## 前回のQ＆A

Q1) eの小数乗のやり方がわかりません。

A1) eがついていない電卓では，普通は小数乗機能もついていませんね。失礼しました。

Q2) カイ二乗は記号の意味とわかりましたが，公式をもう一度教えてください。

A2) 観測値から期待値を引いて二乗して期待値で割ったものを足し合わせるとカイ二乗統計量がでます。

Q3) 何か統計学の本を買って学んでもこの授業と内容は変わりませんよね？　授業中，例題がないところはテストにもでないのですか？

A3) まともな本なら。しかし，まともな本を自習して理解するのは大変なので，講義に出た方が効率はいいと思います。なお，教科書を読めばわかるようなことだけを講義しても意味が無いので，たくさん出ている,「入門」系の本には書かれていないようなことにも，講義ではふれています。例題がないところは，計算問題は出しません。

# 統計学第８回　平均値に関する推定と検定

## （１）量的な変数を分析する

・ 前回まで２回はカテゴリ変数の分析を説明したが，今回は量的な変数の分析を扱う。

・ 量的な変数は，データ数と，個々の値の大きさに意味がある。

## （２）母平均値と標本平均の差の検定

▼ 標本平均 $E(X)=\sum X/n$ と既知の母平均値 $\mu_X$ の差の検定は，母分散 $V_X$ が既知のとき，$z_0=|E(X)-\mu_X|/\sqrt{V_X/n}$ が標準正規分布に従うことを使って検定できる（注：$n$ はサンプル数）。

▼ $V_X$ が未知のときは，標本の不偏分散 $S_X=\sum(X_i-E(X))^2/(n-1)=\mathrm{var}(X)$ を使って，$t_0=|E(X)-\mu_X|/\sqrt{S_X/n}$ が自由度 $n-1$ の $t$ 分布に従うことを使って検定できる（暗黙の仮定として，ランダムサンプルで，母集団の分布が正規分布であることが必要）。

▼ 未知の母平均値の信頼区間の推定はこの裏返し。 つまり，母平均値の95%信頼区間の下限は，不偏分散を標本数 $n$ で割ったものの平方根に自由度 $n-1$ の $t$ 分布の97.5%点を掛けた値を標本平均から引いた値になり，上限は，同じ値を標本平均に足した値になる。なお，R では，変数 $X$ について，`t.test(`$X$`,mu=既知の母平均値)` とすれば，上記の検定と推定を両方やってくれる。

（例題） 昨年，厚生科学研究で行われた「少子化の見通しに関する専門家調査」の結果の一部を見てみる。この調査は,「人口学，経済学，家族社会学，公衆衛生学を中心とした専門家を対象として少子化研究会のメンバーが対象候補者を抽出し，回答者の偏りや不足等について検討を加えた上で，748名を対象として調査を実施した。」もので，回収率は44 %であった。この調査では，2025年の合計出生率がいくつになるかという推定値が尋ねられていて，生データを見ると，

＞ 1.38 1.50 1.30 1.40 1.40 1.15 1.31 1.37 1.50 1.55 1.55 1.56 1.50 1.56 1.50 1.38 1.50 1.20 1.20 1.50 1.25 1.25 1.22 1.40 1.80 1.37 1.35 1.70 1.35 1.50 ...（後略）

のようになっていた（回答数は311，平均値は1.385，不偏分散は0.0252であった）。調査用紙には，厚生労働省『人口動態統計』から，国立社会保障・人口問題研究所による低位推計1.38，中位推計1.61，高位推計1.85という情報が掲載されていた。仮にこれらの値を母平均とすると，専門家たちが出した推定値がそれに一致するかどうかという問題設定ができる。

仮に中位推計を母平均としたとき，得られたデータがそれに一致するかどうかを見てみよう。母分散は不明なので，$S_X$ を使って，$t_0=|1.385-1.61|/\sqrt{0.0252/311}=25.0$ より，自由度310の $t$ 分布で25.0の上側確率はほぼ0なので両側検定のために2倍しても[1] ほぼ0であり，有意に異なるといえる。

低位推計を母平均としたときは，$t_0=|1.385-1.38|/\sqrt{0.0252/311}=0.555$ となり，自由度310の $t$ 分布で上側確率をみると0.289となる。つまり，得られたデータの母平均が1.38と等しいという帰無仮説が成立する確率は $0.289\times 2=0.579$ となり，棄却されない。

