8. 偶然誤差と統計の役割

2019年7月17日

小野裕介

＊疫学データの分析における統計の役割

1. 偶然が果たす役割を評価しつつ、データのばらつきを測定する(本章)

2. 交絡などのバイアスを補正したうえで効果を推定(10, 12章)

＊系統誤差を取り除いても依然として残る誤差：「偶然誤差」

偶然の過程から生じる場合が多いがそれだけではない。

説明困難なデータのばらつき：隠れたバイアス、未知の因子など

#oy 確率的、偶然的、乱数とみなせるようなもの。(例) 正規分布

#oy 7章「研究のサイズを無限大にするとゼロにまで減らすことができる」

**(コラム)偶然**

・偶然という言葉の2つの意味：

無作為的な過程から生じる結果

予測が困難だが必ずしも無作為的な出来事ではない結末

・コインを弾いて表裏を当てるのは、一見、予測不能な無作為的な事象と考えられる。だが、

実際は初期条件と硬貨に加えられた力についての十分な情報があれば、コインの表裏を当てることは可能。

・データ中の誤差がどこから生じるのかを学べば、当初無作為的と思われた誤差も減らすことができる。

・疫学者が取り扱う問題の範疇であればデータ中に観察される偶然誤差の多くは、よりよい情報を得れば説明することができると考えるのが合理的である

**(1) 推定**

＊研究の結果、疫学的指標の「点推定値」が得られる。

＊点推定値の精度は「信頼区間」を用いて示される。

**(2) 点推定値、信頼区間とP値**

＊「信頼区間」の定義 (#oy　ロスマンの表現をいかしつつ、補足)：バイアスはなく、偶然誤差のみがあるとする。データの収集と分析を反復するとしたら、その都度、点推定値と信頼区間が

計算されることになる。偶然誤差があるため、真の値は、信頼区間に含まれたり、含まれていなかったりする。95%信頼区間とは、データの収集と分析を反復したときに真の値が95%の確率で含まれているような範囲である。

＊非実験的な疫学研究では、上の信頼区間の正式な定義は現実的ではない。信頼区間は結果の中に偶然誤差がどれくらい含まれるのかを示す目安と考えるのがよい。

＊統計学的仮説検定に用いられる値として、P値というものがある。P値と信頼区間は同じ式から計算される。#oy P値が0.05以上の領域が95%信頼区間。図8-3参照。

＊例：症例対照研究で2.5というRRの推定値を得たとする。真の値がRR=1.0だとしても

(帰無仮説)、偶然誤差があるので、1.0からある程度ずれた値を観察することになる。P値は、RR=2.5以上にRR=1.0(帰無仮説)からかけ離れた値を得る確率である。

#oy 小さいp値は、観察された値が1.0(帰無仮説)からかけ離れた値であることを意味する。

＊P値の計算に必要な理論的条件が満たされるのは困難なので、通常はP値を意味のある確率値とはみなさない。

＊P値を、あまり専門的な数値とは考えず、むしろ帰無仮説と手元の結果との相対的な一致性を示す指標と考えるのがよい。

＊P値は帰無仮説の正誤の指標でもない。帰無仮説の真偽は、他のデータの有無や帰無仮説と

その対立仮説の相対的な信憑性などによって最終的に判断されるべきものである。

(コラム)帰無仮説が正しい確率とは何か

＊P値は帰無仮説が正しい確率を示すものではない。

＊P値は確率ではあるが、それは帰無仮説が正しいと仮定してはじめて計算できる値である。

＊P値はデータと帰無仮説の一致性の指標とみることはできるが、帰無仮説の真偽にかかわる

ものとみることはできない。

＊P値が小さいとしても、帰無仮説がデータに対して最も合理的な仮説かもしれない。

#oy 他に合理的な仮説がない場合？

#oy 「科学的知識とは、既存の観察事象を説明すると思われる、まだ反証されていない仮説の

集合体」(p.53)

**(3) 統計学的仮説検定と推定**

＊統計学的優位差という言葉は多くの意味をもつように聞こえるが、実際はある任意の値

(ほとんどの場合で0.05)よりもP値が小さいかどうかを言及するにすぎない。

＊いかなるデータ解析を行っても帰無仮説もしくはその他の仮説の真偽を断定することは

できない。「有意さがある」あるいは「有意差がない」と宣言をし、帰無仮説の真偽を強制的に結論づけるものとして、統計学的検定が誤用されている。

＊統計学的有意差は有無を示すだけであり、数値であるP値よりも情報が減っている。

＊関係の強さと精度を表現しているという意味で、P値よりも推定のほうが優れている。

**(4) P値(信頼区間)関数**

＊P値を拡張した「P値関数」あるいは「信頼区間関数」：

表8-1のデータに基づいたP値関数(図8-1)

