Welch 検定って, 役に立つの？

柳川　　堯（久留米大学バイオ統計センター）
山下拓人（久留米大学大学院医学研究科社会医学専攻）

1. 研究の背景

$X_{i1}, X_{i2}, ..., X_{im_i}$ を正規分布 $\mathrm{N}(\mu_i, \sigma_i^2)$ からの確率標本として $(i=1,2)$ 統計的仮説検定問題 H_0: $\mu_1 = \mu_2$ vs H_1: $\mu_1 \neq \mu_2$ を考える. ただし, $\sigma_1 \neq \sigma_2$ とする. H_0 vs H_1 の妥当な統計量は

$$T = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{m_1} + \frac{S_2}{m_2}}},$$

で与えられる. T の分布は, H_0 の下で t -分布にしたがわない. Smith, R.A.(1936), J. Council Sci. Industrial Res), Welch, B.L.(1938, Biometrika, Vol.29), Aspin, A.A. (1948, Biometrika, Vol. 35) は, 自由度を調整することによって T の分布を t-分布で近似することを提唱し, 調整自由度として次を提案した.

$$\nu_w = \frac{(U_1 + U_2)^2}{\frac{U_1^2}{m_1 - 1} + \frac{U_2^2}{m_2 - 1}},$$

ただし, $U_i = S_i^2/m_i$. ν_w は Welch の調整自由度とよばれ, この自由度で調整した t-検定を Welch 検定, または Aspin-Welch 検定という.

2. 研究の目的

Welch の調整自由度は確率標本に依存しているにもかかわらず, その信頼度を評価することなく幅広く適用されてきた. 本研究の目的は, その信頼度を信頼度 95%の信頼区間を構成して評価すること, 特に, この信頼区間に対応する p 値の区間を求め Welch 検定結果の信頼度を評価する方法を提案することである.

3. 問題の定式化

問題を一般化し, K 標本問題として定式化する. 正規分布 $\mathrm{N}(\mu_i, \sigma_i^2)$ からとられた大きさ m_i の確率標本に基づいく標本平均と標本分散を $\bar{X}_i$, S_i^2 $(i = 1, 2, \ldots, K)$ で表し,

$$H_0:\ c_1\mu_1 + c_2\mu_2 + \cdots + c_K\mu_K = 0 \quad vs \quad H_1 : H_0\text{が成立しない}$$

の検定問題を考える. ただし, $c's$ は $\sum_{i=1}^{K} c_i = 0$ をみたす既知の定数（コントラスト）である. 妥当な検定統計量として, 次の T を考える.

$$T = \frac{\sum_{i=1}^{K} c_i \bar{X}_i}{\sqrt{\sum_{i=1}^{K} c_i^2 S_i^2 / m_i}}.$$

$\sigma_1 = \sigma_2 = \cdots = \sigma_k$ のとき T は H_0 の下で t-分布に従うが, そうでないときは t-分布には従わない. このとき, Satterswaite (1946) は Smith-Welch の考えを発展させ, T の分布を自由度を調整した t-分布で近似することを考え, 調整自由度として, 次を提案した.

$$\hat{\nu} = \left(\sum_{i=1}^{K} \frac{c_i^2 S_i^2}{m_i}\right)^2 \left(\sum_{i=1}^{K} \frac{1}{m_i - 1} (\frac{c_i^2 S_i^2}{m_i})^2\right)^{-1}. \qquad (1)$$

この自由度は Satterswaite-Welch df とよばれている.

4. 数学的発展

4.1 仮定と近似の戦略

[仮定] T の分母の二乗 $\sum_{i=1}^{K} c_i^2 S_i^2/m_i$ と等しい 1 次, 2 次のモーメントをもつ, 次の (1),(2),(3) をみたす統計量 ξQ_ν が存在する. (1) ξ は未知パラメータ, (2) Q_ν は自由度 ν のカイ二乗分布に従う統計量, (3) Q_ν は $\{\bar{X}_i\}_{i=1}^{K}$ と独立.

[近似戦略] 仮定が満たされるとき, T の分母の二乗を ξQ_ν で置き換える.

定理 1. 仮定が成り立つとして, 近似戦略を適用するとき, ν 不偏推定方程式は以下で与えられる.

$$\left(\sum_{i=1}^{K} \frac{c_i^2 S_i^2}{m_i}\right)^2 - 2\sum_{i=1}^{K} \frac{1}{m_i+1}\left(\frac{c_i^2 S_i^2}{m_i}\right)^2 - \nu \sum_{i=1}^{K} \frac{1}{m_i+1}\left(\frac{c_i^2 S_i^2}{m_i}\right)^2 = 0. \tag{2}$$

4.2 自由度の信頼区間の構成

$m_i, \bar{X}_i = \bar{x}_i$, および $S_i^2 = s_i^2$ が given という条件の下で正規分布 $\mathrm{N}(\bar{x}_i, s_i^2)$ から乱数 $X_{\ell i1}^*, X_{\ell i2}^*, \ldots, X_{\ell i n_i}^*$, を生成する, $\ell = 1, 2, \ldots, L$, ただし n_i は m_i, $i = 1, 2, \ldots, K$, と等しいとは限らない.

$$\bar{X^*}_{\ell i} = \frac{1}{n_i}\sum_{j=1}^{n_i} X_{\ell ij}^* \quad and \quad S_{\ell i}^{*2} = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(X_{\ell ij}^* - \bar{X^*}_{\ell i})^2.$$

とおく. 次が成り立つ.

定理 2. 仮定が成り立ち, 近似戦略を適用するとき, 自由度 ν の (unconditional) 不偏推定方程式は次で与えられる.

$$\left(\sum_{i=1}^{K} \frac{c_i^2 S_{\ell i}^{*2}}{m_i}\right)^2 - 2\sum_{i=1}^{K} \left(\frac{c_i^2 S_{\ell i}^{*2}}{m_i}\right)^2 \frac{(m_i + n_i)}{(m_i+1)(n_i+1)} = \nu_\ell \sum_{i=1}^{K} \left(\frac{c_i^2 S_{\ell i}^{*2}}{m_i}\right)^2 \frac{(n_i - 1)}{(m_i+1)(n_i+1)},$$

for $\ell = 1, 2, \ldots, L$.

定理 2 から与えられる L 個の推定量 $\hat{\nu}_1, \hat{\nu}_2, \ldots, \hat{\nu}_L$ を大きさの順序に並べ, 小さい方から 25%と 75%番目に対応するものを $\hat{\nu}_{Low}, \hat{\nu}_{Upp}$ とする. このとき, 信頼度 95%の ν の信頼区間は $(\hat{\nu}_{Low}, \hat{\nu}_{Upp})$. さらに, 自由度 $\hat{\nu}_{Low}$ のカイ二乗分布から算出される p 値を p_{Upp}. また, $\hat{\nu}_{Upp}$ から算出される p 値を p_{Low} とおく.

[提案] p 値だけではなく (p_{Low}, p_{Upp}) をみて H_0 を棄却するかどうか決める. 例えば, p_{Upp} 値 $<.05$ なら H_0 を棄却. p 値 < 0.05 でも p_{Upp} 値 > 0.05 なら H_0 を棄却しない, など.

参考文献

Satterswaite, F.E.(1946): Biometrics Bull.,2,110-114.

Smith, R.A.(1936): J. Council Sci. Industrial Res, 9, 211-212.