

p 値に基づく検定の再現性

久留米大学バイオ統計センター 柳川 堯

処置群と、処置群と同じ大きさのサンプルサイズをもつ対照群からなる 2 標本問題を想定し、処置効果を Δ として帰無仮説 $H_0: \Delta = 0$ vs. 対立仮説 $H_1: \Delta = \delta$ ($\delta > 0$) の統計的検定を考える. 検定統計量を T として T の実現値を t_0 で表す. p 値は H_0 の下での $T > t_0$ の確率. すなわち

$$p \text{ 値} = Pr(T > t_0 \mid H_0).$$

p 値は確率変数である. 本稿では、まず p 値の分布を吟味し、その 4 分位範囲が、サンプルサイズおよび δ の大きさによっては極めて大きいこと、したがって p 値のバラツキは、特にサンプルサイズが小さい時、極めて大きいこと、よって p 値による検定結果はサンプルサイズをあらかじめデザインしておかなければ再現性をもたないことを示す. 次に、再現性をある一定以上の確率で保証するためのサンプルサイズ決定式を与え、最後に与えられたサンプルサイズ決定式が Neyman-Pearson 流検定のサンプルサイズ決定式と同一であることを示す.

定義 1 あらかじめ定めた定数 α に対して

$p \leq \alpha$ のとき 「有意な処置効果あり」,

$p > \alpha$ のとき 「有意な処置効果ありとはいえない」

と判定する判定を、**p 値に基づく有意水準 α の検定**とよぶ.

定義 2 有意水準 α で検定結果が有意であった時にかぎり、同じ比較試験を同一のサンプルサイズで最初の試験とは独立に、くり返し実施する. このとき、有意水準 α で再び有意と判定されれば、有意水準 α の検定は**再現性をもつ**という.

定義 3 最初の試験の p 値を p この試験をくり返したときの p 値を p_2 で表す. このとき、条件付き確率 $Pr(p_2 < \alpha \mid p < \alpha)$ を有意水準 α の検定の**再現確率**という.

命題 1 有意水準 α の検定の再現確率は $Pr(p \leq \alpha)$ に等しい.

命題 2 $\delta = \delta_0$ を所与とするとき、有意水準 α の検定の再現確率を γ_0 とするサンプルサイズは、有意水準を α 、検出力を γ_0 として決定される Neyman-Pearson 流検定のサンプルサイズ決定式から導かれるサンプルサイズと一致する.