RRの任意の値についてその適合性の指標となるP値を計算したものがP値関数

帰無仮説の検定で用いられるいわゆるP値は、RR=1のときの値。

データに最も適合する仮説(点推定値)と推定の精度が視覚的にわかる。

P値が最大値をとるRRの値が点推定値

円錐の幅が推定の精度

(統計学的有意性検定などの)P値のみにもとづく推論は誤解をまねく。

RR=1に対するP値はRR=10.5に対するP値と同じ

＊有意性検定をデータの解釈に用いることの警鐘を示す例：

2つのP値関数の比較。

図8-2　表8-1のデータと表8-2の仮想的な症例対照データに関するP値関数

#oy 表8-1 効果はあるが、精度が低い。帰無仮説は棄却できず、有意差なしとされる。

#oy 表8-2 効果はないが精度は高い。帰無仮説は棄却され、有意差ありとされる。

#oy 臨床上意味ある差なのか、という問題。ビッグデータ時代でさらに問題に？

＊P値関数と信頼区間

p値関数から信頼区間をすべて求めることができるので、P値関数は信頼区間関数とも

呼ばれる。

95%信頼区間、90%信頼区間、80%信頼区間(図8-3)。

入れ子になっている(90%CIは95%CIに含まれる)。

単一の信頼区間だけで、P値関数全体を構築できる。

効果の強さと精度という２つの情報は信頼区間によって視覚化されている。よって、

P値関数自体をあえて示す必要はない。

#oy 信頼区間が持つ情報量は、P値関数が持つ情報量と同じぐらい豊か。

信頼区間を有意差の判断にしか用いないのはもったいない。

#oy P値関数はRothmanのtwitter(@ken\_rothman)の先頭でリンクされているepisheet.xlsで

計算できる。

**＊例：フルタミドは前立腺がんの治療に有効か？**

表8-3の結果。アイゼンバーガーらはフルタミドに効果はないと解釈。

OR=0.87

95%CI: 0.70-1.10 #oy 1.0を含む。有意にはならない。

わずかに効果があるとする10の先行研究の結果(統合するとOR=0.89)とは対立。

統計学的有意性検定にとらわれるあまり、データが示すわずかな有益な効果を無視して、間違った解釈をしてしまった。

先行研究とアイゼンバーガーらの研究を含めたP値関数(図8-4). 統計学的有意性検定に

不適切に頼ったため、先行研究を再現する結果だったにもかかわらず先行研究を否定し、結果はフルタミドが無効なことを示していると間違って解釈した。

#oy 有意性検定の観点からは、そもそも先行研究も「効果なし」なのでは？

#oy 先行研究と同じような傾向を示すデータだったのにもかかわらず、それを意識して

いなかったことを、ロスマンは批判している？

**＊例：セント・ジョーンズ・ワート(薬草療法)は大うつ病の改善に有効か？**

表8-4の結果。著者らは統計学的有意差の欠如という理由に基づいて解釈し、

セント・ジョーンズ・ワートは有効でないと結論づけた。

RR=2.0

90%CI: 0.90-4.6 #oy 1.0を含む。有意にならない。

P値関数(図8-5)

破線。RR=1.0と同じP値は、RR=4.1のとき。

裏にあるP値関数の位置と広がりを認識するためには、信頼区間(0.90-4.6)を見ればよい。

**(5) 信頼区間を計算する簡単な方法**

＊推定値は正規分布をとるという前提で計算。データが十分な大きさであれば正規分布に

基づいて推定値を得るのは合理的。

＊95%信頼限界をもとめるためには、1.96×SDを点推定値に加減。ここで、SDとは推定値の

分布の標準偏差(の推定値)で、データから計算される。

＊1.96という値は、正規分布において95%が含まれる間隔、ということで決まっている。

＊90%信頼区間をもとめるときは、1.645を点推定値に加減(式[8-1])

＊データから、点推定値と、点推定値の分布の標準偏差(の推定値)をどのように計算するのかは、次章で扱われる。

＊リスク差RDや発生率RDの場合は式[8-1]でよいが、リスク比RRや発生率比RRの推定値は非対称に歪むので、対数変換(lnRR)をして信頼限界を計算したうえで、元に戻す(式[8-2])

**(コラム)統計学的有意性検定と推定**

・統計学的有意性検定の結果に基づき、結論を出すことはよくない。

#oy 統計学的有意性検定は、バイアスがないと仮定したうえで、偶然誤差を評価するものでしかない。

・交絡、選択バイアス、測定誤差によるバイアスなどについて考慮するほうが重要。

・「有意差がない」ためにさらなる探求を中止した場合、交絡や非差異誤分類のために過少評価された、重要で強い連関が放置されることになるかもしれない。

・信頼区間の解釈においても、区間の境界値をそのまま読みすぎないことが大切。

#oy 区間の境界値を気にするのは、検定的な思考。