



## 海老原誠治 (えびはら せいじ)

いただきます.info事務局、三信化工株式会社、資源と環境と教育を考える会『エコが見える学校』、女子栄養大学短期大学部非常勤講師、元関東学院大学非常勤講師。和食器を用いた出前授業や、テレビ局の撮影クルーの経験を生かして動画作成の研修会の講師も務める。

# 計算せずに、統計を考える／p値の限界

## ▶ 統計の意味

II

今年も学会の季節になりましたが、食育関係の学会といえば、5月の日本家政学会、6月の日本食育学会、9月の栄養改善学会でしょうか。最近参加する栄養教諭の先生方も増えたため、さまざまな食育の可能性を感じることができます。さて、よく発表や論文では、筆者【も】苦手な統計解析が行われます。今回は計算方法を無視しつつ、この意味や問題について考えます。

例えば、食に興味を持ってもらうための授業をクラスで実施したと想定します。

- ・クラスの人数：30名
- ・【介入前】食に興味がある：10人
- ・【介入後】食に興味がある：18人
- ・食への興味の増加：1.8倍

この場合、どのように評価したら良いでしょうか？ 興味の増加（意識変容）が偶然1.8倍になっただけなのか、同じような

対象者に授業を行ったとき、再現性があるのか？ これを考えると、統計解析が広く使われます。今回は計算したくないため、解析結果のみ書くと、前記の事例では次のような結果になります。

$p=0.038$  ( $p<0.05$ )

※  $\chi^2$  (カイ二乗) 方

※ExcelのCHISQ.TEST関数を使用

ここで出てくる「p」とは「p値」とも呼ばれますが、（表現が難しいのですが）偶然に起こりえる確率です。ですから「このプログラムの意識変容の結果は、偶然でしかない確率が0.038、約4%で、概ね再現されることが期待できるプログラム（有意である）」といえます。ここで出てくる「有意」とは、「偶然でなく意味がある」という意味になります。「有意」かどうかの目安として、p値の基本が5%以下 ( $p\leq 0.05$ ) であれば、偶然である確率は5%以下であり、おおよそ間違いない目安とされています。

表1 食に興味を持ってもらうための授業を30人に実施した際の意識変容

介入	食に興味がある (介入前)	食に興味がある (介入後)	意識変容した 人数の増加	p値	p値の意味
A	10/30人	20/30人	2倍に増えた	$p=0.010$	偶然であった確率は約1%
B	10/30人	18/30人	1.8倍に増えた	$p=0.038$	偶然であった確率は約4%
C	10/30人	17/30人	1.7倍に増えた	$p=0.069$	偶然であった確率は約7%
D	10/30人	15/30人	1.5倍に増えた	$p=0.194$	偶然であった確率は約20%
F	10/30人	14/30人	1.4倍に増えた	$p=0.292$	偶然であった確率は約30%



表2 食に興味を持つてもらうための授業を300人・3000人に実施した際の意識変容

介入	食に興味がある (介入前)	食に興味がある (介入後)	意識変容した 人数の増加	p値	p値の意味
G	1000／3000人	1100／3000人	+10%増えた	p=0.007	偶然であった確率は約1%
H	100／300人	125／300人	+25%増えた	p=0.035	偶然であった確率は約4%

す（＝有意水準）。

同じような相手にプログラムを実施する際、介入前後で意識変容の明確な差があれば曖昧さが少なくなるため、p値は小さくなり、偶然に生じた確率も減ります。例えば、A～Fまでの介入によって、「食に興味がある」人数の増加を見ると、最も大きな変容が起きた介入Aでは偶然に生じた可能性がp=約1%で再現性が高いといえ、最も小さな変容が起きた介入Fではp=約30%で、偶然であった可能性も高まります（表1）。以上よりp値が小さい方が、良い結果であるとされる場合もありますが、ここに大きな落とし穴があります。



## p値が証明しないもの



対象者（N数・サンプル数）が増えると、精度が上がり偶然の要素であるp値は小さくなります。介入Aから条件を変え、対象者を30人から、300人・3,000人と増やし、G・Hのような事例を考えます（表2）。

介入Gでは意識変容が+10%（=1.1倍）ですが、大きなサンプル数であるため、偶然に生じた可能性がp=約1%と精度が上がります。一方、介入Hは意識変容が+25%（=1.25倍）ですが、前記に比べて小さなサンプル数であり、偶然の可能性がやや増えて、p=約4%となります。

このとき、どちらが好ましい介入と考えるでしょうか？筆者であれば、介入Hが好ましいと判断します。p値では、介入したプログラムの良し悪しは一切判断できません。p値で証明されるのは、集計結果が偶

然でないことのみです。ですから、イマイチな内容のプログラムを、N数・サンプル数を増やしp値を小さくし偶然の要素を排除しても、高い精度で“イマイチ”であることを証明するだけです。

このような誤使用に関する問題を挙げるには、決して筆者が統計解析が嫌いだからではありません。Wikipediaで“有意”を調べると、専門家による問題の指摘が複数挙げられています（2025.05.30時点）。ソースをたどり詳細を確認すると、2016年、アメリカ統計協会が下記の課題を発信しています。※筆者意訳・抜粋

●p値は、仮説が正しい確率を計るものではない。 ●科学的結論・ビジネス・政策決定はp値が小さいかどうかのみで判断すべきではない。主張や結論を正当化する手段として小さなp値を用いるのは不適切な決定につながる可能性がある。

●統計的に良い結果のみ集めると、透明性に欠き、結果自体から本質的な意味が失われる。 ●p値は、効果の大きさや重要性を測るものではない。 ●p値自体は、仮説などを評価する指標にはならない。

2016, The ASA's Statement on p-Values, <https://www.tandfonline.com/doi/epdf/10.1080/00031305.2016.1154108?needAccess=true>

専門家による日本語訳（日本計量生物学会国際担当理事佐藤俊哉、2017年4月23日、<https://www.biometrics.gr.jp/news/all/ASA.pdf>）

以前も書きましたが、数字自体は嘘をつきませんが、その解釈や使い方によって、人間は数字を使って嘘がつけます。数字は大切ですが、数字に踊らされない姿勢はより大切です。