

二値アウトカムの取り扱い【第30回生物統計学】

1 概要

健康食品の開発における主要アウトカムとして、大小関係や変化量を容易に比較できる連続量の平均値が特に頻繁に用いられていますが、二値データを始めとした離散量を指標として機能性を評価するケースも存在します。今回はその手法を紹介します。

2 クロス集計表を用いた試験結果の可視化

特定の条件を満たしているかいないかを1か0かの二値のみで表現する場合、その結果にどのような傾向があるのか一目で把握するための表現方法としてクロス集計表がよく用いられます。あるバイオマーカーが境界域にある被験者を二群に分け被験食品もしくはプラセボを投与する試験を想定し、模擬データを用いて主要な解析の根本となる考え方を紹介します。

表1は試験の結果、境界域から正常域への改善が見られた、もしくは見られなかった人数を群別にまとめたものです。各群内で改善の見られた被験者がどの程度の割合いるのか比較的容易に把握できます。二値のアウトカムを評価する場合、この分布の差が偶然生じたものか、介入の効果によって生じたものか判断することを目的とします。

表1. 介入と改善判定基準の有無を示したクロス集計表

	改善あり	改善なし	合計
被験食品投与群	18	15	33
プラセボ投与群	6	24	30
合計	24	39	63

3 Fisherの直接確率検定

異なる介入を受けた二群において正常値への改善を示した被験者の割合に差が無いと仮定した場合、被験者全体の中で改善を示した24名の被験者(表1「改善あり」列の「合計」)は、介入の有無に関わらず、どちらの群に割り付けられても改善していたものと考えられ、また改善しなかった39名(表1「改善なし」列の「合計」)についても同様に、どちらの群に割り付けられても改善していなかったものと考えられます。Fisherの直接確率検定は、行と列それぞれの合計を固定した状態で、実際に観測された、あるいはそれ以上に極端な割合の差が生じるパターンの発生する確率が十分に高いかどうか計算することで介入による差が生じているのか検証する手法です。

表2には、被験食品群で改善の見られた人数が、実際に観測された18人より多くなる場合のクロス表の一



表 2. 観測値及びそれ以上に極端なパターンのクロス集計表とその発生確率 Pr

	改善あり	改善なし	合計
被験食品投与群	18	15	33
プラセボ投与群	6	24	30
合計	24	39	63

$$Pr_1 = \frac{33C_{18} * 30C_6}{63C_{24}} = 0.0039$$

	改善あり	改善なし	合計
被験食品投与群	19	14	33
プラセボ投与群	5	25	30
合計	24	39	63

$$Pr_2 = \frac{33C_{19} * 30C_5}{63C_{24}} = 0.0007$$

⋮
⋮
⋮

	改善あり	改善なし	合計
被験食品投与群	24	9	33
プラセボ投与群	0	30	30
合計	24	39	63

$$Pr_7 = \frac{33C_{24} * 30C_0}{63C_{24}} = 2.4 * 10^{-10}$$

例と、それぞれのケースが実際に観測される確率を示しました。Fisher の検定の P 値はこの確率の総和であり、この場合 $P \approx 0.0048$ になります。これは片側 2.5% で有意であり、改善の見られた 24 人のうち 18 人以上が偶然一方の群に偏る可能性は十分に低く、改善した被験者割合の群間差は介入の効果によるものであると考えられます。

4 カイ二乗検定

カイ二乗検定は、「介入効果に差が無いならば、各群の改善あり/なしの比率は全体の比率と等しくなることが期待できる」という考えに基づき、期待値と実際に観測された値の差が十分に小さいかどうか検証する検定です。帰無仮説に乗っ取った場合、表 1 に示した試験結果から期待される各象限に属する被験者数は表 3 の通りになります。この期待値と実際に観測された数値の相対的なズレの大きさを各象限ごとに算出します。それらの総和が「カイ二乗値」と呼ばれる統計量です。

$$\frac{\left(18 - 33 * \frac{24}{63}\right)^2}{33 * \frac{24}{63}} + \frac{\left(15 - 33 * \frac{39}{63}\right)^2}{33 * \frac{39}{63}} + \frac{\left(6 - 30 * \frac{24}{63}\right)^2}{30 * \frac{24}{63}} + \frac{\left(24 - 30 * \frac{39}{63}\right)^2}{30 * \frac{39}{63}} \approx 7.9521$$



表 3. 各象限に属する被験者数の期待値

	改善あり	改善なし	合計
被験食品投与群	33*24/63	33*39/63	33
プラセボ投与群	30*24/63	30*39/63	30
合計	24	39	63

観測された結果と期待値とのズレが大きいほどカイニ乗値が大きくなるので、得られたカイニ乗値よりも大きくなる確率が十分に低ければ、期待値とのズレが偶然に生じたものではなく介入の効果であると考えられます。この場合、自由度1のカイニ乗分布に従うと7.9521を上回る確率は約0.0048になり、片側2.5%で有意であることがわかります。

確率を直接算出する Fisher の検定と異なり、カイニ乗分布への近似で P 値を算出するのが特徴です。サンプル数が少ない場合正確性が損なわれるリスクがある代わりに必要な計算量が少なく済むため、各象限に割り振られたサンプル数が十分に多いケースに適している検定手法です。

5 まとめ

今回は二値データの分布に差があるか検定するためによく使われる手法を紹介しました。連続量の平均値によって統計的に差が検出されてもそれが臨床的意義のある差とは限りません。「臨床的にクリアしたい目標を達成した被験者の割合」を比較する手法も多くの試験で用いられています。目的に応じて適切な解析計画を設定することが重要です。