- ・目に見えること
- ・じっと眺めること
- ・視点を変えること

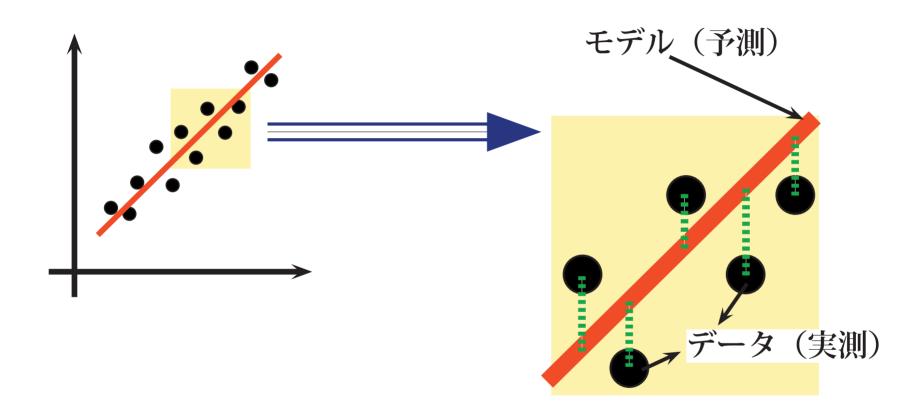
どのように「視覚化」するか

実験データをいきなり 「数式」でいじるのではなく、 まずはじめに 「絵」を描いてみる。

データをモデルで説明する



データをモデルで説明する



モデルはデータから「ずれる」ものである

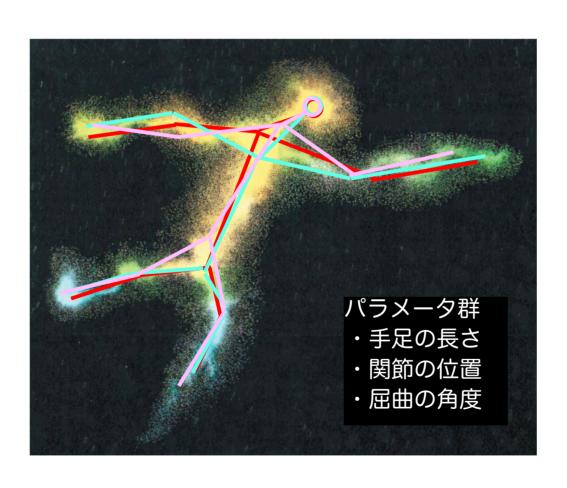
パラメトリック統計学の出発点

データのもつ「誤差」の性質は,

「誤差関数」によって数式表現できる.

線形統計モデル(LM)への入り口

モデルの構造とパラメータ



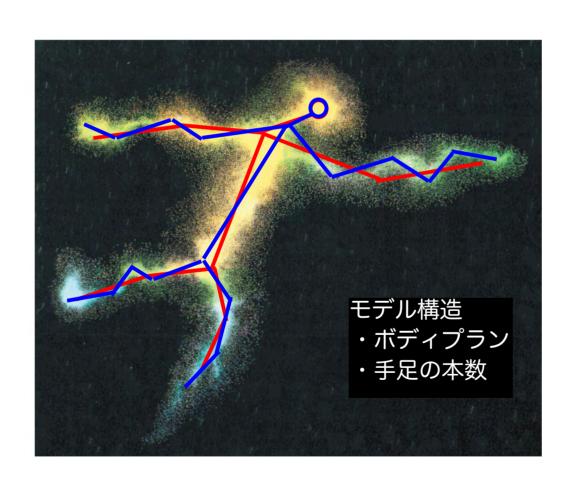
パラメータ推定問題

モデルの「構造」を固定 したとしても、パラメー タ群の値は可変である. どのような基準で最適な パラメータ値を決定する のか.

