数理統計学まとめ（その10）：第 6 章 検定

1 検定の手順

1. 仮説の設定

   母集団のパラメータについて二つの仮説を設定

   (a) 仮説 H_0 ：**帰無仮説** $A = B$ (二つのものが同じである：変化がない)

   (b) 仮説 H_1 ：**対立仮説** $A \neq B$ (二つのものが異なる：変化がある）これにはいくつかの型がある．

   $A \neq B$ の型の仮説を**両側仮説**， $A > B$ の型の仮説を **右片側仮説**，$A < B$ の型の仮説を左 **左片側仮説** という．どれをとるかによって検定の方法が異なってくる (両側 α 点をとるか上側 α 点をとるか，など）

   対立仮説は データを見る前に決める．

2. 検定統計量の選定

   $T(X_1, \ldots, X_n)$ 仮説 H_0 を検定するのにもっとも適切な統計量を選ぶ．このとき，仮説 H_0 の下での T の分布についての知識が必要．（いままで勉強してきた）

3. **有意水準・棄却域**の決定．検定では仮説が正しいか間違っているかデータから結論を出す．この結論は間違う可能性もある．この間違う確率を**有意水準**という．普通，$\alpha = 0.05$ がよくとられるが，より厳しく $\alpha = 0.01$ をとることもある．

   有意水準 α が決まると，**棄却域** R を決める．これは対立仮説 H_1 のとり方により，$A \neq B$ の型のときは，棄却域は両側 α 点を使って決まる．$A < B$ や $A > B$ の型のときは片側 α 点を使って決める．決め方はそれぞれの場合に後で説明する．

4. 帰無仮説 H_0 の棄却または採択（保留）

   データ $(x_1, \ldots, x_n)$ から検定統計量の値 $T(x_1, \ldots, x_n)$ が決まり，これが棄却域 R に入るときは帰無仮説 H_0 は棄却され対立仮説 H_1 が

採択される．$T(x_1, \ldots, x_n)$ が R に入らないときは H_0 は棄却できないとしてが採択（保留）される．

2　平均の検定

「母集団の平均は μ_0 と見てよいか？」

という問題を考える．帰無仮説は

$$H_0 : \mu = \mu_0$$

対立仮説は

$H_1 : \mu \neq \mu_0$　(片側仮説：$\mu < \mu_0$，または $\mu > \mu_0$ をとることもある)

として考える．検定統計量は $\bar{X}$ を使う．母集団は大きいので正規母集団であるとしてよい（中心極限定理）．

2.1　母分散 σ^2 がわかっている (既知の) 場合

いま，**われわれは仮説 H_0 の元で考える**．このとき，大きさ n の標本 $(X_1, \ldots, X_n)$ の標本平均

$$\bar{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

は平均 μ_0 で分散 $\frac{\sigma^2}{n}$ の正規分布に従う．したがって

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

は標準正規分布に従うので，その両側 100α %点 z_α により，

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

である．$|Z| > z_\alpha$ を $\bar{X}$ について解くと

$$\bar{X} < \mu_0 - z_\alpha \times \frac{\sigma}{\sqrt{n}} \quad \text{または}\ \bar{X} > \mu_0 + z_\alpha \times \frac{\sigma}{\sqrt{n}}$$

となり，棄却域 R は二つの区間

$$(-\infty, \mu_0 - z_\alpha \times \frac{\sigma}{\sqrt{n}}],\ \ [\mu_0 + z_\alpha \times \frac{\sigma}{\sqrt{n}}, \infty)$$

の和集合となる．$\bar{X}$ が R に入らなければ帰無仮説 H_0 は採択（保留）され，入れば棄却され、対立仮説 H_1 が採択される．

対立仮説を

$$H_1 : \mu < \mu_0 \quad (\text{左片側仮説})$$

とすると，棄却域 R は 両側 2α 点 $z_{2\alpha}$ を使って

$$(-\infty, -z_{2\alpha})$$

と変わる．

$$H_1 : \mu > \mu_0$$

とすると，棄却域 R は

$$(z_{2\alpha}, \infty)$$

となる．

注意 2.1 教科書の正規分布表は両側 α 点 z_α の表にはなっていない．$0 < Z < z$ となる確率を表にしてある．両側検定，片側検定どちらを行うにしても，求めたい確率がこの表とどう対応するかをよく考える必要がある．

例 2.1 （両側検定：教科書 p.119 例題 6.1）　ある中学校で 1 年生 44 名に知能テストをした．偏差値の平均は 55.4 であった．偏差値の分布は $N(50, 10^2)$ に従うことが知られている．このとき，この中学校の 1 年生は平均的な生徒といえるか?

解　結果を見る前に仮説を設定するので， H_0 としては $\mu = 50$, H_1 としては $\mu \neq \mu_0$ ととる．両側検定をやる．有意水準は 5 %で定めて良いだろう．分散が 10^2 と既知なので，$n = 44$ より

$$Z = \frac{\bar{X} - 50}{\frac{10}{\sqrt{44}}}$$

が仮説 H_0 の下で標準正規分布になる．(正規両側検定）$z_{.05} = 1.96$ だから，Z にデータを入れて計算すると

$$Z = \frac{55.4 - 50}{\frac{10}{\sqrt{44}}} = 3.582 > 1.96$$

となり，Z は棄却域に入る．よって，帰無仮説 $H_0 : \mu = 50$ は棄却される．従ってこの中学の 1 年生は平均的ではない．

例 2.2 (片側検定：教科書 p.120 例題 6.2)　あるメーカーの電化製品の寿命はカタログによると平均が 1200 時間，標準偏差 150 時間とかかれている．標準偏差は正しいとしてこのカタログが正しいかどうかを検定する．この場合，カタログより長い寿命の製品については問題はないと考えて，対立仮説は

$$H_1 : \mu < 1200$$

とおき，左片側検定を行う．有意水準は 5 %でとることにする．サンプルを 10 個とって調べた所，標本平均 $\bar{x}$ は 1100 時間だった．

$$Z = \frac{\bar{X} - 1200}{\frac{150}{\sqrt{10}}}$$

が標準正規分布にしたがうので，棄却域は $Z < -z_{0.1}$.

$\bar{X} = 1100$ を代入してみると，$Z = -2.18$ となり，正規分布表で $z_{0.1}$ を求めると，z_α が両側 α 点だから，

$$P(|Z| > z_{0.1}) = 0.1$$

となる $z_{0.1}$ を探す．教科書の標準正規分布表を使うため上の式は標準正規分布が左右対称だから

$$P(Z > z_{0.1}) = 0.05, \quad \therefore P(0 < Z < z_{0.1}) = 0.5 - 0.05 = 0.45$$

となり，付表 1 の 4 桁の値が 0.4500 になるところを見ればよい．表には $z = 1.64$ のときの確率（表の上の斜線部の面積）が 0.4495 で，
$z = 1.65$ のときの確率が 0.4505 とあるので，求める z の値は 1.64 と 1.64 の間の値.

比例配分で z について

$$z - 1.64 : 1.65 - 1.64 = 0.4500 - 0.4495 : 0.4505 - 0.4495$$

が成り立つとして z を求めると，

$$x = 1.64 + \frac{0.45 - 0.4495}{0.4505 - 0.4495} = 1.645$$

だから $Z = -2.108 < -z_{0.1} = -1.645$ となり，H_0 は棄却され，カタログは正しくかかれていないと結論される.（比例配分の計算をしなくても $-z_{0.1} > -1.65$ は分かるので，これだけでも $Z = -2.18$ が棄却域に入っているのは分かる）