日本教育心理学会第58回総会自主企画シンポジウム

<ポストp値時代>の統計学

かがわ国際会議場

2016年10月9日(日)16:00ー18:00

企 画：秋山 隆（早稲田大学 ）

司 会：久保沙織（早稲田大学）

話題提供 ：豊田秀樹 （早稲田大学）

指定討論：楠見孝 （京都大学）

指定討論 ：向後千春 （早稲田大学 ）

- p値の誤解や誤用に対処する6つの原則に関する声明
  - アメリカ統計学会ASAが、2016年3月7日に発表
  - 「p値偏重主義に対する決別を統計学者のコミュニティがその会長とboardの名において宣言する」のは恐らく史上初めて（ASA president, Jessica Utts）
  - 声明は「『ポストp < 0.05 時代』へ向けて研究方法の舵を切らせることを意図している」(ASA executive director R. Wasserstein) ものだと言明

- ６つの原則
  - 1.p値は「そのデータがある特定の統計モデルとどれくらい適合しないか」を示し得る
  - 2.p値は「その仮説が真である確率」は与えないし「ランダムな偶然だけからそのデータが得られる確率」も与えない
  - 3.科学的結論及びビジネス・政策上の意思決定は「p値がある特定の閾値を切ったかどうか」だけに拠るべきではない
  - 4.適切な推論は完全な（データ及びモデルに関する）レポーティングと透明性を要するべきである
  - 5.単一のp値もしくは統計的有意性はその結果の効果や重要性の大きさを測るものではない
  - 6.p値そのものだけではモデルや仮説に関するエビデンスの良い指標たり得ない

訳：尾崎隆

- 検定のp値とは
  - 0.05を下回ったら卒論が書ける数値、ではない
  - 帰無仮説が正しい確率、でもない
  - １ーpが対立仮説が正しい確率、でもない

  - 正解
  - 帰無仮説が正しいと仮定したときに、手元のデータから計算した検定統計量が、今以上に甚だしい値をとる確率

  - さっぱりわからん、確率なのに、具体的イメージがわかない

- 何故p値は分かりにくいのか？
  - 帰無仮説が偽であることがデータを取る前から明白だから。
  - p値は有りえないことを前提として導いた確率なので、確率なのに抽象的で実感が持てない。
  - 有意性検定は、帰無仮説が偽であるという結論の下で、
    - 採択だったらnが小さかった
    - 棄却ならnが大きかった、
  - ということを判定する方法と言い換え可能。
  - 「帰無仮説を採択する」というときの「採択」という用語はジャーゴンである。
  - 有意性検定では、有意にならないからといって、差がないとは積極的にいえない。ましてや、母平均が同じなどと言ってはいけない。（言っている誤った教科書が多い）

- なぜp値(仮説検定)をやめるのか
  - p値は平均的にいくらでも0に近づくという性質があるから。
    - 有意になるまでデータをとるのが人情である。再現性のない、有意なだけで役に立たない論文が生産され続ける。
  - 多重比較は、さまざまな観点から多数の方法が提案され、収拾がつかなくなっているから。
    - 有意になる多重比較の方法を探す。
    - 多重比較という考え方は、統計学を隘路に迷い込ませる。
  - bigデータに対しては、あらゆる意味で有意性検定は無力だから。
    - 有意になっても、nが大きい場合には意味のある差とは限らない。

- 検定力分析は纏足（てんそく）である。
  - 「大切なのは健全な成長・発達なのに、歪んだ男の愛玩的欲望を満たすために、むりやり木靴をはかせて骨を折り、女性の足を大きくしないようにする。」のが纏足。
  - 「大切なのはデータなのに、道具でしかない有意性検定の制度を守るために、効果量と検定力と有意水準から逆算して、データ数を大きくしないようにする。」のが検定力分析。
  - 適切なデータは、多いほうがよい。間違った制度・道具は捨ててしまうが吉。

初級コースでは

不偏分散はよくわからないからとにかく暗記しましょう。
検定統計量も、どうせ導けないから暗記しましょう。

暗記・暗記・暗記で、自分で分析方法を工夫するなんてこと思いもよらないさ。

有意性検定で心理統計学に入門した場合

- 不偏分散がいらなくなる
  - 「データの要約的記述にはnで割った標本分散を利用し、母集団の推定にはn-1で割った不偏分散を利用する。」という誤った教科書の流布。
  - 不偏分散は検定統計量に多数必要。
  - 多種類の不偏分散は、数学的に高度なので、そういうものだと丸暗記しなさいと心理統計学では習えてきた。
  - 有意性検定をやめてしまえば、n-1や、その他の不偏分散を使わなくて済むし、習う側は暗記しなくて済む。
  - ベイズ的アプローチでは、母数は未知なる固定された値ではない。母数は分散に限らず確率的に分布する。だから通常の意味での不偏性という概念は、ベイズ的アプローチには存在しない。

