No.16 2標本検定

NAKADA Masayuki

Kobe University Secondary School

January 13, 2022

仮説検定の方法（復習）

e.g. 与えられたコインを投げたとき，表裏が同等に出やすいか否かを知りたい．

仮説の設定: 表の出る確率を p とする．

帰無仮説 $H_0: p = 0.5$，**対立仮説** $H_1: p \neq 0.5$（両側検定）

有意水準の設定: $\alpha = 0.05$ とする（両側検定なので，閾値は上下 0.025 ずつ）．

データの収集: 実際にコイントスを 20 回行ったところ，そのうち 15 回表が出た．

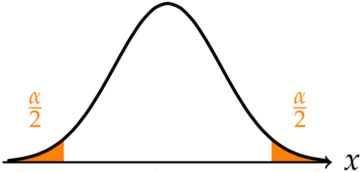
仮説検定の進め方

1. 一旦帰無仮説 H_0 を仮定し
2. $H_0: p = 0.5$ のもとで観測事実の起こる確率 $P(表 \geqq 15)$ を算出し
   - $p < 0.025$（i.e.「観測事象が十分稀である」）と言えるならば，帰無仮説 H_0 を**棄却**し，対立仮説を**採択**する．
   - 「観測事象が十分稀である」とは言えないならば，帰無仮説 H_0 は棄却されない．

$P(表 \geqq 15) \fallingdotseq 0.021 < 0.025$ であるから，帰無仮説 H_0 は有意水準 5% で棄却される．

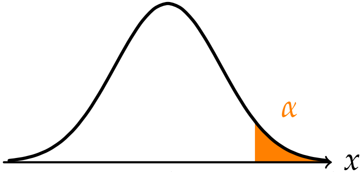
両側検定，片側検定について

例のコイントスにおいて，表裏の出方が公平であるか否か，すなわち，p と 0.5 に差があるか否かを検定したい場合，観測事実（表の出た回数）が極端に多くても少なくても，帰無仮説 H_0 を棄却すべきである．この場合，**両側検定**を用いる．両側検定では，有意水準 α を上位 $\frac{\alpha}{2}$ と下位 $\frac{\alpha}{2}$ に分けて閾値を設定する．



両側検定（$H_1: p \neq p_0$）

表が出やすくなるようにコインに細工を仕掛けた結果，本当に表が出やすくなっているのか否か，すなわち p が 0.5 より大きくなっているか否かを検定したい場合，観測事実（表の出た回数）が有意に大きいときに，帰無仮説 H_0 を棄却すべきである．この場合，**片側検定**を用いる．



右片側検定（$H_1: p > p_0$）

1 標本問題（z 検定，t 検定）における母平均の検定（復習）

正規分布 $N(\mu, \sigma^2)$ に従うとされる母集団から，n 個の標本 $[X_1, X_2, \ldots, X_n]$ をとる．
母平均 μ に関する仮説の設定は

帰無仮説 $H_0\colon \mu = \mu_0$，対立仮説 $H_1\colon \mu \neq \mu_0$（両側検定）

帰無仮説 $H_0\colon \mu = \mu_0$ を仮定する．

母分散 σ^2 が既知である場合: 標本平均を $\overline{X}$ とすると，標準化変数 $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ は標準正規分布 $N(0,1)$ に従う．得られた標本から $z = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}$ を計算し，閾値 $\frac{\alpha}{2}$ と比較する．
※先行研究や予備調査などから σ^2 が分かっている場合はこちらの検定方法を使う．

母分散 σ^2 が未知である場合: （不偏）標本分散 $\hat{S}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ で代用した標準化変数 $T = \frac{\overline{X}-\mu_0}{\hat{S}/\sqrt{n}}$ は，自由度 $n-1$ の t 分布 t_{n-1} に従う．得られた標本から $t = \frac{\overline{x}-\mu_0}{\hat{s}/\sqrt{n}}$ を計算し，閾値 $\frac{\alpha}{2}$ と比較する．

検定における誤り

e.g. 表裏が平等に出るように造られたコインを品質テストのために 20 回投げたところ，15 回表が出た．表の出る確率を p とし，次の仮説に基づき検定を行う．

