

医療統計学リテラシー(1)

P 値と仮説検定



大阪市立大学大学院医学研究科

医療統計学教室

新谷 歩

しんたに あゆみ

▶大阪市立大学大学院医学研究科医療管理医学講座医療統計学教授 ▶PhD(米国Yale大学)
 ▶国立病院機構本部総合研究センター生物統計室長, 東海大学客員教授, 名古屋第一赤十字統計アドバイザー, 藤田医科大学統計アドバイザー, 米国Vanderbilt大学客員教授, 厚生労働省厚生科学審議会臨床研究部会委員・患者申出療養評価会議評価員 ▶1996年Yale大学修士課程修了, 2000年同大学博士課程修了, 01年Vanderbilt大学医療統計学部講師, 07年同准教授, 14年大阪大学大学院医学系研究科臨床統計疫学寄附講座教授, 16年より現職 ▶1968年生まれ, 兵庫県出身 ▶著書: 今日から使える医療統計, みんなの医療統計, あなたの臨床研究応援します ▶主な専門: ICUにおけるせん妄研究, 糖尿病, リウマチ, 癌, 感染症, 腎臓病など多分野にわたる臨床データの統計解析。臨床研究ジャーナルに300以上に上る論文を執筆。最近では米国国立衛生研究所(NIH)推奨の電子データ集積システムREDCapの日本への普及に尽力 ▶受賞: 2013年Vanderbilt大学医学部ティーチングアワード

臨床研究, 基礎研究を問わず, 大切な研究結果をいざ世に送りだそうという時, 統計解析で大変苦労した, 査読者から「解析が間違っている」と指摘を受けて途方に暮れた, という経験を持つ方は多いのではないだろうか。近年多くの国際ジャーナルでは統計専門家による査読が行われ, 統計をいかに適切に行うかが研究の質を決めると言っても過言ではない。

本コラムでは, 研究者が陥りやすい統計解析上のポイントとその克服法について, 数式を一切用いず, 基本から分かりやすく解説する。第1回は, 論文で大変よく使われるP値と仮説検定について説明する。

* *

たとえば「新薬で治療された10人の血圧は, 既存薬で治療された10人と比べると平均で10低かった($P = 0.35$)」と記載があったとしよう。このPという値について, 何となく小さいほうがいいのは分かるけど, なぜ小さいほうがいいのか分からないという人も多いのではないだろうか。

そもそもP値とは何だろう。P値のPとは

Probability(確率)の頭文字からきている。P値とは「本当は差がない時に, 集めたデータで偶然差が出てしまう確率」を示している。

前述の例におけるP値は, 「本当は新薬の効果が既存薬と同じ場合, 集めたデータでこれくらいの差が偶然出てしまう確率が35%くらいある」という意味である。35%といえば結構高い確率である。つまりこれくらいの差は, 差がなくとも偶然起こりうることであって, 「このデータをもとに, 差があるとは言えない」と解釈できる。

次は, データを増やしてみる。すると「新薬で治療された100人の血圧は, 既存薬で治療された10人と比べると平均で10低かった($P = 0.01$)」となった。今度は, 「仮に, 新薬と既存薬の効果に差がないときには, こんな差が観察される確率は1%くらいしかない」, つまり, この差は偶然起こるわけがなく, 必然だった。「効果に差があったのではないか」と結論づけられる。

P値とは, 「偶然的確率」, 「まぐれ当たりの確率」さらに言い換えれば「データをもとに, 差があると判断した時の間違いの確率」と読み解ける。間違いの確率が低ければ, データをもとに「差がある」といっても間違いではないので, 「P値が小さければ

