

PHC Curve による 統計教育の教程によせて

豊田秀樹

早稲田大学文学学術院



PHC: probability that the research hypothesis is correct
https://wcms.waseda.jp/em/67afc5d4c836d?authkey=1739576571669&user_id=6010107&lg=ja

講演内容

- 現在の標準的な有意性検定を柱とした統計学の入門教程は学問の発展にとって望ましくない、と講演者は考えている
- 批判だけでなく、オリジナルの入門教程を提案
- 本講演では、PHC による統計教育の教程を紹介

講演内容の詳細については以下の文献をご参照ください

文献[1] 統計学入門 I: 生成量による実感に即したデータ分析 2022 朝倉書店 ISBN978-4254122664

文献[2] 統計学入門 II: 尤度によるデータ生成過程の表現 2022 朝倉書店 ISBN978-4254122725

文献[3] 瀕死の統計学を救え! —有意性検定から「仮説が正しい確率」へ— 2020 朝倉書店 ISBN978-4254122558



抗原検査： 標準的な治療法がいまだ確立していない新型コロナウイルスに対する治療薬の開発を行っている。この病気は、幸いにも重症化しなかった場合には、発症から起算して平均 15.0 日で抗原検査の結果が陰性になる。この平均日数は膨大なケースから計算され、値が安定している。治験に参加してくれた患者に新薬 A を投与し、半日ごとに抗原検査を行った。結果が陰性になるまでの日数が表 1.1 である。この新薬には効果があるといえるだろうか。

表 1.1 測定結果 (日数)

1	2	3	4	5	6	7	8	9	10
11.5	14.0	15.0	10.0	14.0	14.5	12.5	12.5	12.5	12.0
11	12	13	14	15	16	17	18	19	20
13.0	12.0	12.5	13.5	15.0	13.5	12.5	12.0	12.0	17.0

- 新薬Aを投与された患者が陰性になるまでの平均日数は15日より有意に短い。
- 統計的検定の結果、 $p < 0.05$ であり、有意な差が認められた。

表 5.1 $\text{phc}(\mu \leq c)$ の phc テーブル (新薬 A)

$c(\text{日})$	11.5	12	12.5	13	13.5	14	14.5	15
$\text{phc}(\mu \leq c)$	0.000	0.004	0.061	0.419	0.876	0.991	1.000	1.000

