資　料

実　務　と　統　計（1）

―P 値中心主義の問題点―

山内俊平†（豊橋市立看護専門学校 非常勤講師，
豊橋市食肉衛生検査所 元所長）

1　は　じ　め　に

統計と聞けば耳を塞ぎたい，そんなイメージを持つ人は意外と多いのではないかと思います．私もその一人でした．多くの人にとって，難解な統計理論を理解するのはハードルが高く，深掘りしすぎると挫折の要因にもなりかねません．「統計解析は，実務上の価値を見出すための道具の一つにすぎない」と考えると，ずいぶん気が楽になった記憶があります．基本を押さえて，統計理論と齟齬のない方向性で解析結果を解釈し，合理的な実務判断ができればいい．こんな緩やかな考え方も必要かと思います．

ところが，この基本の部分があやしいケースが意外に多いと感じています．私自身も若い頃に，「**有意差あり（P＜0.05），だから効果がある**」といった業績発表をした苦い経験があります．今回は，統計初心者が陥りやすいP値中心主義の問題点について述べてみたいと思います．

2　P 値

P値は「帰無仮説（差がない）が正しいと仮定したときに，観察したデータと同じ，またはそれよりも極端な結果が得られる確率」というふうに定義されます．

例1：10回コインを投げて「表」が9回出たとします

帰無仮説（表出現確率＝0.5）が正しいと仮定すると，期待値は「表5回」です．

① 観察値の表9回（＋4回の偏り）と同じか，それより極端な結果（表10回）となる確率の合計（9回＋10回）を二項分布の確率から計算すると，0.01075です（表1）．

② 観察値と同じだけ偏った結果（表1回や表0回，－方向に極端な結果）も含めて考えると，二項分布の確率から同じく0.01075となり，両側合計でP値＝①＋②＝0.0215になります．なお，仮に観察値が「表8回」ならP値＝0.1094，「表7回」ならP値＝0.3438です．このことから，差が大きいほどP値は小さくなるように思えますが，実はそうともいえません．

表1　コインを10回投げたときの「表」がでる確率

表回数	試行回数	帰無仮説（表出現確率＝0.5）	事象発生確率
0	10	0.5	0.00098
1	10	0.5	0.00977
2	10	0.5	0.04395
3	10	0.5	0.11719
4	10	0.5	0.20508
5	10	0.5	0.24609
6	10	0.5	0.20508
7	10	0.5	0.11719
8	10	0.5	0.04395
9	10	0.5	0.00977
10	10	0.5	0.00098

例2：

① A，B群の各10人の血圧を測定したところ，A群の平均がB群より10 mmHg高かった．

② A，B群の各1,000人の血圧を測定したところ，A群の平均がB群より10 mmHg高かった．

帰無仮説（差がない）が正しいと仮定したときに，10人測定して10 mmHgの差が生じるのは，特に不自然とは思えませんが（P値は比較的大きい），1,000人も測って10 mmHgの差がでることは考えにくく，こんな結果を得る偶然性はきわめて低い．明らかに①よりも②のP値が小さいことは，直感的にも理解できるかと思います．

P値は，「差の大きさ」，「サンプルnの大きさ」，「標準偏差」などの影響を受けます．差が大きいからといってP値が小さいとは限りません．P値からは，どれほどの差があるのかはわかりません．

† 連絡責任者：山内俊平（豊橋市立看護専門学校 非常勤講師，豊橋市食肉衛生検査所 元所長）

〈注意点：差の方向性〉

例1で帰無仮説（表出現確率＝0.5）が正しいと仮定すると，得られた観察値と同じか，それより極端な結果（±の両側）を得る確率はP＝0.0215でした．

ここで仮にあなたが，このコインは表が出やすいのではないか（＋側）について検証したいとすると，「表9回以上」の極端な結果を得る確率ということになりP＝0.01075，反対に－の片側（このコインは表が出にくい）について調べたい場合は，「9回以下」の方向となりP＝0.99902ということになります．

3　有意差検定（仮説検定）

P値は，統計学的仮説検定（母集団に対する仮説を立てて，これを標本から検証する手法）で用いられる指標です．帰無仮説（差はない）が否定したい方の仮説で，主張したい方の仮説に対立仮説（差がある）を置きます．有意水準は5%とするのが一般慣例です．