計算すればなんとかなる

線形統計モデル(LM)への入り口

モデルの構造とパラメータ

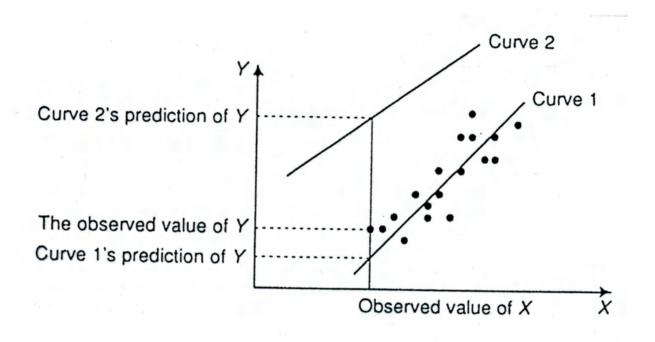


モデル選択問題

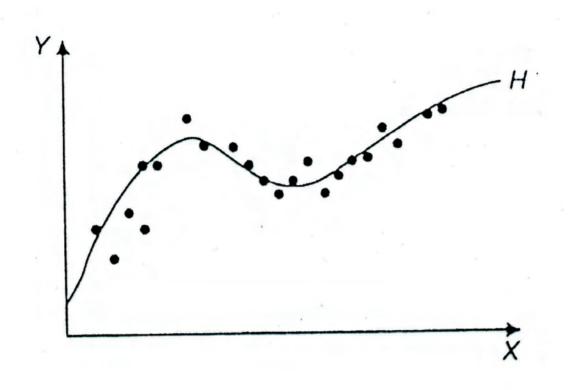
モデルの基本となる「構造」そのものをどのよう な基準で決定すればいい のか.

やや難しいかも・・・

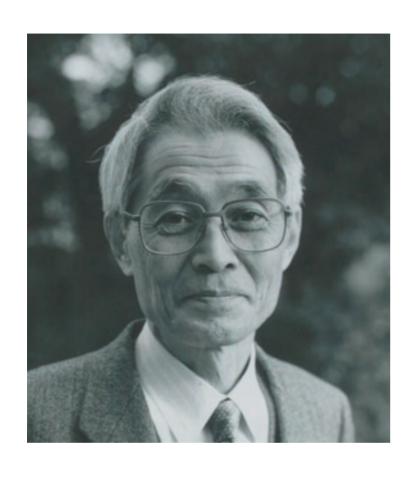
Curve-fitting problem (2)



Curve-fitting problem (3)



複雑なモデルは「最良の説明」 を必ずしも与えないだろう.単 純なモデルの方がこの点で勝っ ている可能性がある.

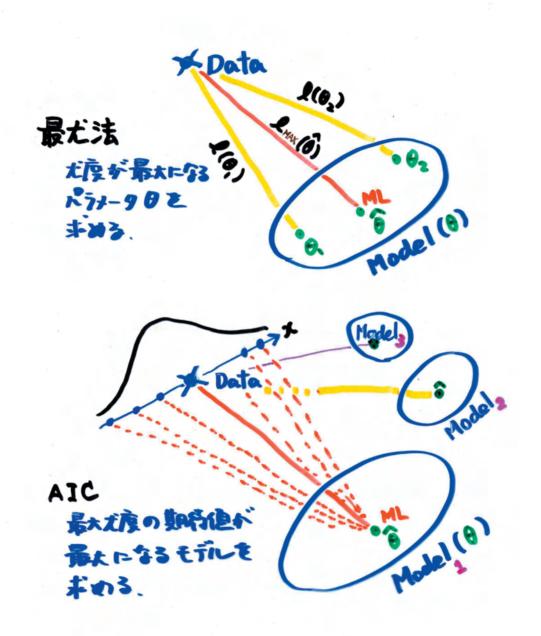


良いモデルは 「予測確度」 (predictive accuracy) が大きい

[赤池弘次]

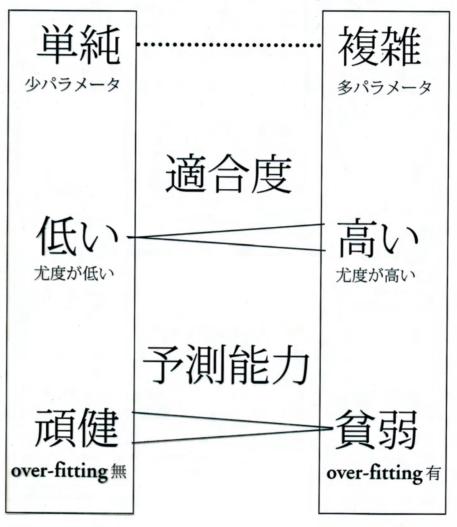
AIC

(赤池情報量基準)