帰無仮説 H_0: $p = 0.5$，対立仮説 H_1: $p \neq 0.5$

帰無仮説を棄却する／しないに関して，次の 4 つの場合が考えられる．

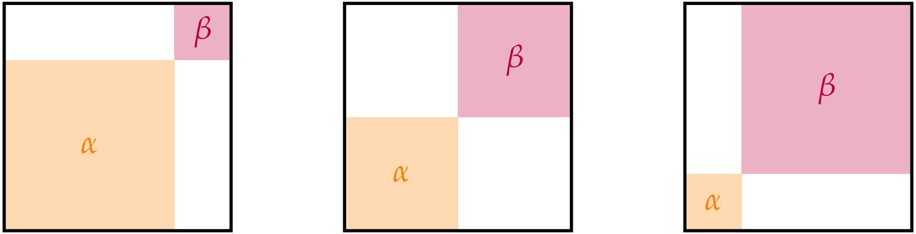
	H_0 が正しい	H_0 が誤り
H_0 を棄却しない	検定結果〇	第 2 種の誤り
H_0 を棄却する	第 1 種の誤り	検定結果〇

第 1 種の誤り　コインが正常であるにも関わらず，品質テストに不合格となる確率．
第 1 種の誤りを犯す確率は，有意水準 α である．

第 2 種の誤り　コインが不良品であるにも関わらず，品質テストに合格となる確率．
第 2 種の誤りを犯す確率 β は，後述の検出力に関係する．

	H_0 が正しい	H_0 が誤り
H_0 を棄却しない	検定結果○	第２種の誤り
H_0 を棄却する	第１種の誤り	検定結果○

※これらの誤りを犯す確率 α と β を同時に 0 にすることはできない.



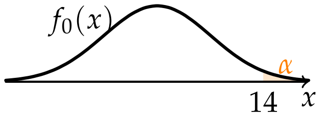
第１種と第２種の誤りの関係（イメージ図）

検出力

e.g. 表が多く出るように造られたコインを品質テストのために 20 回投げたところ，15 回表が出た．コインの表の出る確率を p とすると，次の仮説に基づき，有意水準 5% の右片側検定を行うときを考える．

帰無仮説 H_0: $p = 0.5$，対立仮説 H_1: $p > 0.5$

第 1 種の誤りを起こす確率: この確率は，H_0 を採択すべきであるのに棄却する確率であるから，有意水準である $\alpha = 0.05$ である．

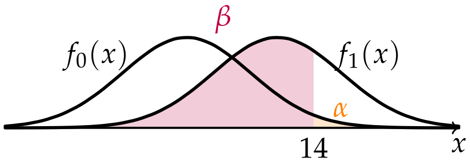


第 1 種の誤りを起こす確率

第 2 種の誤りを起こす確率: この確率を計算するために，p が 0.5 よりどの程度大きいと表が出やすいと判断してよいかという閾値 Δ を決めておく（**効果量**という）．今回は，$\Delta = 0.1$ として，対立仮説 $H_1 : p = 0.5 + \Delta = 0.6$ を仮定したときに H_0 を採択する（条件付）確率を求める．$B(20, 0.5)$ の上側 5% 点は 14 回であるから，

$$\beta = P_{H_1 :\, p=0.6}(X \leqq 14) = 0.8744$$

よって，H_0 を棄却すべきであっても誤って H_0 を採択してしまう可能性が 87.44% もあることになる．



第 2 種の誤りを起こす確率

第2種の誤りを起こす確率 β に対して，$\mathbf{1-\beta}$ を**検出力**という．仮説検定では一般的に，検出力が0.8（すなわち $\beta=0.2$）程度であると十分とされる（実は高すぎてもダメ）．

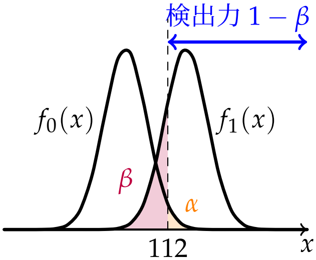
検出力を高めるには，

標本の大きさ n を増やす，効果量 Δ を工夫する

などの方法がある．
試しにコイントスの回数を200回に設定すると，
$B(200, 0.5)$ の上側5% 点は112回となり，

$$\beta = P_{H_1:\,p=0.6}(X \leqq 112) = 0.1397$$

よって，検出力は $1-0.1397=0.8603$ となり，この検定は十分な検出力をもつといえる．



n を大きくすると，「山が尖る」ため，検出力が上がる．

※例えばビッグデータを扱うなど n が巨大になると，検出力が強過ぎて，帰無仮説 H_0 はほぼ棄却できてしまう．このような事情から，ビッグデータの解析には仮説検定は向かないとされている．代替の分析手法として，信頼区間による推定などが用いられる．

2 標本問題における母平均の検定

対応のある検定 データが対をなす場合や，同一の標本における事前事後での平均の変化などを検定する場合． → 実質的には 1 標本問題.