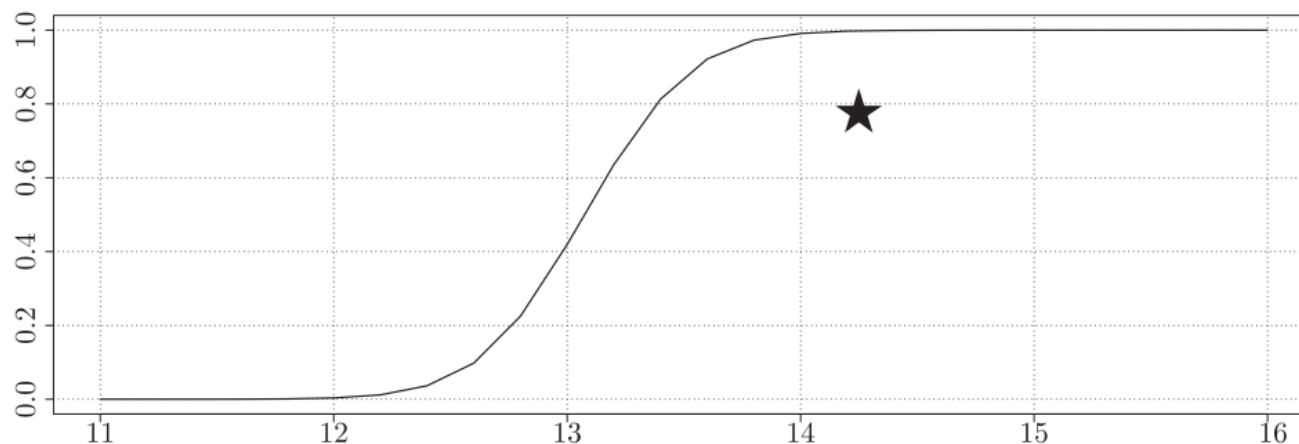


図 5.2 $\text{phc}(\mu \leq c)$ の phc 曲線 (新薬 A) $n=20$

- 「陰性になるまでの平均日数は14日より短い」
という主張は99.1%正しい。99.1%の確信をもてる。
- 「陰性になるまでの平均日数は12日より短い」
という主張は99.6%間違っている。0.4%の確信しか持てない

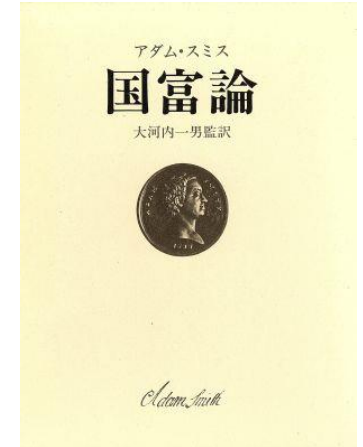
PHCは、考え方が自然である

- 有意性検定の論法
 - 陰性になるまでの期間が旧薬と同じと仮定すると、検定統計量が、これ以上甚だしい値になる確率は5%以下である。
 - 複雑で統計学の普及を妨げている。理解が難しいのでクリアが目的化
- PHCの論法
 - 「陰性になるまでの平均日数は14日より短い」
という主張は99.1%正しい。99.1%の確信をもてる。
 - 言葉のまま、日常会話である。素直で小学生にだって理解できる。

有意差は学問発展の必要条件に過ぎない

- 「新薬を服用した患者が回復するまでの平均日数は対照群と同じ」という帰無仮説を棄却できたとしても、それは新薬に有効性があるための必要条件を確認しただけに過ぎない。
- 有意性検定の大きな罪は「学問発展のための必要条件を確認すれば論文は査読を通る」という誤った文化を定着させたこと。
- 必要条件をクリアしているけれど、学問発展に寄与しない論文をあの手この手で掲載。Pハッキングと総称

神の見えざる手



- 価格が安いと売れやすい（ n が大きいと有意になりやすい）。
- 価格が高いと売れにくい（ n が小さいと有意になりにくい）。
- 高品質なら売れやすい（学問的に価値があるなら、もちろん有意にもなりやすい）。
- しかし高品質商品は品薄で仕入れが困難
（しかし価値のある発見は、創造的偉業であり、極めて高コストで実現が難しい）。
- 価格は売り手が決められる（観測対象の数 n は研究者が決められる）。
- 商人（研究者）としての生き残りをかけて、
売れた売れない（公刊できた、できない）の無数の経験を積む。
- 神の見えざる手が働く。
- 低品質の品物が売れる値段に落ち着く
（学術的に無意味でも有意差が得られる程度に、その分野の研究の平均的な n が落ち着く）。

学問発展の十分条件を最初から目指す

- 「短縮期間はゼロではない」等の必要条件などには目もくれず
- 「新薬を服用した患者が回復するまでの平均日数は、対照群と比較して、医学的観点から評価して十分に短い」等の学問発展に対する確率的十分条件を示す。

表 5.1 $\text{phc}(\mu \leq c)$ の phc テーブル (新薬 A)

$c(\text{日})$	11.5	12	12.5	13	13.5	14	14.5	15
$\text{phc}(\mu \leq c)$	0.000	0.004	0.061	0.419	0.876	0.991	1.000	1.000

p値は抽象的

- 2群の入院期間の平均値には5%水準の有意差があった($p < 0.05$)
- 抽象的なp値が、0.05を下回ることが、ドメインにおける研究にどのような価値をもたらすのかが不明
- 有意性検定のもう一つの罪は「研究分野/文脈によらずに研究の価値を自動的に判定できる統計指標 p 値が存在する」という誤解を、利用する多くの学問分野に蔓延させたこと
- 学術的に意味のある差と、統計的有意差は連動しない