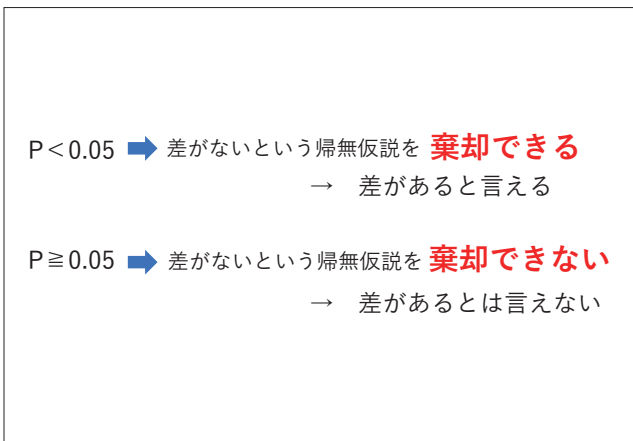


図1 エビデンスは、帰無仮説を棄却することによって作られる

小さいほど差を見るエビデンスとしては良い」というのは納得できる。

医学・歯学分野の研究では、この「差がないのに偶然差が出る確率であるP値」が慣習的に5%未満であれば、偶然ではなく必然、つまり「新薬と既存薬の効果に差があると言って良い」とされており、この差を「統計的に意味のある差、統計的有意差」と呼んでいる。

P値は「差があるかどうか」を判定する仮説検定に欠かせない指標であるが、なぜ偶然の確率であるP値が重要なのか。次にP値についてくわしく理解するために、仮説検定について考えてみよう。

* *

「カラスは黒い」という仮説を証明してみよう。「カラスは黒い」という仮説は、たとえ1万羽のカラスを観察しても、1万1羽目のカラスが白ければ「カラスは黒い」という仮説は証明できない。しかし「カラスは黒い」という仮説は1羽の白いカラスを観察すれば否定できる。つまり、仮説を正しいと証明するためにはエビデンスはいくらあっても足りないが、仮説を否定するには1つのエビデンスで十分である。仮説を否定することを、専門用語で「仮説を棄却する」という。仮説とは棄却するものであって、正しいと証明するものではないのである(この仮説を帰無仮説と呼ぶ)。エビデンスとは、仮

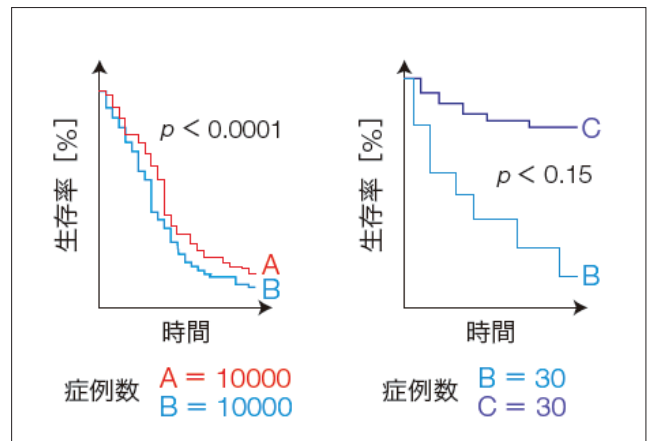


図2 観察された差、症例数によってP値が決まる。正しい統計解析には、データの背景を知ることが必要である(新谷 歩『みんなの医療統計, 12日間で基礎理論とEZRを完全マスター』講談社から引用)

説を棄却することによって作られる。「薬に効果がある」ことを示したい場合は、「薬に効果がない」という帰無仮説を棄却することによって達成できる(図1)。

P値に影響を及ぼす要因は大きく二つある。一つは観察された差、もう一つは症例数である。

新薬と既存薬で、平均血圧の差が10の場合と3の場合では、差が大きいほうが、偶然差が観察される確率は小さくなる。また症例数が大きいほど偶然起こる確率は小さくなる。ここにP値の落とし穴がある。

例えば図2では、左の例のほうで有意差が出ているが、右では出ていない。これをもって「C剤はB剤に比べて効果に差がないが、A剤は差がある」とするのは誤りである。B剤とC剤では、観察された生存率は大きく異なるが、統計的有意差が観察できなかったのは、症例数が少ないことによる。臨床的な差があっても、観察したデータの症例数が少ない場合、統計的有意差が出ないことはよくある。臨床的な差と統計的な差とは全く異なるものである。図2右の例では、臨床的な差は観察できたが、症例数が足りなかったため、統計的な有意差には至らなかったと解釈できる。

P値のみに頼らず、臨床的な差は確認できたのか、有意差が出なかった原因は何かなど、結果を広く理解することが重要である。