**統計の手法レポート No.4 — 解説と解答例**

**担当： 渕野 昌** (2005 年 7 月 11 日)

**最終更新: 16. Juli 2005**

---

**以下は，講義で出さなかった仮想のレポートの解説です．ここでは，教科書 3 章と 4 章の例題のパターンのうちの幾つかについて解説をしています．試験では，以下の例題の類題が出題される可能性が高いので，よく理解しておいてください．**

**この文書は**

http://math.cs.kitami-it.ac.jp/~fuchino/chubu/statistics-05s-report03.pdf

**として downloadable です．なお この文書はまだ完成版ではなく，拡張中です 近日中に最終版と差替られることになる予定ですので注意してください．**

http://math.cs.kitami-it.ac.jp/~fuchino/chubu/statistics-05s.html

**にはこのファイルをはじめ，講義に関連する資料のリンクがあります．**

---

以下では，ある母集団の特性値について調べるために，無作為抽出した標本の個体での特性値を測定した状況を想定して，次の記法を固定して使うことにします:

(1) 母平均（母集団での特性値の平均）: $\mu$

(2) 母分散（母集団での特性値の分散）: $\sigma^2$

(3) 母標準偏差（母集団での標準偏差 $= \sqrt{\text{母分散}}$）: $\sigma$

(4) 標本のサイズ : $n$

(5) サイズ $n$ の標本をとったときの，標本の中の $i$ 番目の個体 $(1 \leq i \leq n)$ の特性値を返す確率変数 : $X_i$

(6) 標本平均（サイズ $n$ の標本をとったときのその標本での特性値の平均を返す確率変数）: $\bar{X} = \dfrac{X_1 + \cdots + X_n}{n}$

(7) 標本分散（サイズ $n$ の標本をとったときのその標本での特性値の分散を返す確率変数）: $S^2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(X_k - \bar{X})^2$

(8) 標本不偏分散（サイズ $n$ の標本をとったときのその標本での特性値の不偏分散を返す確率変数）:

$$U^2 = \frac{n}{n-1}S^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

(9) 実際にサイズ $n$ の標本をとったときの標本平均の値 : $\bar{x}$

(10) 実際にサイズ $n$ の標本をとったときの標本分散の値 : $s^2$

(11) 実際にサイズ $n$ の標本をとったときの不偏分散の値 : $u^2 = \dfrac{n}{n-1}s^2$

このとき次が成り立つ:

(12) $E(\bar{X}) = E\left(\dfrac{X_1 + \cdots + X_n}{n}\right) = \dfrac{1}{n}(\mu + \cdots + \mu) = \mu$ （教科書 p.66, 1° による）

(13) $$S^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k-\bar{X})^2$$
$$= \frac{1}{n}\sum_{k=1}^{n}\left((X_k)^2 - 2X_k\bar{X} + \bar{X}^2\right)$$
$$= \frac{1}{n}\sum_{k=1}^{n}(X_k)^2 - 2\cdot\frac{1}{n}\left(\sum_{k=1}^{n}X_k\right)\bar{X} + n\cdot\frac{1}{n}\bar{X}^2$$
$$= \frac{1}{n}\sum_{k=1}^{n}(X_k)^2 - \bar{X}^2$$

(14) $$V(\bar{X}) = V\left(\frac{X_1+\cdots+X_n}{n}\right) = \frac{1}{n^2}(V(X_1)+\cdots+V(X_n)) = \frac{1}{n}\sigma^2$$
（教科書 p.66, $2^\circ$ による）

(15) $$V(X_k) = E\left((X_k)^2\right) - (E(X_k))^2;\ \ V(\bar{X}) = E(\bar{X}^2) - \left(E(\bar{X})\right)^2$$
（教科書 p.55, (2) による）

(16) $$E(S^2) = E\left(\frac{1}{n}\sum_{k=1}^{n}(X_k)^2 - \bar{X}^2\right)$$
$$= \frac{1}{n}\sum_{k=1}^{n}E((X_k)^2) - E(\bar{X}^2)$$
$$= \frac{1}{n}\sum_{k=1}^{n}\left(V(X_k) + (E(X_k))^2\right) - \left(V(\bar{X} + (E(X))^2\right) \qquad \text{((15) による)}$$
$$= \frac{1}{n}\cdot n(\sigma^2+\mu^2) - \left(\frac{\sigma^2}{n}+\mu^2\right)$$
$$= \frac{n-1}{n}\sigma^2$$

**(1a) 母集団の特性値の分布が正規分布と仮定でき，母分散が判っているときに，この値と標本平均を用いて，母平均の信頼度 $t$ % の信頼区間を求める．**

このときには，仮定から各 $X_i$ は正規分布に従うから $\bar{X}$ も正規分布に従う．したがって，(12) と (14) から，

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$$

は標準正規分布 $N(0,1)$ に従う．教科書 p.201 の数表で $I(z) = \frac{1}{2}\cdot\frac{t}{100}$ となる $z$ を求めると，そのような $z$ に対し，$P\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| \leq z\right) = \frac{t}{100}$ が成り立つ．この式は，不等式,

$$\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| \leq z$$

が $t$ % の確率で成り立つことを主張しているが，$\bar{X}$ の実現値の 1 つである $\bar{x}$ もこの確率で,

$$\left|\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right| \leq z$$

を満たしていると考えてよい．ここで上の式を $\mu$ について解くと,

(17) $$\bar{x} - z\cdot\frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + z\cdot\frac{\sigma}{\sqrt{n}}$$