対応のない検定 処置群と対照群の比較や男女比較など，一般的な 2 標本の平均を比較する場合.

- 2 集団の母分散が既知 2 標本の標本平均の差が従う分布を求める.
- 2 集団の母分散が未知 ウェルチの近似式を用いる（Welch 検定）.

対応のある検定

例題-問題 1

20 名の生徒が，10 点満点の漢字テスト①と漢字テスト②を受験した．漢字テスト①と漢字テスト②の間には，補習を受講した．20 人の漢字テスト①と②の成績をそれぞれ X，Y で表すと以下の表になる．この補習に効果があったと言えるだろうか，有意水準 5% の右片側検定により，判断せよ．

ID	X	Y	ID	X	Y	ID	X	Y	ID	X	Y
1	6	6	6	5	9	11	7	8	16	6	8
2	8	5	7	9	8	12	7	6	17	7	8
3	9	7	8	5	7	13	7	5	18	5	5
4	8	8	9	8	9	14	7	8	19	8	6
5	5	8	10	8	7	15	7	8	20	9	10

対応のある検定

例題-問題 1

20 名の生徒が，10 点満点の漢字テスト①と漢字テスト②を受験した．漢字テスト①と漢字テスト②の間には，補習を受講した．20 人の漢字テスト①と②の成績をそれぞれ X，Y で表すと以下の表になる．この補習に効果があったと言えるだろうか，有意水準 5% の右片側検定により，判断せよ．

ID	X	Y	ID	X	Y	ID	X	Y	ID	X	Y
1	6	6	6	5	9	11	7	8	16	6	8
2	8	5	7	9	8	12	7	6	17	7	8
3	9	7	8	5	7	13	7	5	18	5	5
4	8	8	9	8	9	14	7	8	19	8	6
5	5	8	10	8	7	15	7	8	20	9	10

$d = Y - X$ を考える．

ID	X	Y	d	ID	X	Y	d	ID	X	Y	d	ID	X	Y	d
1	6	6	0	6	5	9	4	11	7	8	1	16	6	8	2
2	8	5	−3	7	9	8	−1	12	7	6	−1	17	7	8	1
3	9	7	−2	8	5	7	2	13	7	5	−2	18	5	5	0
4	8	8	0	9	8	9	1	14	7	8	1	19	8	6	−2
5	5	8	3	10	8	7	−1	15	7	8	1	20	9	10	1

$d = Y - X$ より，$E(d) = E(Y - X) = E(Y) - E(X) = \mu_Y - \mu_X$ であり，$\mu_d = E(d)$（差の母平均）とおくとき，

帰無仮説 $H_0 \colon \mu_d = 0$，対立仮説 $H_1 \colon \mu_d > 0$

として，データ d について t 検定を行う．帰無仮説 $H_0 \colon \mu_d = 0$ を仮定すると

$$\text{標本平均}\ \bar{d} = 0.25, \quad \text{不偏分散}\ \hat{s}^2 = 3.25, \quad \text{t 統計量}\ t = \frac{\bar{d} - \mu_d}{\hat{s}/\sqrt{20}} = 0.620$$

であるから，自由度 19 の t 分布における上側 5% 点の値 1.729 より小さい．ゆえに，有意水準 5% で H_0 は棄却されないので，補習は効果があったとは言えない．