最尤法 期待 期待 尤度 **AIC**

トレード・オフ



モデルの予測確度

期待尤度ーパラメータ数

(適合度)

(複雑度)

AIC

(赤池情報量基準)

『心理学評論』特集号(2016)



目次(原著論文)

巻頭言

原著論文

- 池田功毅・平石界心理学における再現可能性危機:問題の構造、現状と解決策
- 山田祐樹 認知心理学における再現可能性の認知心理学
- 森口佑介発達科学が発達科学であるために:発達研究における 再現性と頑健性
- 鮫島和行システム神経科学における再現可能性
- 澤幸祐・栗原彬 動物心理学における再現可能性の問題
- 大久保街亜帰無仮説検定と再現可能性
- 小塩真司 心理尺度構成における再検査信頼性係数の評価—「心理学研究」に掲載された文献のメタ分析から—
- 藤島喜嗣・樋口匡貴 社会心理学における"p-hacking"の実践例
- 渡邊芳之心理学のデータと再現可能性

コメント論文

- 小島康生 人間の観察研究における再現可能性の問題
- 松田一希フィールド研究の再現性とは何か?
- 平井啓 心理学研究におけるリサーチデザインの理想
- 三中信宏統計学の現場は一枚岩ではない
- 武田美亜 再現可能性の問題から始める心理学研究の「バック ヤードツアー」
- 東島仁 研究公正から見た再現可能性問題
- 佐倉統 科学的方法の多元性を擁護する

「統計学的有意性を再定義する」(2017)

Science Home News Journals Topics Careers

SHARE



It will be much harder to call new findings 'significant' if this team gets its way



By Kelly Servick | Jul. 25, 2017, 2:30 PM



31



500

A megateam of reproducibility-minded scientists is renewing a controversial proposal to raise the standard for statistical significance in research studies. They want researchers to dump the long-standing use of a probability value (p-value) of less than 0.05 as the gold standard for significant results, and replace it with the much stiffer p-value threshold of 0.005.

Backers of the change, which has been floated before, say it could dramatically reduce the reporting of false-positive results—studies that claim to find an effect when there is none—and so make more studies reproducible. And they note that researchers in some fields, including genome analysis, have already made a similar switch with beneficial results.

「統計学的有意性を再定義する」(2017)



nature.com

NEWS • 26 JULY 2017

Big names in statistics want to shake up muchmaligned P value

One of scientists' favourite statistics — the P value — should face tougher standards, say leading researchers.

Dalmeet Singh Chawla

Science is in the throes of a reproducibility crisis, and researchers, funders and publishers are increasingly worried that the scholarly literature is littered with unreliable results. Now, a group of 72 prominent researchers is targeting what they say is one cause of the problem: weak statistical standards of evidence for claiming new discoveries.

SUBJECTS

PUBLISHING

仮説検定史

• Pure Significance Test (R. A. Fisher 1925)

ある仮説 H のデータ x に対する「整合性」の程度を調べる. 有意確率: $p = \Pr[t(X)>t(x)/H]$ が小さいほど整合性が低い. 仮説の棄却・受容に関する意思決定ではない.