となるが，この不等式で与えられる $\mu$ の値の区間が，$\mu$ の信頼度 $t$ % の信頼区間である．

| $t$ | $z$ |
|---|---|
| 95 % | 1.96 |
| 99 % | 2.58 |

上の $z$ の値は教科書 p.201 の表から，上の表のとなる．

**例 1．** 10 名の日本人成人男性を無作為に選び身長を測定したところ，測定値の平均は 170.4cm だった．過去の測定では，日本人成人男性の身長の標準偏差は 6.7 cm だった．身長が正規分布に従い，この標準偏差が現在の日本人成人男性の身長に対しても有効だと仮定するとき，日本人成人男性の平均身長の信頼区間を信頼度 85 % で推測せよ．

(17) により，信頼区間は

$$170.4 - 1.96 \times \frac{6.7}{\sqrt{10}} \leq \mu \leq 170.4 + 1.96 \times \frac{6.7}{\sqrt{10}}$$

したがって,（計算結果を小数点下 3 桁以下を四捨五入して）$166.2 \leq \mu \leq 174.6$ $(cm)$ となる．

**(1b) 母集団の特性値の分布のタイプは不明だが，母分散が判っているときに，この値と十分に大きなサイズの標本の標本平均を用いて，母平均の信頼度 $t$ % の信頼区間を求める．**

$n$ が十分に大きいときには，中心極限定理により，母集団の特性値の分布のタイプにかかわらず，$\bar{X}$ の分布は近似的に $N(\mu, \sigma^2/n)$ に従う．したがって，この場合にも (1a) と同様に (17) によって信頼区間を求めることができることがわかる．

**(2) 母集団の特性値の分布が正規分布と仮定でき，母分散も未知のとき，標本平均と標本分散を用いて母平均の信頼度 $t$ % の信頼区間を求める．**

(7) と (16) から，$E(U^2) = E(\frac{n}{n-1}S^2) = \frac{n}{n-1}E(S^2) = \frac{n}{n-1} \cdot \frac{n-1}{n}\sigma^2 = \sigma^2$ となる．このことから，$u^2$ の値を母分散の代用として用いることが考えられる．$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ は $N(0,1)$ に従うが，定数 $\sigma$ を確率変数 $U$ で置き換えて得られる確率変数 $T = \dfrac{\bar{X}-\mu}{U/\sqrt{n}}$ を考えると，これは $N(0,1)$ と類似の，**自由度 $n-1$ の $t$ 分布** ($t(n-1)$) と呼ばれる分布に従う[1]．したがって，(1a) と同様に，教科書 p.202 の数表から，$t_{n-1}(\alpha) = \frac{1}{2} \cdot (1 - \frac{t}{100})$ となる $\alpha$ を求めると．$P\left(\left|\dfrac{\bar{X}-\mu}{U/\sqrt{n}}\right| \leq \alpha\right) = \dfrac{t}{100}$ が成り立つ．この式は，不等式,

$$\left|\frac{\bar{X}-\mu}{U/\sqrt{n}}\right| \leq \alpha$$

が $t$ % の確率で成り立つことを主張しているが，$\bar{X}$ と $U$ の実現値の 1 つである $\bar{x}$ と $u$ もこの確率で,

$$\left|\frac{\bar{x}-\mu}{u/\sqrt{n}}\right| \leq \alpha$$

を満たしていると考えてよい．ここで上の式を $\mu$ について解くと,

[1] $n$ が十分に大きいときには，$t(n-1)$ は $N(0,1)$ とほとんど等しくなるので，$t(n-1)$ を $N(0,1)$ で代用することができる．

$$(18) \quad \bar{x} - \alpha \cdot \frac{u}{\sqrt{n}} \leq \mu \leq \bar{x} + \alpha \cdot \frac{u}{\sqrt{n}}$$

となるが，この不等式で与えられる $\mu$ の値の区間が，$\mu$ の信頼度 $t$ % の信頼区間である．

**例 2.** （教科書の練習問題 3.2.2 (p.122)） 八重山列島のタイワンカブトムシの雌 8 匹の体長を計ったところ，35, 39, 41, 34, 44, 35, 31, 37 $(mm)$ だった．八重山列島のタイワンカブトムシの雌の平均の体長の信頼度 95% の信頼区間を求めよ．

八重山列島のタイワンカブトムシの雌の平均の体長は正規分布に従うと仮定する．このとき，測定した 8 匹の体長の平均は

$$\bar{x} = \frac{35+39+41+34+44+35+31+37}{8} = 37$$

不偏分散は，

$$u^2 = \frac{1}{8-1}\left((35-37)^2 + (39-37)^2 + \cdots + (37-37)^2\right) = 17.42857\cdots$$