研究の価値判断は実感できる指標で

- **ア** 新薬使用の患者の平均入院期間は従来より1日半以上短縮
- **イ** 新薬使用の4分の3の患者の平均入院期間の点推定は14.2日
- **ウ** 新薬を使うと、入院期間が、比率で従来の 0.95以下
- 医師も患者も理解できる結論の表現を用いる
- その言明が正しい確率をPHCで示す
- $\text{PHC}(\text{ア}) = 87.6\%$ $\text{PHC}(\text{ウ}) = 99.8\%$

文献[1]参照。アは表5.1, イは表4.2, 表ウは表5.4

帰無仮説はデータを取る前から偽

- 帰無仮説が成立する確率はゼロ
- 帰無仮説の真偽を論じている文脈で、データを取る前から帰無仮説が義であるという性質は奇妙
- 有意差がなかった場合に、実験群と対照群の平均は等しいと結論してしまう誤り
- 差がないことを積極的に示したい状況が、科学研究における要請として存在するから、しかたなく有意性検定は誤用される

事実上同じ範囲 ROPE

- 母平均は、およそ15日の近辺に存在する
- ROPE region of practical equivalence
- $|\mu - 15| < c$ $\text{PHC}(|\mu - 15| < 0.25) = 0.997$
- 確率的評価を可能にするために差がない状況を区間で表現
- 効果がない（実験が失敗した）確率の評価

John Kruschke (2014) Doing Bayesian Data Analysis, Second Edition: A Tutorial with R, JAGS, and Stan. Academic Press.
前田和寛, 小杉考司(翻訳) (2017) ベイズ統計モデリング: R, JAGS, Stanによるチュートリアル 原著第2版, 共立出版. 第12章.

ビッグデータの時代に有意性検定は相応しくない

- 検定力：帰無仮説が偽であるときに帰無仮説を棄却する確率
- n を増価させれば検定力が上がりいつかは必ず有意に
- 万を超えるビッグデータを扱う際には、全て有意
- 転職に強いスキルとして、データ分析力は今後益々重要視
- Big Dataに無力な有意性検定は教育資源・教育機会の無駄使い
- 検定力分析は纏足（てんそく）である。

纏足：昔、中国で、女の足に子供の時から布を堅く巻きつけ、足を大きくしないようにした風習。

PHCは n が小さいと「自分に有利な主張がしにくい」

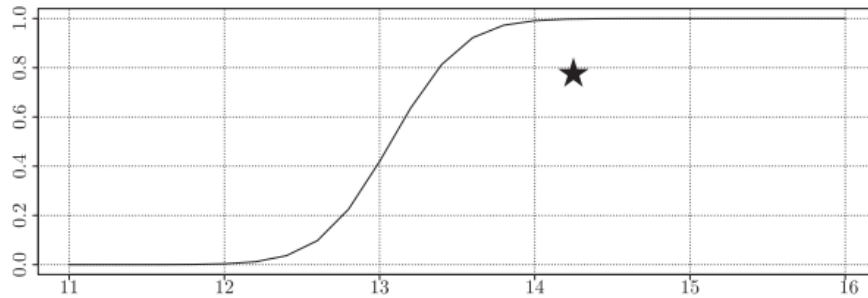


図 5.2 $\text{phc}(\mu \leq c)$ の phc 曲線 (新薬 A) $n=20$

表 5.1 $\text{phc}(\mu \leq c)$ の phc テーブル (新薬 A)

$c(\text{日})$	11.5	12	12.5	13	13.5	14	14.5	15
$\text{phc}(\mu \leq c)$	0.000	0.004	0.061	0.419	0.876	0.991	1.000	1.000