95%信頼区間を計算すると，下限が $1.385-1.968\cdot\sqrt{0.0252/311}=1.367$，上限が $1.385+1.968\cdot\sqrt{0.0252/311}=1.403$ となる。

（練習１） 昭和53年の国民栄養調査によれば３歳男児の平均身長は95.5cmであった。昭和53年にある保健所の３歳児健診に来所した326人の男児の平均身長が96.3cmで不偏分散が25.7であったとき，全国平均と比べて発育に差があるか検定せよ（出典：豊川裕之，柳井晴夫「医学・保健学の例題による統計学」現代数学社，1982）。

## （３）独立２標本の平均値の差の検定[2]

---

[1] 両側検定と片側検定については後で触れる。

[2] 分布がひどく歪んでいる場合には，Mann-WhitneyのU検定（Wilcoxonの順位和検定ともいう）を行う。２群を混ぜて個々の値に小さい方から順位を与え，その順位を使って検定統計量Uを計算するのだが，詳細は省略する。Rでは `wilcox.test(変数１,変数２,paired=F)` で計算可能である。この場合は，代表値としても平均値と標準偏差でなく，中央値と四分位偏差を表示する。また，分布の違いを検出したい場合は，コルモゴロフ＝スミルノフ検定を用いることができる。説明は省略するが，Rでは `ks.test(変数１,変数２)` で実行可能である。

▼ 標本調査によって得られた独立した２つの量的変数 $X$ と $Y$（サンプル数が各々$n_X$ と $n_Y$ とする）について，母分散が既知で等しい $V$ である場合は，$z_0 = |E(X) - E(Y)|/\sqrt{V/n_X + V/n_Y}$ が標準正規分布に従うことを使って検定する。

▼ 母分散が未知の場合（調査データを分析する場合はこれが普通）は以下の通り。

1. $F$ 検定（分散が等しいかどうか）：２つの量的変数 $X$ と $Y$ の不偏分散の大きい方を小さい方で割った $F_0 = S_X/S_Y$ が第１自由度 $n_X - 1$，第２自由度 $n_Y - 1$ の $F$ 分布に従うことを使って検定する。[3]
2. 分散に差があるか差がないかによって，平均値が等しいかどうかの検定法は異なる。[4]
   - 2-1. 分散に差がない場合：母分散 $S$ を $S = [(n_X - 1)S_X + (n_Y - 1)S_Y]/(n_X + n_Y - 2)$ として推定し，$t_0 = |E(X) - E(Y)|/\sqrt{S/n_X + S/n_Y}$ が自由度 $n_X + n_Y - 2$ の $t$ 分布に従うことを利用して検定する。[5]
   - 2-2. 分散が差がある場合（Welch の方法）：$t_0 = |E(X) - E(Y)|/\sqrt{S_X/n_X + S_Y/n_Y}$ が自由度 $\phi$ の $t$ 分布に従うことを使って検定する。但し $\phi = (S_X/n_X + S_Y/n_Y)^2/\{(S_X/n_X)^2/(n_X - 1) + (S_Y/n_Y)^2/(n_Y - 1)\}$。[6]

（例題） 先の調査では，少子化に対するイメージを問う質問項目もあった。「明るいイメージ」「どちらかといえば明るいイメージ」「どちらかといえば暗いイメージ」「暗いイメージ」の４つから選ぶのだが，仮に「明るいイメージ」または「どちらかといえば明るいイメージ」と答えた人（楽観主義と呼ぶことにする）と,「どちらかといえば暗いイメージ」または「暗いイメージ」と答えた人（悲観主義と呼ぶことにする）に分けると，楽観主義と悲観主義で比べたら，2025 年の合計出生率の推定値が違ってくるのではないかという仮説が考えられる。

楽観主義のデータは,

＞ 1.40 1.15 1.55 1.56 1.50 1.50 1.20 1.80 1.30 1.54 1.40 1.60 1.50 1.25 1.38 1.50 1.30 1.35 1.50 1.20 1.70 1.60 1.40 1.30 1.05 1.62 1.25 1.40 1.50 1.10 ... （後略）

となっていて（サンプル数 $n_X = 68$，平均 $E(X) = 1.384$，不偏分散 $S_X = 0.0337$），悲観主義のデータは,

＞ 1.38 1.50 1.30 1.40 1.31 1.37 1.50 1.55 1.56 1.50 1.38 1.20 1.50 1.25 1.25 1.22 1.40 1.37 1.35 1.70 1.35 1.55 1.60 1.70 1.20 1.31 1.40 1.40 1.60 1.10 ... （後略）