- 心理統計学は暗記科目であった
  - 平均値の差を検定するための検定統計量は平均値の差ではなくt値である(どうして？)教えない。 →暗記
  - 何故t値はt分布するのか(どうして？)教えない。 →暗記
  - こういう場合は、この検定を使う →暗記
  - 理由を教えないから暗記しかできない。自分で工夫するなんて夢のまた夢。

  - 平均値の差を推測する生成量は平均値の差。 自然
  - MCMCの本質は高校数学でも理解可能。 数学的に平易
  - BMIの事後分布は、BMIの生成量。変動係数や比も推測可能
  - 学生が先生を追い越して、自分で分析を工夫することさえできる。 ドメイン知識が大切

上級コースでは

上級の心理モデルでは、せっかく習った検定はもう使わないみたいね。その上、尤度という新しい概念が出てきたわ。

木に竹をついだような感じだな。どうせ分かんないから因子分析も使い方だけ暗記しよう。

学習性無力感

有意性検定で心理統計学に入門した場合

初級コースでは

ベイズの定理しか使わないから、統一感があって理解しやすいわ。

暗記じゃなくて理解をしているから、
習っていない分析も自分で工夫できるよ。

ベイズ的アプローチで心理統計学に入門した場合

心理統計の中級･上級の
教材はまだない！

ベイズで入門すると
あとで困るのでは
ないだろうか

上級コースでは

上級になると、有意性検定とやらは出てこない。やっぱり勉強する必要は、なかったのね。

上級のモデルは、みんな尤度で記述されているぞ。尤度は、もう十分に習っているからよく分かるな。

読める！
読めるぞ！！

ベイズ的アプローチで心理統計学に入門した場合

- t検定，F検定，X2検定相当の分析をどうするか

- 2015年日本心理学会
  - 独立した2群
  - 対応ある2群
- 今日の話題は
  1要因実験計画
  - 例は違います
  - 2要因とクロス表の話は書籍を参照願います

- １要因の実験データの分析

- 1要因計画のモデル式

$$y_{ij} = \mu_j + e, \qquad e \sim N(0, \sigma_e)$$

- y_{ij} は、要因 A の j 番目の水準における i 番目の測定値
- μ_j は、水準 j の母平均
- e は、水準内の散らばりを表現する誤差変数
- e は、平均 0、標準偏差 σ_e の正規分布に従うことが仮定

- 母数の事後分布（ベイズの定理）

$$f(\boldsymbol{\theta}|\boldsymbol{x}) \propto f(\boldsymbol{x}|\boldsymbol{\theta})f(\boldsymbol{\theta})$$

[事後分布]　比例　[尤度] × [事前分布]

- 測定値の確率分布

$$f(y_{ij}|\mu_j, \sigma_e)$$

- 水準内の測定値の同時分布

$$f(\boldsymbol{y}_j|\mu_j, \sigma_e) = f(y_{1j}|\mu_j, \sigma_e) \times f(y_{2j}|\mu_j, \sigma_e) \times$$

$$\cdots \times f(y_{n_j j}|\mu_j, \sigma_e)$$

- 尤度

$$f(\boldsymbol{y}|\boldsymbol{\theta}) = f(\boldsymbol{y}|\boldsymbol{\mu}, \sigma_e)$$
$$= f(\boldsymbol{y}_1|\mu_1, \sigma_e) \times \cdots \times f(\boldsymbol{y}_j|\mu_j, \sigma_e) \times$$
$$\cdots \times f(\boldsymbol{y}_a|\mu_a, \sigma_e)$$

- 事前分布

$$\mu_j \sim U(0, 100), \qquad \sigma_e \sim U(0, 50)$$
$$f(\boldsymbol{\theta}) = f(\mu_1) \times \cdots \times f(\mu_j) \times \cdots \times f(\mu_a) \times f(\sigma_e)$$

- 事後分布

$$f(\boldsymbol{\theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{\theta})$$

