2005 年 07 月 05 日（今野）

総合科目　数理・自然 8 「統計学入門」—　検定について

検定の考え方

$$\{\text{データ}\} = \{\text{決定的（非確率的）な要因}\} + \{\text{誤差}\}$$

と考える[1]．決定的（非確率的）な要因は理論的な仮説とみなせる．決定的（非確率的）な要因は理論的に予測できる量であれば，データと理論値とのずれは誤差のためかそれとも「それ以上の何らかの意味あるずれ」なのかを知りたい．後者の場合を「仮説からのずれが有意である」という．今後は「仮説が有意」と簡単にいうことにする．統計的仮説検定では，仮説のもとで求められた標本の確率が「稀」であるとき，すなわち，仮説からすれば，標本はでるはずのないかなりはずれた値（ずれが有意）のときに，仮説は有意である（仮説は誤っている）と判断する．ここで,「仮説があやまっている」と判断することを「仮説を棄却する」ということにする．稀であると判断する確率を有意水準という．

例（コイン投げの実験）　「コインに歪みがない」という仮説は「表の出る確率が 1/2 」であらわそう．そのために 20 回コイン投げの実験をおこなうとし，20 回中に表のでる回数を X としよう.「表の出る確率が 1/2 」という仮説を仮定すれば，X は二項分布 $Bin(1/2, 20)$ に従うので

x	14	15	16	17	18	19	20
確率	0.0370	0.0148	0.0046	0.0011	0.00018	0.00002	0.0000
累積確率	0.9793	0.9941	0.9987	0.9998	1.0000	1.0000	1.0000

となる[2]．表から

$$\mathbb{P}(X \geq 14) = 1 - 0.9432 = 0.0577$$

となるので，X の値が 14 以上となれば，たとえば,「稀な現象」を $X \geq 14$ とすることができる．

帰無仮説と対立仮説

- 帰無仮説 — もとの仮説（棄却の対象となるもの）．以後，H_0 と記す．
- 対立仮説 — 帰無仮説が棄却されたときの対案．以後，H_1 と記す．
- 誤りのパターン

[1] この式は模式的なもので，データの値は二つの項の和で必ず表されるというわけではなく，二つの要因によるというくらいの意味である．

[2] $P(X \leq 18) = 1 - P(X = 19, 20) = 1 - P(X = 0, 1) = 1 - P(X \leq 1) = 1 - (0.00001907 + 0.0000009357) = 1.0000$

		真実	
		H_0 が正しい	H_1 が正しい
決	H_0 を棄却しない	正しい判断	第二種の誤り
定	H_0 を棄却する	第一種の誤り	正しい判断

両側検定と片側検定

コイン投げの例を用いて説明する．帰無仮説と対立仮説を

$$H_0 : p = 1/2 \qquad H_1 p \neq 1/2$$

としたとき，これを両側検定と呼ぶ．また，帰無仮説と対立仮説を

$$H_0 : p = 1/2 \qquad H_1 p > 1/2$$

としたとき，これを両側検定という．もちろん，不等号の向きは逆でもいい．

正規母集団に対する検定

$X_1, X_2, \ldots, X_n$ を母集団分布 $N(\mu, \sigma^2)$ からの大きさ n の単純無作為標本とする．

両側検定

帰無仮説と対立仮説は

$$H_0 : \mu = \mu_0, \qquad H_1 : \mu \neq \mu_0$$

で与えられる．ただし，μ_0 は具体的な値である．

母分散が既知のとき

帰無仮説のもとでは，$\bar{X}_n$ は正規分布 $N(\mu_0, \sigma^2/n)$ に従うので，標準変換[3]をすれば，

$$Z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \tag{1}$$

は標準正規分布 $N(0, 1)$ に従う．検定方式は

$$|Z| > z_{\alpha/2} \qquad \text{ならば，} H_0 \text{ を棄却}$$

となる．ただし，$z_{\alpha/2}$ は標準正規分布の上側 $\alpha/2$ 点とである．

[3]標準変換は

$$\frac{\text{確率変数} - \text{平均}}{\text{標準偏差}}$$

である．

$\boxed{\text{例}}$ あるメーカーの製品の寿命は，平均が 1200 時間（$\mu_0 = 1200$），標準偏差が 150 時間（$\sigma = 150$）であると言われている.．この製品 10 個の標本（$n = 10$）についてテストした結果は

944, 1161, 1268, 1057, 1140, 1095, 1107, 1072, 955, 1110

であった．果たして，言われていることに偽りはないだろうか？

２つの仮説は

$$\text{帰無仮説 } H_0 : \mu = \mu_0 = 1200$$
$$\text{帰無仮説 } H_0 : \mu < \mu_0 = 1200$$

と設定し，H_0 のもとで

$$Z = \frac{\bar{X}_n - \mu_0}{\sqrt{\dfrac{\sigma^2}{n}}}$$

は標準正規分布に従うことを利用して検定を行う．有意水準 $\alpha = 0.05$ とすると検定方式は

$$Z < -1.645 \Longrightarrow H_0 \text{ を棄却}$$

となる．

上のデータより標本平均 $\bar{X}_n$ の実現値を求める：

$$\frac{944 + 1161 + 1268 + 1057 + 1140 + 1095 + 1107 + 1072 + 955 + 1110}{10} = 1090.9$$

となるので，Z の実現値は

$$\frac{1090.9 - 1200}{\sqrt{\dfrac{150^2}{10}}} = -2.30003$$

となる．したがって，帰無仮説 H_0 は棄却される．このデータの平均寿命は，主張される平均寿命よりも短く，それがランダムな影響と考えるよりもべつの理由（有意）で短くなっていると考えた方が妥当である．

しかし，有意水準を $\alpha = 0.01$ としてみると検定方式は

$$Z < -2.326 \Longrightarrow H_0 \text{ を棄却}$$

となるので，帰無仮説 H_0 は採択される．

$\boxed{\text{注意}}$：有意水準の選び方によって，結論が変わる可能性があることに注意する．