

検定の過誤、有意水準、検出力。

仮説検定においては表に示すような4通りの判断が想定され、「検定にもとづく解析者の判断が間違っているケース」として第一種の過誤、第二種の過誤が定義される。第一種の過誤とは『実際には有意差がないにも関わらず検定の結果「有意差あり」と間違っ判断してしまうこと』であり、その確率は有意水準あるいは危険率と呼ばれ α で定義される。第二種の過誤とは『実際には有意差があるにも関わらず検定の結果「有意差なし」と判断し有意な差を見落としてしまうこと』であり、その確率は β で定義される。

どちらの過誤も「判断の間違い」なのでできるだけ小さくなるのが好ましいが、一般に限られたデータでは第一種の過誤の確率 α を小さくすると第二種の過誤の確率 β は大きくなる。通常行う有意差検定では第一種の過誤を犯す確率 α をある数値(0.05、5%など)に定めて判定を行い、第二種の過誤については厳密な評価は行わず、検出力($1-\beta$ で与えられる)として値を示すだけの場合もある。検出力とは『差がある』ときに正しく「差がある」と検出できる確率』である。

データのばらつき(標準偏差 σ)、検出したい差(最小検出差 $\Delta\mu$)、有意水準(α)、検出力($1-\beta$)、および症例数(n)はお互いに影響しあっていて、例えば n 以外の4つが定めれば、検定に必要な「必要症例数」あるいは試験で目標とするという意味を込めて「目標症例数」を計算で定めることができる。関係式は統計学の書籍を参照。

検出力に関して次のことがいえる: 1) 平均値の差が大きいほど検出力は高まる、2) 例数が大きいほど検出力は高まる、3) 分散が小さいほど検出力は高まる、4) 有意水準が大きいほど検出力は高まる。

事実	「(帰無) 仮説が正しい」と判断する (「有意差なし」と判断)	「(帰無) 仮説が誤り」と判断する (「有意差あり」と判断)
「(帰無) 仮説は正しい」 (事実が「有意差なし」)	○ (問題なし)	第一種の過誤 (Type I Error) 有意水準・危険率 (α) で定義
「(帰無) 仮説は誤り」 (事実が「有意差あり」)	第二種の過誤 (Type II Error) β で定義: 検出力は $1-\beta$	○ (問題なし)

通常 $\alpha=0.05$ 、 $\beta=0.2$ が多く用いられる。このとき検出力は0.8。

特に「ひと」を対象とした臨床試験においては、『試験実施前に』必要症例数の計算を行い解析計画書にその根拠とともに記載する必要がある。後ろ向きカルテ調査などの場合には症例数が定めにくい場合もあるが、上に示した設定条件を考慮しつつ、実際にデータを活用できる症例数やデータ集積期間を定めるなど、必要十分な症例数が得られるような綿密な計画が必要である。

臨床第1相試験のように比較試験ではない場合でも、推定する評価指標(例えばAUCなどの薬物動態パラメータ値)の推定精度が妥当になるような例数を選択する。また動物実験等でも計画時に必要例数を妥当な根拠とともに設定すべきであり、不必要な資源の消費は避けるべきである。