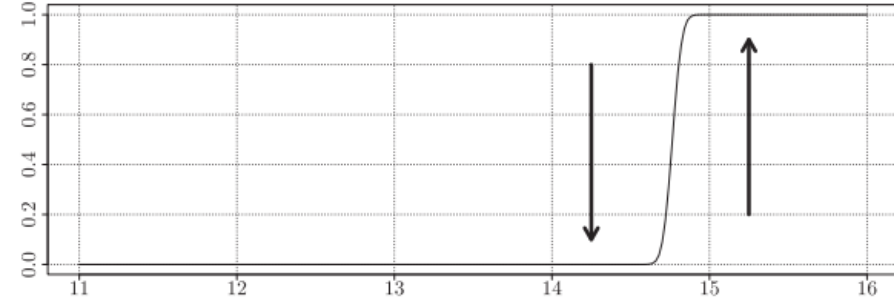


図 5.3 $\text{phc}(\mu \leq c)$ の phc 曲線 (新薬 B) $n=1000$

表 5.2 $\text{phc}(\mu \leq c)$ の phc テーブル (新薬 B)

$c(\text{日})$	14.5	14.6	14.7	14.8	14.9	15	15.1	15.2
$\text{phc}(\mu \leq c)$	0.000	0.000	0.089	0.772	0.998	1.000	1.000	1.000

- PHC曲線は n が小さいと緩慢 ($n=20$ 左図)、大きいと急峻 ($n=1000$ 右図)
- 小さいほうが望ましい母数のPHC曲線は単調増加関数
- n が小さいときは分析者に不利な★の領域でしか phc は高くない

データ数は多ければ多いほどよい

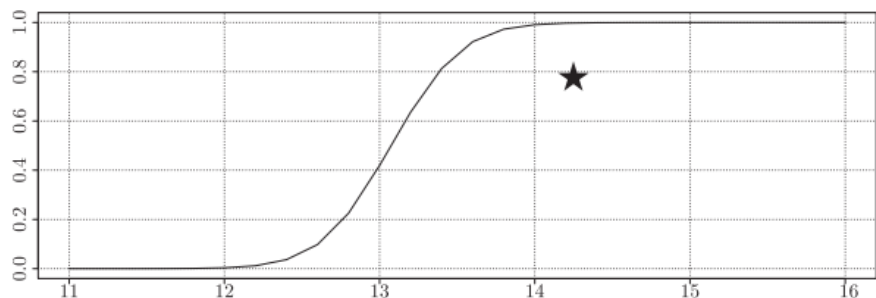


図 5.2 $\text{phc}(\mu \leq c)$ の phc 曲線 (新薬 A) $n=20$

表 5.1 $\text{phc}(\mu \leq c)$ の phc テーブル (新薬 A)

$c(\text{日})$	11.5	12	12.5	13	13.5	14	14.5	15
$\text{phc}(\mu \leq c)$	0.000	0.004	0.061	0.419	0.876	0.991	1.000	1.000

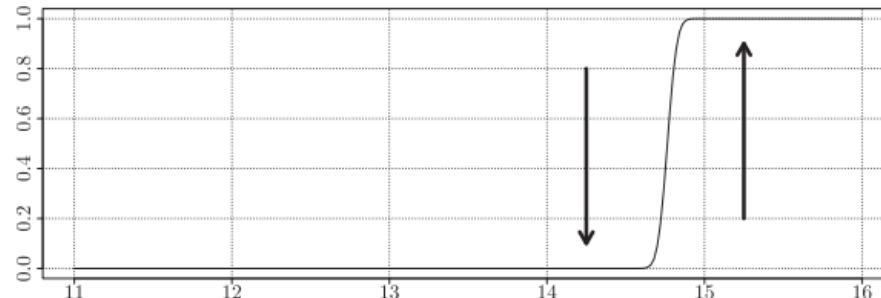


図 5.3 $\text{phc}(\mu \leq c)$ の phc 曲線 (新薬 B) $n=1000$

表 5.2 $\text{phc}(\mu \leq c)$ の phc テーブル (新薬 B)