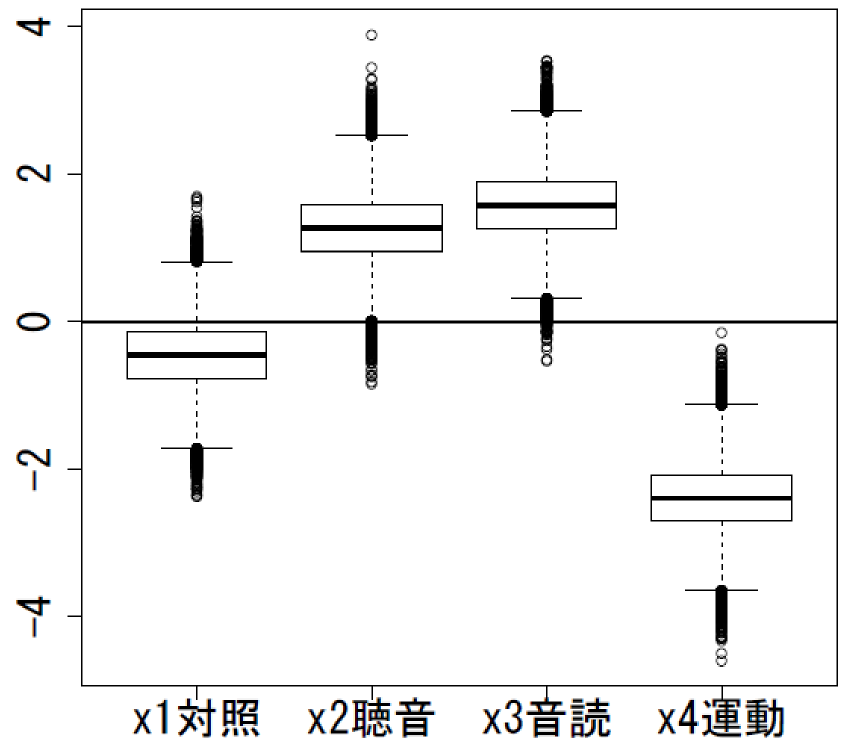
- 母数の推定結果

	EAP	post.sd	2.5%	5%	50%	95%	97.5%
対照 μ_1	31.04	0.55	29.96	30.14	31.04	31.94	32.12
聴音 μ_2	32.76	0.55	31.69	31.87	32.76	33.66	33.83
音読 μ_3	33.07	0.55	31.99	32.17	33.07	33.97	34.15
運動 μ_4	29.11	0.55	28.03	28.21	29.11	30.00	30.18
σ_e	2.43	0.20	2.08	2.12	2.41	2.78	2.86

- 水準の平均と水準の効果

$$\mu = \frac{1}{a}(\mu_1 + \cdots + \mu_a)$$

$$a_j = \mu_j - \mu$$



$$a_1 + \cdots + a_a = \mu_1 + \cdots + \mu_a - a\mu = 0$$

- 水準と要因の考察
-
- 水準の効果の有無
- 要因の効果の大きさ
- 水準間の比較(どの対に差があるのか)
- 連言命題が正しい確率
- 特に興味のある2水準間の比較

- 水準の効果の評価

「研究仮説 $U_{a_j>c}$: a_j は c より大きい」

が正しい確率は、生成量

$$u^{(t)}_{a_j>c} = \begin{cases} 1 & a^{(t)}_j > c \\ 0 & \text{それ以外の場合} \end{cases}$$

の EAP で評価

水準の効果が 0 より大きい（小さい）確率

条件	対照	聴音	音読	運動
$U_{a_j>0}$	0.169	0.996	1.000	0.000
$U_{a_j<0}$	0.831	0.004	0.000	1.000

- 要因の効果の評価

- 分散の分解

$$\sigma_y^2 = \sigma_a^2 + \sigma_e^2$$

- 要因の分散

$$\sigma_a^2 = \frac{1}{a}(a_1^2 + \cdots + a_a^2)$$

- 分散説明率

$$\eta^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}$$

- 効果量

$$\delta = \frac{\sigma_a}{\sigma_e}$$

効果の大きさに関する生成量の推定結果

	EAP	post.sd	2.5%	5%	50%	95%	97.5%
σ_a	1.628	0.268	1.102	1.188	1.628	2.071	2.158
η^2	0.311	0.076	0.160	0.184	0.313	0.434	0.456
δ	0.675	0.122	0.437	0.476	0.674	0.876	0.915

