数理統計学まとめ（その 11）：第 6 章 検定

2.2 母分散 σ^2 がわかっていない (未知の) 場合の平均の検定

母分散が未知の場合は，推定値として不偏分散

$$\hat{\sigma}^2 = U^2 = \frac{nS^2}{n-1} = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \bar{X})^2$$

を使う．このとき

$$T = \frac{\bar{X} - \mu}{\frac{U}{\sqrt{n}}}$$

が自由度 $n-1$ の T-分布に従っているので，T-分布表を使って検定を行うことができる．両側検定，片側検定の考え方は母分散が既知の場合と同じで，違うのは正規分布の代わりに T-分布を使う所だけ.

帰無仮説 $H_0 : \mu = \mu_0$ と対立仮説 $H_1 : \mu \neq \mu_0$ について，両側検定を有意水準 α で行うことを考える．このとき帰無仮説 H_0 の下で，

$$T = \frac{\bar{X} - \mu_0}{\frac{U}{\sqrt{n}}} = \frac{\bar{X} - \mu_0}{\frac{S}{\sqrt{n-1}}}$$

は自由度 $n-1$ の T-分布に従うので，

$$P(|T| > t_{n-1}(\alpha)) = \alpha$$

となる自由度 $n-1$ の T-分布の両側 α 点 $t_{n-1}(\alpha)$ をとって，棄却域

$$R = \{|T| > t_{n-1}(\alpha)\}$$

が決まる．無作為標本 n 個をとり，データ $(x_1, \ldots, x_n)$ から $\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n)$ を作り，$t = T(x_1, \ldots, x_n)$ が棄却域 R に入るかどうかを調べる.

$t \in R$ つまり $|t| > t_{n-1}(\alpha)$ ならば仮説 $H_0 : \mu = \mu_0$ は棄却される.

$t \notin R$ つまり $|t| \leq t_{n-1}(\alpha)$ ならば仮説 $H_0 : \mu = \mu_0$ は採択される.

例 2.1 (教科書 p.122 例題 6.3)
ある県での統計によると，満六歳の児童の平均身長は $108.6cm$ であるという．A 小学校 6 才児 27 名について調べた所，$\bar{x} = 109.7cm, s = 3.98cm$ であった．この結果から，同校 6 才児童の身長は県平均に比べ高いと言えるか？

この問題を考えるとき，帰無仮説 H_0 は $H_0 : \mu = 108.6$ だが，対立仮説 H_1 として $H_1 : \mu > 108.6$ としてはいけないと教科書にある．この意味は，データをとる前に遡って考える必要がある．帰無仮説 H_0 および対立仮説 H_1 はデータをとる前に決めなくてはいけない．最初からなにか別の根拠がない場合，片側仮説はとらない．上の例の場合，別の根拠は示されていないので，両側検定を行う． 有意水準を 5 %として，データの t-値は

$$\frac{109.7 - 108.6}{\frac{3.98}{\sqrt{27-1}}} = 1.409$$

となる．$t_{26}(0.05) = 2.056 > 1.409$ だから，データの t-値 1.409 は棄却域に入らない．つまり， A 小学校 6 才児の身長は県平均と同じと考えるべきであるという結論になる.（「データから見て県平均に比べ高いとは言えない」と表現される.）

3　平均の差の検定

二つの正規母集団 (独立と思う）$N(\mu_1, \sigma_1^2)$, $N(\mu_2, \sigma_2^2)$ があり，そこから無作為抽出された n_1 個，n_2 個の標本をもとに，二つの母平均を同じとみなしてよいか否かを検定する事を考える．これを**平均の差の検定** という.

帰無仮説 $H_0 : \mu_1 = \mu_2$
対立仮説 $H_1 : \mu_1 \neq \mu_2$
として，次のように考える.

3.1　二つの母分散が既知の場合

第一の母集団からとった標本の平均

$$\bar{X}_1 = \frac{1}{n_1}(X_{1,1} + X_{1,2} + \cdots + X_{1,n_1})$$

は $N(\mu_1, \frac{\sigma_1^2}{n_1})$ に従い，第二の母集団からとった標本の平均

$$\bar{X}_2 = \frac{1}{n_2}(X_{2,1} + X_{2,2} + \cdots + X_{2,n_2})$$

は $N(\mu_2, \frac{\sigma_1^2}{n_2})$ に従う．$\bar{X}_1$ と $\bar{X}_2$ は独立なので，$\bar{X}_1 - \bar{X}_2$ は平均 $\mu_1 - \mu_2$，分散 $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$ の正規分布に従う．帰無仮説 $H_0 : \mu_1 = \mu_2$ の下では

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

が標準正規分布に従うので，標準正規分布の両側 α 点 z_α に対して

$$P(|Z| > z_\alpha) = \alpha$$

を使って両側検定をすることができる．

例 3.1 (教科書 p.64 例題 6.4)
ある学年で知能指数を測定し，男女別に集計したところ次の結果が得られた．

	平均	標準偏差	人数
男生徒	103	17	40
女生徒	101	12	35

男女差ありと言えるか？ただし知能指数の分布は $N(100, 15^2)$ と言われている．

解　表には標準偏差があるが，これは標本の標準偏差 S^2 と思われるが，これよりは後に書いてある知能指数の分散の方が信頼できると考え，ここでは知能指数の分散 15^2 を採用して，平均差の検定を行う．有意水準は 5 %としておく．男生徒の標本平均を $\bar{M}$ と書くと，これは $N(\mu_1, \frac{15^2}{40})$ に従い，女生徒の標本平均 $\bar{F}$ は $N(\mu_2, \frac{15^2}{35})$ に従う．帰無仮説 $H_0 : \mu_1 = \mu_2$，対立仮説 $H_1 : \mu_1 \neq \mu_2$ として，帰無仮説 $\mu_1 = \mu_2$ の下では $\bar{M} - \bar{F}$ は $N(0, 15^2(\frac{1}{40} + \frac{1}{35}))$ に従うので，

$$Z = \frac{\bar{M} - \bar{F}}{15\sqrt{\frac{1}{40} + \frac{1}{35}}}$$

が標準正規分布に従う．表から Z の値を求めると

$$\begin{aligned} Z &= \frac{103 - 101}{15 \times \sqrt{\frac{3}{56}}} \\ &= \frac{2\sqrt{56}}{15\sqrt{3}} = \frac{2\sqrt{168}}{45} = 0.576 \end{aligned}$$

標準正規分布の両側 5 %点は $z_{.05} = 1.96$ で，上の z 値 0.576 は棄却域

$$R = \{z \in \mathbb{R}; z < -1.96 \text{ または } z > 1.96\}$$

にいらないので，帰無仮説 $H_0 : \mu_1 = \mu_2$ は採択される．つまり，得られたデータからは男女差は認められない．(「男女の知能指数に有意の差があるとはいえない」)