対応のない検定

2 集団の母分散が既知の場合

2 種類の母集団 U_X，U_Y の母平均 μ_X，μ_Y に差があるかどうかが知りたい．ただし，それぞれの母分散 σ_X^2，σ_Y^2 は（先行研究等により）既知であるとする．

$$帰無仮説\ H_0\colon \mu_X - \mu_Y = 0,\quad 対立仮説\ H_1\colon \mu_X - \mu_Y \neq 0$$

とし，有意水準 α の両側検定を行う．
標本 $[X_1, \ldots, X_m]$，$[Y_1, \ldots, Y_n]$ の標本平均を $\overline{X}$，$\overline{Y}$ とすると，$V(\overline{X}) = \frac{1}{m}\sigma_X^2$，$V(\overline{Y}) = \frac{1}{n}\sigma_Y^2$ であり，

$$E(\overline{X} - \overline{Y}) = E(\overline{X}) - E(\overline{Y}) = \mu_X - \mu_Y = 0 \quad (\because\ H_0)$$

$$V(\overline{X} - \overline{Y}) = V(\overline{X}) + V(\overline{Y}) = \frac{1}{m}\sigma_X^2 + \frac{1}{n}\sigma_Y^2 \quad (\because \overline{X}\ と\ \overline{Y}\ の独立性，\text{No.7}\ 参照)$$

より，帰無仮説 $H_0\colon \mu_X - \mu_Y = 0$ を仮定すると，統計量 $Z = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{1}{m}\sigma_X^2 + \frac{1}{n}\sigma_Y^2}}$ は標準正規分布 $N(0,1)$ に従う．あとは，（両側検定なので）閾値の $z_{\frac{\alpha}{2}}$ との大小を比較する．

2 集団の母分散が未知の場合

2 種類の母集団 U_X，U_Y の母平均 μ_X，μ_Y に差があるかどうかが知りたい．ただし，それぞれの母分散 σ_X^2，σ_Y^2 は未知であるとする．

帰無仮説 $H_0\colon \mu_X - \mu_Y = 0$，対立仮説 $H_1\colon \mu_X - \mu_Y \neq 0$

とし，有意水準 α の両側検定を行う．
母分散 σ_X^2，σ_Y^2 を（不偏）標本分散 $\hat{S}_X^2$，$\hat{S}_Y^2$ で代用する．統計量 $T = \frac{\overline{X}-\overline{Y}}{\sqrt{\frac{1}{m}\hat{S}_X^2+\frac{1}{n}\hat{S}_Y^2}}$ は，

$$\nu = \frac{\left(\frac{1}{m}\hat{S}_X^2 + \frac{1}{n}\hat{S}_Y^2\right)^2}{\frac{(\hat{S}_X^2/m)^2}{m-1} + \frac{(\hat{S}_Y^2/n)^2}{n-1}}$$

に最も近い整数 ν^* の自由度の t 分布 $t(\nu^*)$ に，近似的に従うことが知られている（**ウェルチの近似**）．あとは，（両側検定なので）閾値の $t_{\frac{\alpha}{2}}(\nu^*)$ との大小を比較する．

例題-問題 2

ある化学物質の濃度を 2 つの方法で測定した．その結果は次の通りであった（単位: %）．標準法に比べて，簡便法は濃度を過小評価しているだろうか．

標準法	25	24	25	26				
簡便法	23	18	22	28	17	25	19	16

2 群の母分散が分からないので，不偏標本分散で代用する．標準法，簡便法の濃度をそれぞれ X%，Y% とする．

帰無仮説 $H_0\colon \mu_X - \mu_Y = 0$，対立仮説 $H_1\colon \mu_X - \mu_Y > 0$

を，有意水準 5% で右片側検定する．

$\overline{x} = 25,\ \overline{y} = 21,\ \hat{s}_X^2 = 0.6667,\ \hat{s}_Y^2 = 17.7143,\ m = 4,\ n = 8$ として，ウェルチの近似式により自由度を計算すると

$$\nu = \frac{\left(\frac{0.6667}{4} + \frac{17.7143}{8}\right)^2}{\frac{(0.6667/4)^2}{4-1} + \frac{(17.7143/8)^2}{8-1}} = 7.988 \quad \text{より} \quad \nu^* = 8$$

よって，自由度 8 の t 分布において統計量は

$$t = \frac{25 - 21}{\sqrt{\frac{0.6667}{4} + \frac{17.7143}{8}}} = 2.592 > 1.860 = t_{0.05}(8)$$

であるから，有意水準 5% で帰無仮説 $H_0\colon \mu_X = \mu_Y$ は棄却され，簡便法は濃度を過小評価していることが分かる．