したがって $u \approx 4.175$ となる．教科書 p.202 の数表から，$t_{8-1}(\frac{1}{2} \cdot (1-95)) = 2.365$ だから，(18) から，

$$37 - 2.365 \cdot \frac{4.175}{\sqrt{8}} \leq \mu \leq 37 + 2.365 \cdot \frac{4.175}{\sqrt{8}}$$

したがって，約 $33.5 \leq \mu \leq 40.5$ $(mm)$ が，信頼度 95% の信頼区間となる．

**例 3.** ある国の男子幼稚園児（5 才児）150 人を無作為抽出して身長を計ったところ，平均 110.6 $cm$, 標準偏差 5.2 $cm$ だった．この国の男子幼稚園児の平均身長を信頼度 99 % で推定してください．

この国の男子幼稚園児の平均身長は正規分布に従うと仮定して議論する．標本のサイズが十分に大きいので，$t$ 分布は正規分布で近似することができる．また標本の不偏標準偏差も $\sqrt{\frac{150}{149}} \cdot 5.2 \approx 5.2$ なので標本の標準偏差で置き換えてよい．
したがって (18) と，p.3 の表（教科書 p.201 の表から求めた値）から，

$$110.6 - 2.58 \cdot \frac{5.2}{\sqrt{150}} \leq \mu \leq 110.6 + 2.58 \cdot \frac{5.2}{\sqrt{150}}$$

これを計算して，約 $109.5 \leq \mu \leq 111.7$ $(cm)$ が信頼度 99 % の信頼区間となる．

### (3a) 母平均の検定 （離散的確率分布の場合）

**例 4. (教科書 p136〜の例の翻案)** 「4 本に 1 本は当たりくじが入っている」と主催者がうたっているくじを 11 回ひいて 11 回ともはずれくじだった．このことから，くじの主催者の言っていることが嘘であることを有意水準 0.05 で主張することができるか?

十分に大きな数のくじの中からくじを引いて，復元抽出していると考えてよいとする．つまり「4 本に 1 本は当たりくじが入っている」という状況はくじをひいた結果変らないとする．

「4 本に 1 本は当たりくじが入っている」を帰無仮説 $H_0$ として議論する．このとき，11 回ひいて 11 回ともはずれくじが出てくる確率は

$$\left(\frac{3}{4}\right)^{11} = 0.0422351360321044\cdots \; < \; 0.05$$

だから,「4 本に 1 本は当たりくじが入っている」という主催者の主張は有意水準 0.05 で棄却される．つまり，くじの主催者の言っていることが嘘であることを，有意水準 0.05 で主張できる!

### (3b) 母平均の検定 （正規分布の場合）

**例 5. (教科書 p.142〜143 の例の変形)** 従来，寿命が 1280 時間だった $N$ 社の蛍光灯をあちこちの店から計 16 本買い（無作為抽出）測定したところ，それらの平均寿命は 1205 時間で，寿命の標準偏差（標本標準偏差）は 155 時間でした．この蛍光灯の最近の製品のクオリティーは劣化したと言えるでしょうか? 有意水準 0.05 で考えてください.

このデータの不偏標準偏差は $\sqrt{\dfrac{16}{15}} \cdot 155 \approx 160.1$ 母集団 (つまり蛍光灯全製品) の寿命は正規分布に従うと仮定する．帰無仮説 $H_0$ として，寿命は短かくなっていない，と仮定してみる．つまり，$\mu \geq 1280$ である（実は製品は改良されていて平均寿命は延びているかもしれない）． $n = 16$ だから，$\dfrac{\bar{X} - \mu}{U/\sqrt{16}}$ は自由度 15 の $t$ 分布に従うが，教科書 p.202 の数表から（縦軸 15 横軸 0.050 のところを見ると 1.753 になっている），$\dfrac{\bar{x} - \mu}{u/\sqrt{16}} < -1.753$ が有意水準 0.05 での棄却域の条件であることがわかる.

$$\frac{\bar{x} - \mu}{u/\sqrt{16}} < -1.753 \quad \Leftrightarrow \quad \bar{x} < -1.753 \cdot \frac{u}{\sqrt{16}} + \mu$$

だから，$\bar{x} < -1.753 \cdot \dfrac{u}{\sqrt{16}} + 1280$ なら，ここでの標本のデータは棄却域に入ることになるが，$\bar{x}$ と $u$ に 1205 と 160.1 を代入すると，$1205 < -1.753 \cdot \dfrac{160.1}{\sqrt{16}} + 1280 \approx 1210$ となり，この等式が実際に成り立っていることが分る．したがって，$H_0$ は有意水準 0.05 で棄却され，対立仮説の「この蛍光灯の最近の製品のクオリティーは劣化した」ということが有意水準 0.05 で採択される.

## 教科書:

[1] 小寺 平治: ゼロから学ぶ統計解析，講談社, (2002).