両側検定における一般ルールと判定及びその解釈は次のとおりです．

帰無仮説（差はない）
（例：A群とB群で平均は同じ→平均差＝0）
対立仮説（差がある）
（例：A群とB群で平均が違う→平均差≠0）
帰無仮説棄却：P＜0.05

〈判定と解釈〉

P＜0.05
→偶然では説明しにくい差である→帰無仮説棄却
→統計的に有意（有意差あり）と推論する．

P ≧ 0.05
→偶然で説明できそうな差である→帰無仮説非棄却
→統計的に非有意（有意差なし）と推論する．

〈注意点：有意差なし（P ≧ 0.05）の誤解釈〉

「有意差なし（P ≧ 0.05）」を「差がない」と誤解釈している例をしばしば見かけますが，仮説検定で証明するのは対立仮説の方です．正しくは，「差があるとはいえない」となります．有意差なし（P ≧ 0.05）は，実際に得た結果からは，「帰無仮説を棄却し，差がある（有意差あり）」と推論するだけの十分な証拠が得られなかったという意味になります．

例1を有意水準5%で検定すると以下となります．

帰無仮説：コインの表出現確率＝0.5
対立仮説：コインの表出現確率≠0.5
帰無仮説棄却：P＜0.05

10回中9回表がでたときのP＝0.0215，P＜0.05より帰無仮説棄却，統計的に有意（有意差あり）です．対立仮説を表出現確率＞0.5と置くとP＝0.01075（統計的に有意），表出現確率＜0.5ならP＝0.99902（非有意）となります．平均や比率に差があるかを見たい（差の方向性は問わない）場合は両側検定，どちらか一方が大きい（または小さい）ことを検証したい場合には，片側検定となります．

4　信　頼　区　間

「**差の大きさ」の指標が信頼区間で**，一般に95%信頼区間を求めます．95%信頼区間とは，同じ手法で何度もサンプルを抽出し毎回信頼区間を計算すると，そのうち約95%の区間が母平均や母比率などの真値を含むという意味です．サンプルデータ（平均，標準偏差及びn数）を次のように設定して，母平均の95%信頼区間を計算してみます（表2）．nが大きい，あるいは標準偏差が小さい（データのばらつきが小さい）ほど，信頼区間の幅は狭まります．

表2　母平均の95%信頼区間

サンプルデータ			母平均の95%信頼区間
平均	標準偏差	n数	
50	10	8	41.64-58.36（50±8.36）
50	10	20	45.32-54.68（50±4.68）
50	5	20	47.66-52.34（50±2.34）
60	5	20	57.66-62.34（60±2.34）

〈信頼区間の見方と解釈〉

あるクラスの生徒から無作為に10人を選び平均身長を求めたら，170 cmであったとします．統計ソフトの解析で「95%信頼区間167～173 cm」と出力されたとしたなら，クラス全体の平均身長（真値）は95%の確からしさでこの範囲に含まれると推定される．こんな解釈でいいかと思います．

5　ま　と　め

「**有意差あり（P＜0.05）**」は，得られた差が偶然とは考えにくいと統計的に推論した結果ですが，実務面においては「差の大きさ」が重要です．**信頼区間**からどの程度の差（違い）がありそうなのか，そして何よりも，実務的に意味のある差なのかを評価することが大切になります．次のような**考察例**が考えられます．

考察例1：新薬の有効率は80%（n＝30）で既存薬より30%高かった（P＜0.05：有意差あり）

新薬有効率の95%信頼区間は0.614～0.923，既存薬より少なくとも11.4%以上も有効率が高いと推定される．P＝0.001（統計的に有意）から，この差は偶然ではないと考えられる．新薬の選択には実務的な価値があると判断できる．

考察例 2：新薬の有効率は 70%（n＝20）で既存薬より 20%高かった（P ≧ 0.05：有意差なし）

新薬有効率の 95%信頼区間は 0.457～0.881，区間上限は 88.1%で，非常に高い有効率の可能性もある．その一方で P＝0.115（統計的に非有意）から，この差は偶然の可能性も残る．もう少し例数（n）を増やして継続調査し，新薬の有効性を判断することとしたい．

考察例 3：新薬の有効率は 53.2%（n＝1,000）で既存薬より 3.2%高かった（P＜0.05：有意差あり）

新薬有効率の 95%信頼区間は 0.501～0.563，既存薬との差は数%程度しかなく，実務的には無視できるレベルである．P＝0.046 で統計的には有意だが，あえて新薬を選択する実務的価値は乏しい．

こんな感じで，信頼区間と P 値を実務に照らして評価することが望まれます．**「有意差あり（P＜0.05），だから効果がある」ではなく，信頼区間と P 値を両方セットで確認し，総合的に「差」の実務的意義を考察することが重要です．**