$c(\text{日})$	14.5	14.6	14.7	14.8	14.9	15	15.1	15.2
$\text{phc}(\mu \leq c)$	0.000	0.000	0.089	0.772	0.998	1.000	1.000	1.000

- どの点も n の増加に伴って、PHCは0か1に確率的に近づく
- $\mu \leq 14.7$ のphcは 0.089 であり、 $\mu \leq 14.9$ のphcは0.998
- n が大きいと分析者に有利な仮説か否かとは無関係に明確な結論が出せる

検定統計量は自力で導けない

- 検定統計量の導出には解析学の知識が必要
- 「統計学はデータ分析のための単なる道具であり，ツマラナイ手続きの集まりを暗記する学問である」との決定的な印象
- 2群の肥満度BMIの差の有意性検定を自分で工夫できるか
- コロナワクチンの有効率の有意性検定を自分で工夫できるか
- セリグマンの犬のような学習性無力感を学生に与えてしまう

学習後に登場した指標（ワクチン有効率）

- ファイザー社のコロナワクチン
- 有効率：ワクチン接種によって、接種せずに発症したはずの何%が発症せずに済むか
- PHC($c \leq$ 有効率)

海外6カ国(米国, ドイツ, トルコ, ブラジル, アルゼンチン, 南アフリカ)において実施されました. ワクチンを接種する人とプラセボ(生理食塩水)を接種する人に分け, 約3週間の間隔で2回接種したとき, 新型コロナウイルス感染症の発症がどの程度抑制されるか比較されました. なお, 発症の確認に当たっては, 発熱や咳, 息切れ等, 感染が疑われる症状が1つ以上あり, PCR等の核酸増幅検査で陽性となった人が, 新型コロナウイルス感染症が発症した人と定義されました. (中略) 2回目の接種後7日以降の発症の有無が比較されました.

表 12.12 臨床試験の概要

	発症者数	追跡人数 (人年)
実験群	8	2214
対照群	162	2222

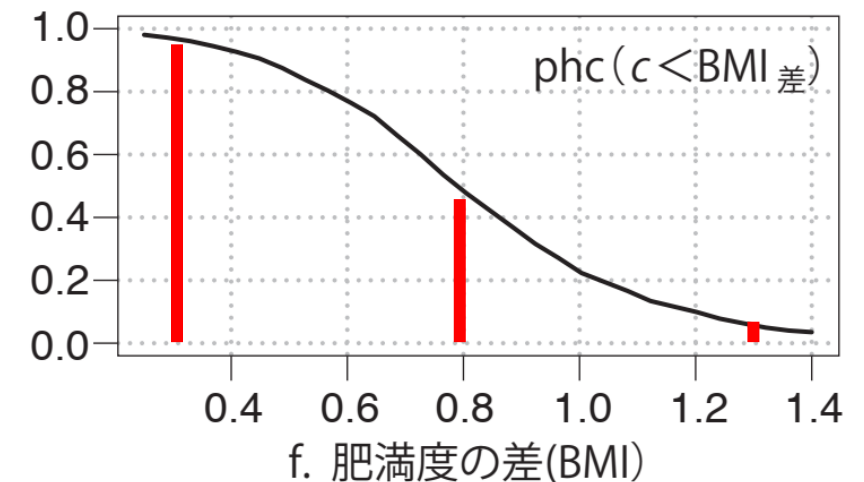
0.9	0.91	0.92	0.93	0.94	0.95	0.96	0.97	0.98	0.99	1.00	有効率
0.977	0.949	0.893	0.791	0.632	0.424	0.210	0.063	0.007	0.000	0.000	PHC