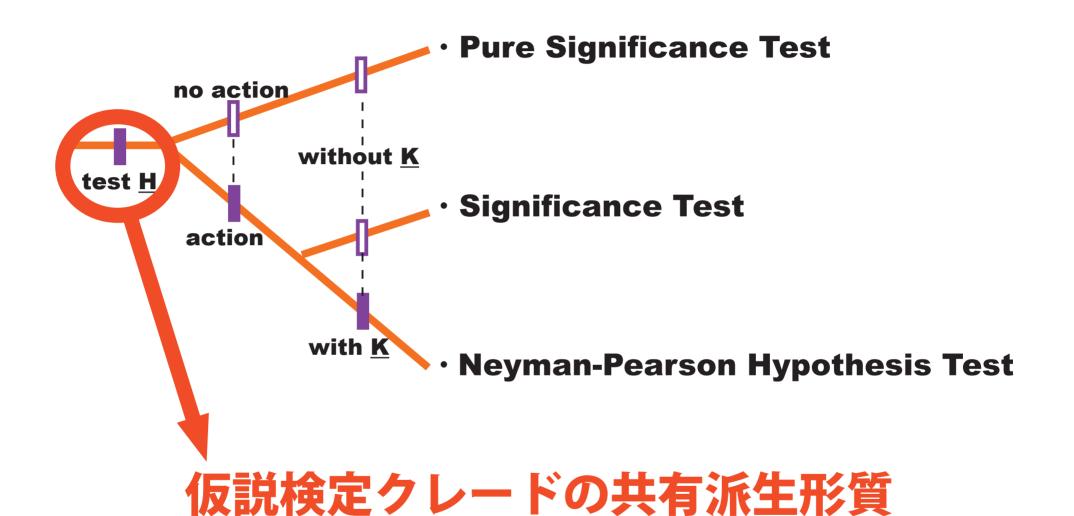
Significance Test

ある「有意水準」 α のもとで、仮説 H の棄却・受容を意思決定する. % 「p< α のとき、H を棄却する」という行動基準を置く.

Neyman-Pearson Hypothesis Test

(J. Neyman & E. S. Pearson 1933)

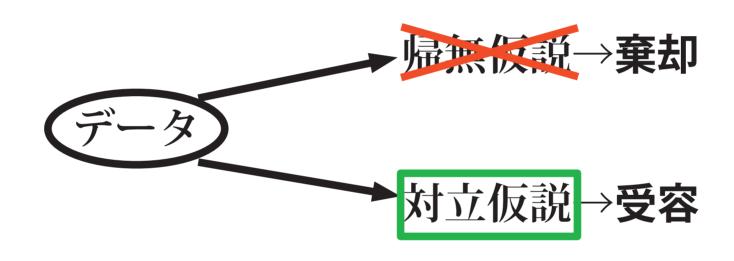
ある帰無仮説 H に対して、対立仮説 K を設定し、有意水準 α の もとでの仮説を棄却・受容するための検定を構築する. 尤度比の 導入. 検定力 (power) の重視などなど.



Neyman-Pearson 的に「仮説をテストする」ためには:

- 1) モデルを指定する.
- 2) そのモデルに含まれるパラメータθに関する帰 無仮説 H と対立仮説 K を設定する.
- 3) 有意水準(type I error) α を指定する.
- 4) 検定の手続きをつくる. 尤度比と検定力に着目.
- 5) 以上のすべてをデータ収集に先立って行なう.

「意思決定パラダイム」の中の統計学



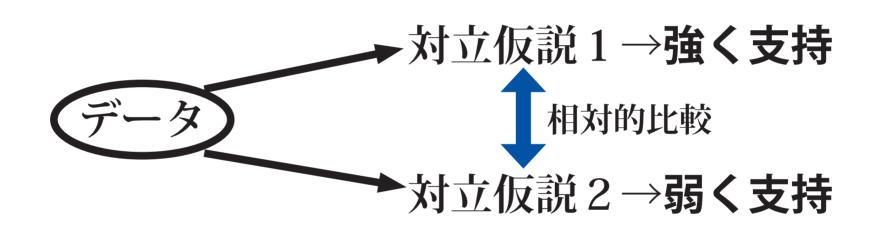
Data as Executioner

テスト対象である仮説の 占めるべき「地位」が 変わりつあるのではないか?

絶対ランキング→相対ランキング

(true or false) (better or worse)

「法科学的パラダイム」



Data as Evidence

Elliott Sober

観察されたデータから,ある仮説が 真であること(強確証),または偽 であること(強反証)を結論するの は現実的ではない.むしろ,データ がそれぞれの対立仮説をどれくらい 支持するか(弱確証/弱反証)を比 較する必要がある.

[『過去を復元する』1996, 蒼樹書房]





Richard Royall

どのような状況で、与えられた観察が、ある統計的仮説ではなく、別の仮説を支持すると言えるのか.この根本問題を統計学は長い間無視し続けてきた.それは、Neyman-Pearson 統計学が、データを仮説の証拠とみなすのではなく、意思決定の理論として定式化されてきたからである.

[『Statistical Evidence: A Likelihood Paradigm』 1997, Chapman & Hall]

非演繹的推論のもつ多様なあり方