- 水準の対の比較

「研究仮説 $U_{\mu_j-\mu_{j'}>c}$: μ_j と $\mu_{j'}$ の差は c より大きい」

が正しい確率 $p(\mu_j - \mu_{j'} > c)$ は、生成量

$$u^{(t)}_{\mu_j-\mu_{j'}>c} = \begin{cases} 1 & \mu_j^{(t)} - \mu_{j'}^{(t)} > c \\ 0 & \text{それ以外の場合} \end{cases}$$

の EAP で評価

行 j の水準が列 j' の水準より大きい確率

条件	対照 (μ_1)	聴音 (μ_2)	音読 (μ_3)	運動 (μ_4)
対照 (μ_1)	0.000	0.013	0.004	0.994
聴音 (μ_2)	0.987	0.000	0.341	1.000
音読 (μ_3)	0.996	0.659	0.000	1.000
運動 (μ_4)	0.006	0.000	0.000	0.000

- 連言命題が正しい確率

研究仮説「$\mu_3 > \mu_2 > \mu_1 > \mu_4$」が正しい確率（0.642）

$$u^{(t)}_{\mu_3>\mu_2} \times u^{(t)}_{\mu_2>\mu_1} \times u^{(t)}_{\mu_1>\mu_4}$$

研究仮説「$(\mu_3, \mu_2) > \mu_1 > \mu_4$」が正しい確率（0.977）

$$u^{(t)}_{\mu_3>\mu_1} \times u^{(t)}_{\mu_2>\mu_1} \times u^{(t)}_{\mu_1>\mu_4}$$

研究仮説「$(\mu_3, \mu_2, \mu_1) > \mu_4$」が正しい確率（0.994）

$$u^{(t)}_{\mu_3>\mu_4} \times u^{(t)}_{\mu_2>\mu_4} \times u^{(t)}_{\mu_1>\mu_4}$$

- 特に興味のある2水準間の比較

音読と運動に関する差の推定結果

	EAP	post.sd	2.5%	5%	50%	95%	97.5%
$\mu_3 - \mu_4$	3.968	0.772	2.457	2.695	3.971	5.238	5.489
δ	1.645	0.345	0.974	1.078	1.644	2.214	2.320
U_3	0.940	0.041	0.835	0.860	0.950	0.987	0.990
π_d	0.871	0.051	0.755	0.777	0.877	0.941	0.950
$\pi_{2.0}$	0.713	0.076	0.552	0.579	0.718	0.830	0.847

それでもやっぱり心配は尽きない

なお残る3つの心配・躊躇

- ★1

- 有意性検定なら、結果が白黒はっきりつくので楽です。5%で有意なら差があると認められ、異論がでません。

- 査読をする場合も「有意なら差がある、そうでなければ、差が示されていない。」と自動的に判断でき、単純明快です。

- 「差がある確率が93%」などと論文に書いても、査読者が認めてくれるかどうかわかりません。

- ★1返答

- そもそも機械的に「5%で有意なら差があると自動的に判定すること」自体が誤りであり、それは誤った単純明快さである。

- 自動的に差の有無が判定可能であるという（誤った）単純明快さのために、無駄に多数の有意性検定が論文中にあふれてしまった。

- 差があるか否かという、実質科学的判定を、純粋に統計学の範囲内で済ませようという考え方は根本的に誤っている。

- ★２

- 初等的な心理統計部分は有意性検定を残し、高度な心理統計をベイズにしたらいいのでは？

- 現在活躍している心理学者は、全員、有意性検定で教育を受けてきているので、スイッチングコストが大きい。

- これから勉強する人が２度手間になる。最初から尤度の概念を教えたほうが、概念的に一貫した教育になる。習うほうは助かる。せめて教える側はベイズを使うべき。

- ★3

- これまでの論文の蓄積がある。有意性検定を教えないわけにはいかない。

- 早稲田大学心理学コースでは、有意性検定の
  - 理論
  - デメリット
  - 手続き的知識
- は今後も教授し続けます。

写真はイメージです

- メリットは分かったけれど、具体的な授業の方法が分からない。
- 2017年４月から放送大学にて、心理統計法の授業をベイズ的アプローチのみで行います。
- 有意性検定の理論とデメリットも講義します。
- t検定，F検定，X2乗検定相当の部分に加えて回帰分析が入ります。
- １つの授業例として参考になさってください。