肥満度 BMI の差の推測統計的分析

- 体重差ではなく、肥満度の差
- BMIが0.36 以上減少する確率95%
BMIが0.80 以上減少する確率50%
- 統計学とは直接関係のない指標が星の数ほどある
- 統計学者の知らない指標のPHCを自分で工夫して計算できる醍醐味

表 7.2 減量法に関する 30 人のデータの要約統計量

		平均値	標準偏差	中央値
before	(x_b) kg	51.9	4.27	52.1
after	(x_a) kg	49.9	4.40	50.5
身長	(x_h) cm	156.7	6.66	156.2
減量	$(x_b - x_a)$ kg	2.0	0.54	1.9



PHCに対する批判

The Annual Report of Educational Psychology in Japan
2023, Vol. 62, 143-164

測定・評価・研究法

「教育心理学研究」におけるオープンサイエンスと

統計的分析方法の現状

——厳密なテストの観点からの考察——

山口 一 大
(筑波大学)

Current Status of Open Science and Statistical Analysis in *The Japanese Journal of Educational Psychology*: Consideration From the Severe Testing Perspective

KAZUHIRO YAMAGUCHI
(UNIVERSITY OF TSUKUBA)

This study reviewed 28 research papers published in *The Japanese Journal of Educational Psychology*, from issue 3 of volume 69 to issue 2 of volume 70, to assess new research practices such as pre-registration, the use of the Open Science Framework, and statistical methods, including Bayesian methods. The results indicated that these new research practices were not pervasive in the journal. We also discussed new research practices and statistical methods from the severe testing perspective. In addition, a confidence curve was explained to visualize the results of statistical hypothesis testing. Finally, based on the insights from our results, several approaches were proposed to improve these new research practices and statistical analyses.

「仮説が正しい確率」(e.g., Kruschke, 2014) が算出できることが、ベイズ統計学的方法の利点として述べられることがある。ここで言及されている確率は、「パラメタの事後分布を用いた確率」であり、先に述べた仮説検定の場合と同様に、サンプリングの仮定や想定した統計モデル（事前分布の設定を含む）がそもそも適切ではない場合には事後分布に基づく範囲が真のパラメタを含む確率とならないことがある点には注意が必要である。このような点から、「仮説が正しい確率」という文言をみたとき、真に仮説が正しい確率を考えているのではなく、「事前分布を含めた当該の統計モデルが定めるパラメタの事後分布を用いた確率」と補足して理解する必要がある。

山口一大(2023) 教育心理学年報 第62集(2022年度) p.156,left.

批判に対する返答

- 数理的前提が妥当でなければ、その程度に応じて、統計量が示す内容が妥当でなくなる性質は、PHCに限った性質ではなく、統計量があまねく有する性質であり、同時に限界である
- 実データから計算された統計量に関して言うなら、山口流の真のt値、真のp値、etc. など、どこにも存在しない。誰にも計算できない。PHCばかりではなく、すべての統計モデルの数理的前提はアイデアであり、飽くまでも『仮定や想定』でしかない
- あらゆる統計量に共通する限界を、この短いスペース全体を使って、さもPHCに特異な欠点のように『注意が必要である』とのレトリックは、バランスを欠いた論評である。長所も述べて欲しかった

原発不要論

- 原発はトイレなきマンション
- 数百年というほんの短い期間の人類の利便のために、その数百倍もの期間、危険物を子々孫々に背負わせるな
- ある主張が認められるか否かは、正しさだけでなく、現状からの距離も重要
- 現実政策とは、日々山積する懸案を卓越した調整能力で解決し続けること



