第 8 章

平均に関する推定と検定

本章では，量的な変数の代表値である平均についての分析法を説明する。まず，何らかの標本データの平均が母集団の平均に一致するかどうかを探る場合を考えよう。たとえば，山口市のある保健所で 100 人の 3 歳児の体重を測ったときに，その平均が全国平均として報告されている値と一致するかどうかを検討するような場合がこれに当たる。

8.1 母平均と標本平均の差の検定

サイズ n の標本 X について，標本平均 $E(X) = \sum X/n$ と既知の母平均 μ_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。

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）。

未知の母平均の信頼区間の推定はこの裏返しである。つまり，母平均の 95% 信頼区間の下限は，不偏分散を標本サイズ 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` と既知の母平均 `mu` について，`t.test(X,mu=mu)` とすれば，上記の検定と推定を両方やってくれる。

例題 1

2001 年に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は，それに一致しているといえるだろうか？

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。なお，元のデータを `x<-c(1.38,1.50,...)` のように `x` に付値しておけば，`t.test(x,mu=1.61,alternative="two.sided")` で母平均との差の検定結果が出力される（`,alternative="two.sided"`は省略しても同じ)。

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。

[*1] 両側検定と片側検定については第 6 章脚注 2 でも触れたが，後で詳述する。

例題 2

平成 10 年の国民栄養調査によれば，50–59 歳男性の平均 BMI は 23.6 であった。同じ年に A 社の職員健診を受診した 50–59 歳男性 248 人の平均 BMI が 24.6 で，その不偏分散が 8.6 であったとき，A 社の 50-59 歳男性は全国平均に比べて BMI に差があるといえるかどうか検定せよ。

母分散が未知なので，標本の不偏分散で代用すれば，$t_0 = |24.6-23.6|/\sqrt{8.6/248} = 5.37$ より，自由度 247 の t 分布で 5.37 の上側確率はほぼ 0*2なので，両側検定のために 2 倍しても有意に異なるといえる。

8.2 独立 2 標本の平均の差の検定

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*3。

8.2.1 母分散が未知の場合

調査データを分析する場合は母分散が既知であることはほとんどなく，これが普通である。手順としては以下の通りである。

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。*4
2. 分散に差があるか差がないかによって，平均が等しいかどうかの検定法は異なる。*5

---

*2 R で `1-pt(5.37,247)` を計算すると結果が `9.081154e-08` と表示される（これは浮動小数点表示となっていて，9.081154×10^{-8} を意味する）。

*3 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。詳細は次章で説明するが，その場合は，代表値としても平均と標準偏差でなく，中央値と四分位偏差を表示するのが相応しい。

*4 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。

*5 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。

8.2.2 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。*6

8.2.3 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 ϕ の t 分布に従うことを使って検定する*7。ただし

$$\phi = \frac{(S_X/n_X + S_Y/n_Y)^2}{\{(S_X/n_X)^2/(n_X - 1) + (S_Y/n_Y)^2/(n_Y - 1)\}}$$

例題 3

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に有意な差はあるだろうか？

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)
```

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

*7 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。

となっていて（サンプル数 $n_Y = 235$，平均 $E(Y) = 1.383$，不偏分散 $S_Y = 0.0234$），生のデータを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。実際はどうなっているだろうか？

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。

例題 4

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。
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）。2 群間に違いがあると言っていいか？

R で計算すると，

```
F0 <- 0.0275/0.0211
1-pf(F0,57,220)
```

とすれば，0.0916 という結果が得られるので分散に有意差はないといえる。したがって，Welch の方法にしなくていい。続けて R で計算すると，

```
S <- ((58-1)*0.0275+(221-1)*0.0211)/(58+221-2)
t0 <- abs(1.487-1.356)/sqrt(S/58+S/221)
2*(1-pt(t0,58+221-2))
```

の結果として `8.97506e-09` が得られ，平均には有意差があるといえる。もっとも，Rでは，`X` と `Y` に各群の生データを付値して，`t.test(X,Y,var.equal=T)` とすれば，同じ結果が得られる。通常はそれで十分である。

8.3 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。

2つの量的変数 X と Y の平均の差の検定をする場合，それぞれの母平均を μ_X, μ_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*8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，たとえば，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。たとえば，$X > Y$ 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，つまり両側検定になる）。

---

*8 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，すなわち，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)`。

8.4 対応のある 2 標本の平均の差の検定

たとえば，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。この場合は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情報量が失われない。このような場合は，独立 2 標本の平均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）[*9]。対応のある 2 標本の差の検定は，paired-t 検定（または対応のある t 検定）と呼ばれ，意味合いとしてはペア間の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 1.37×10^{-7} となるが，paired-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% よりずっと小さいので，2025 年の予測値と 2005 年の予測値は 5% 水準で有意に異なるといえる。

[*9] 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，Wilcoxon の符号付き順位和検定を用いる。詳しくは次章で説明するが，対応のある 2 変数 `X` と `Y` について，R では `wilcox.test(X,Y,paired=T)` で実行できる。この場合も U 検定のときと同じく，代表値は中央値と四分位偏差で表示するべきである。