となっていて（サンプル数 $n_Y = 235$，平均 $E(Y) = 1.383$，不偏分散 $S_Y = 0.0234$），ぱっと見ただけでは差があるかないかさっぱりわからない。そこでまず思いつくのが，平均値を比べてやろうということである。楽観主義でも悲観主義でも同じ母集団に属していて，合計出生率の推定そのものには差がないと仮定すると，これらのデータは平均も分散も一致するはずである。

母分散は不明なので，まず $F$ 検定を行う。$F_0 = 0.0337/0.0234 = 1.443$ なので，第１自由度 67，第２自由度 234 の F 分布で上側確率を計算すると，0.0246 となり，分散が等しいという帰無仮説は棄却される。そこで，Welch の方法によって検定を行うと，$t_0 = 0.031, \phi = 95.465$ となるので，有意確率は 0.9753 であり，２群の平均値に差がないという帰無仮説は採択される。よって，楽観主義でも悲観主義でも 2025 年の合計出生率の推測値には差がないといえる。

（練習 2） 件の専門家調査には，出生率がそのうち回復するとみるか，低下し続けるとみるかという質問項目もあり，この答えの違いによって，2025 年の予測値には違いがありそうである。そこで，この２群間で比較を行ってみると，どうなるか計算してみよう。

回復するとみる人たちの 2025 年の合計出生率の予測値は,

＞ 1.40 1.40 1.56 1.50 1.40 1.80 1.37 1.40 1.40 1.60 1.60 1.25 1.50 1.50 1.70 ... （後略）

となっており（サンプル数 58，平均 1.487，不偏分散 0.0275），低下し続けるとみる人たちの予測値は,

＞ 1.38 1.30 1.15 1.31 1.37 1.50 1.55 1.55 1.56 1.50 1.50 1.38 1.20 1.20 1.25 ... （後略）

となっている（サンプル数 221，平均 1.356，不偏分散 0.0211）。これを元に計算すると？

## （４）両側検定と片側検定

・ これまでも何度か両側検定をしてきたが，ここで両側検定と片側検定の意味をもう一度きちんと押さえておこう。

---

[3] R では `1-pf(`$F_0$`,`$n_X-1$`,`$n_Y-1$`)` が有意確率になる。しかし，$F_0$ を手計算しなくても，`var.test(`$X$`,`$Y$`)` で等分散かどうかの検定が実行できる。この場合は，R が勝手に入れ替えてくれるので，$X$ の不偏分散の方が $Y$ の不偏分散より大きいかどうか気にしなくてもよい。

[4] 分散に差があるだけでも，別の母集団からとられた標本であると判断して，平均値が等しいかどうかを検定する意味はないとする考え方もありうるが，Welch の方法を使うか，ノンパラメトリックな方法を使って検定するのが普通である。

[5] R では，`t.test(X,Y,var.equal=T)` とすれば検定してくれる。

[6] R では，`t.test(X,Y)`（または `t.test(X,Y,var.equal=F` だが，var.equal の指定を省略した時は等分散でないと仮定して Welch の検定がなされるので省略していい）とすれば検定してくれる。