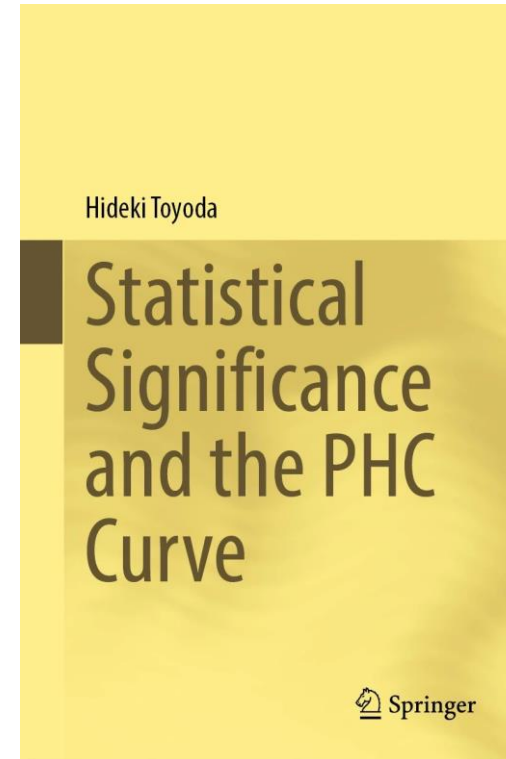
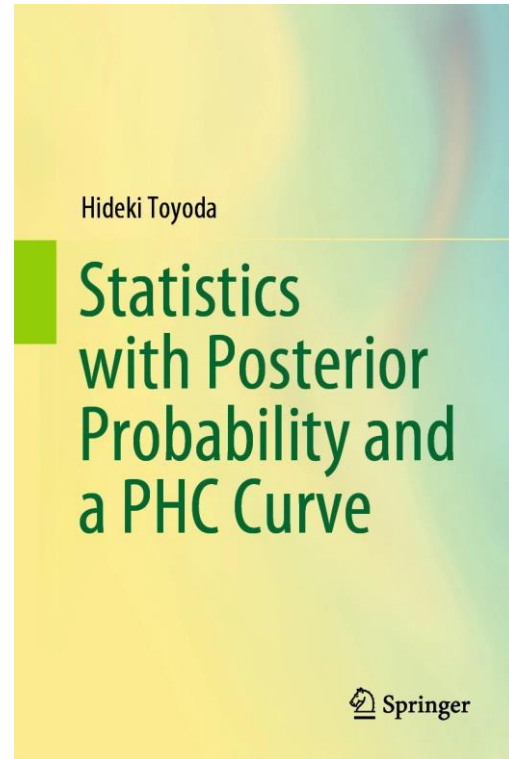
ちやぶ台返しは必要

- 大勢が動かなくても教程の変更を主張し続けることは大切
- 大幅な変更を唱える非主流の改革者が一定数存在するからこそ、長期的にみれば大きな変化が社会に生じる
- 少し前までは、電車や飛行機の中に灰皿があった
- 「論文の採否を、主にp値が5%を切るか否かで判定していた時代があったんだよ」という笑い話が、授業中に語られる日が、きっと来る

必修における教材として

- 早稲田大学で担当しているのは必修の心理統計学の入門的授業
- 典型的な学問体系を講義するべきでは？
- 有意性検定の概念を授業の中できちんと教える
- 有意性検定を否定している研究室の所属で、大学院生は差別されないだろうか
- 教え子たちには、頼まれたら有意性検定の講義して欲しい、主流派に留まっていて欲しい

A message to the future



Statistics with Posterior Probability and a PHC Curve. Springer. Hideki Toyoda, 2024. ISBN 978-981-97-3093-3
Statistical Significance and the PHC Curve. Springer. Hideki Toyoda, 2024. ISBN 978-981-97-7747-1

自己紹介とお礼

- 豊田秀樹（早稲田大学文学学術院教授）1961年生まれ
- 専門：心理統計学・教育測定学・マーケティングサイエンス
- このたびは栄えある「日本テスト学会賞」という名誉ある賞を賜り、誠に光栄に存じます。関係者の皆様のご指導とご支援に心より感謝申し上げます。これを励みにさらに精進してまいります。誠に有難うございました。
- 以下のURLにて動画を再視聴することが可能です。

https://wcms.waseda.jp/em/67afc5d4c836d?authkey=1739576571669&user_id=6010107&lg=ja