- ２つの量的変数 $X$ と $Y$ の平均値の差の検定をする場合，それぞれの母平均を $\mu_X$，$\mu_Y$ と書けば，その推定量は $\mu_X = \mathrm{mean}(X) = \sum X/n$ と $\mu_Y = \mathrm{mean}(Y) = \sum Y/n$ となる。
- 両側検定では，帰無仮説 H0 : $\mu_X = \mu_Y$ に対して対立仮説（帰無仮説が棄却された場合に採択される仮説）H1 : $\mu_X \neq \mu_Y$ である。H1 を書き直すと，「$\mu_X > \mu_Y$ または $\mu_X < \mu_Y$」ということである。つまり，$t_0$ を「平均値の差を標準誤差で割った値」として求めると，$t_0$ が負になる場合も正になる場合もあるので，有意水準 5%で検定して有意になる場合というのは，$t_0$ が負で $t$ 分布の下側 2.5%点より小さい場合と，$t_0$ が正で $t$ 分布の上側 2.5%点（つまり 97.5%点）より大きい場合の両方を含む。$t$ 分布は原点について対称なので，結局両側検定の場合は，上述のように差の絶対値を分子にして，$t_0$ の $t$ 分布の上側確率（$t$ 分布の確率密度関数を $t_0$ から無限大まで積分した値，即ち，$t$ 分布の分布関数の $t_0$ のところの値を 1 から引いた値。R では `1−pt(t0, 自由度)`）を２倍すれば有意確率が得られることになる。
- 片側検定は，先験的に $X$ と $Y$ の間に大小関係が仮定できる場合に行い，例えば，$X$ の方が $Y$ より小さくなっているかどうかを検定したい場合なら，帰無仮説 H0 : $\mu_X \geq \mu_Y$ に対して対立仮説 H1 : $\mu_X < \mu_Y$ となる。この場合は，$t_0$ が正になる場合だけ考えればよい。有意水準 5%で検定して有意になるのは，$t_0$ が $t$ 分布の上側 5%点（つまり 95%点）より大きい場合である。R で片側検定をしたい場合は，alternative という指定を追加する。例えば，$X > Y$ が対立仮説なら，`t.test(X,Y,alternative="greater")` とする。指定しなければ両側検定である。alternative に指定できる文字列は，greater の他には less と two.sided がある（指定しない場合は two.sided を指定したのと同じ意味になる）。

## （５）対応のある２標本の平均値の差の検定[7]

- 対応のある $t$ 検定：例えば，先に説明した専門家調査の結果で，2005 年の予測値と 2025 年の予測値に差があるかないかという問題を考えよう。同じ人について両方の値があるので，全体の平均に差があるかないかだけをみるのではなく，個人ごとの違いを見るほうが筋がいい。このような場合は，独立２標本の平均値の差の検定をするよりも，対応のある２標本として分析する方が切れ味がよい（差の検出力が高い）。対応のある２標本の差の検定は，paired-$t$ 検定と呼ばれ，意味合いとしては値の差を計算して値の差の母平均が 0 であるかどうかを調べることになる。R で対応のある変数 $X$ と $Y$ の paired-$t$ 検定をするには，`t.test(X,Y,paired=T)` で実行できるし，それは `t.test(X-Y,mu=0)` と等価である。2025 年の予測値は，

  > 1.38 1.50 1.30 1.40 1.40 1.15 1.31 1.37 1.50 1.55 1.55 1.56 1.50 1.56 1.50 1.38 1.50 1.20 1.20 1.50 1.25 1.25 1.22 1.40 1.80 1.37 1.35 1.70 1.35 1.50 ...（後略）

  のようになっていた（回答数は 311，平均値は 1.385，不偏分散は 0.0252）。2005 年の予測値は，

  > 1.30 1.35 1.34 1.35 1.32 1.25 1.34 1.34 1.40 1.40 1.35 1.30 1.30 1.32 1.35 1.39 1.30 1.30 1.30 1.20 1.33 1.35 1.30 1.37 1.40 1.33 1.39 1.35 1.35 1.30 ...（後略）

  であった（回答数は 311，平均値は 1.334，不偏分散は 0.00259）。これを普通に $t$ 検定するなら，明らかに分散が異なるので，Welch の検定によって $t_0 = 5.37$，自由度が 373.1 より両側検定の有意確率は $1.37 \times 10^{-7}$ となるが，対応のある $t$ 検定をすると，2025 年と 2005 年の予測値の差が，

  > −0.08 −0.15 0.04 −0.05 −0.08 0.10 0.03 −0.03 −0.10 −0.15 −0.20 −0.26 −0.20 −0.24 −0.15 0.01 −0.20 0.10 0.10 −0.30 0.08 0.10 0.08 −0.03 ...（後略）

  となりサンプル数 311，平均 −0.0508，不偏分散 0.0192 より，$t_0 = 6.46$ となり自由度 310 の $t$ 分布で上側確率を求めて２倍すれば，$p = 3.942 \times 10^{-10}$ となり，こちらの方が有意確率は小さくなる。いずれにせよ 5%水準で有意なので，2025 年の予測値と 2005 年の予測値は有意に異なるといえる。

[7] 分布が歪んでいる場合や，分布が仮定できない場合の対応のある２標本の分布の位置の差があるかどうか検定するには，ウィルコクソンの符号順位検定を用いる。説明は省略するが，R では `wilcox.test(変数１, 変数２,paired=T)` で実行できる。この場合も U 検定と同じく，代表値は中央値と四分位偏差で表示するべきである。