第 VI 部　統計的検定

目次

10 t 分布による区間推定

10.1 はじめに

- 事前情報はありません。
- 使えるものは集められた情報のみです。
- それらを駆使して母集団を推定します。

10.1.1 ポイント

- t 分布
- t 分布の自由度 ν と形状
- t 分布の臨界値
- t 分布を利用した母平均の区間推定

10.1.2　ネイピア数

定義

以下で定義される無理数を e であらわし、**ネイピア数**という。

$$\mathrm{e} = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n \tag{1}$$

ネイピア数 e は底とすることが多い。そこで

$$\mathrm{e}^x = \exp(x) \tag{2}$$

とあらわす。

ネイピア数は、

$$\mathrm{e} = 2.718\,281\,828\,459\,045\,235\,36 \cdots \tag{3}$$

と続く無理数である。

10.1.3 標準正規分布

平均 0, 分散 1 の正規分布 $N(0,\ 1)$ を『**標準正規分布**』という。標準正規分布の確率密度関数は

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \tag{4}$$

である。

標準正規分布 $N(1,\ 0)$ に従う確率変数を z であらわす。

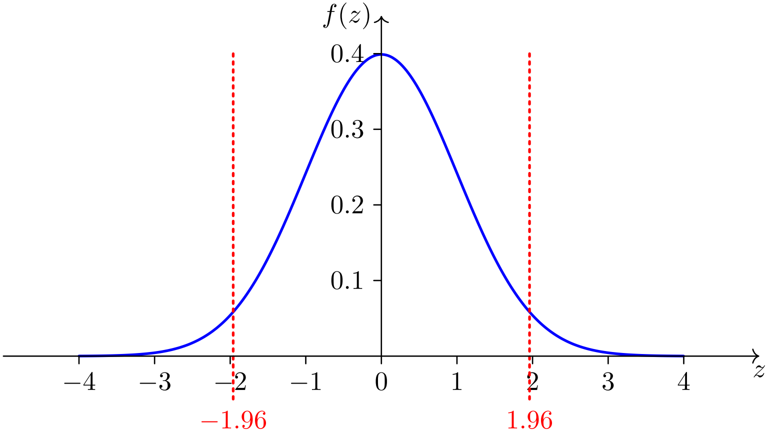
- 標準正規分布の確率密度関数は、$z = 0$ を中心とする左右対称の釣り鐘型・単峰の形状をしている。
- 確率変数 z が区間 $(-1.96, 1.96)$ に属する確率は 95%である。

$N(\mu,\ \sigma^2)$ に従う確率変数 x を**標準化**

$$z = \frac{x - \mu}{\sigma} \tag{5}$$

すると z に変換できる。

図 1　標準正規分布の確率密度関数



10.1.4 標本平均

$N(\mu,\ \sigma^2)$ に従う n 個の確率変数 x_i の平均を『**標本平均**』といい、$\bar{X}$ であらわす。標本平均 $\bar{X}$ は、平均 μ, 分散 $\dfrac{\sigma^2}{n}$ の正規分布 $N\left(\mu,\ \dfrac{\sigma^2}{n}\right)$ に従う。

標本平均を標準化した

$$z = \frac{\bar{X} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} \tag{6}$$

は標準正規分布に従う。

10.1.5 標本分散

標本平均からの偏差の二乗の平均を『**標本分散**』といい $V_{(\bar{X})}$ であらわす。

$$V_{(\bar{X})} = \frac{1}{n}\sum_{x_i \in A}^{n}\left(x_i - \bar{X}\right)^2 \tag{7}$$

標本分散の正の平方根

$$D_{(\bar{X})} = \sqrt{\frac{1}{n}\sum_{x_i \in A}^{n}\left(x_i - \bar{X}\right)^2} \tag{8}$$

は標本の『**標準偏差**』である。

10.1.6 χ^2 分布

標準正規分布 $N(0,\ 1)$ に従う n 個の独立な確率変数 $z_1,\ z_2, \cdots,\ z_n$ の平方和から統計量を作り $\chi^2_{(z)}$ であらわす。このとき、n を『**自由度**』といい ν を使ってあらわす。

$$\chi^2_{(z)} = \sum_{i=1}^{n} {z_i}^2 \tag{9}$$

$\chi^2_{(z)}$ が従う分布を『**自由度** $\nu = \mathbf{n}$ **の** χ^2 **分布**』という。

- χ^2 分布の確率密度関数は自由度 ν によって大きく形を変える。
- χ^2 分布の確率密度関数は、左右非対称の形状を持つ。
- χ^2 分布の確率密度関数の定義域は $[0,\ \infty)$
- 95%信頼区間は自由度 ν ごとに異なり、95%信頼区間の左右の臨界値もそれぞれ異なる。

図 2　χ^2 分布の確率密度関数

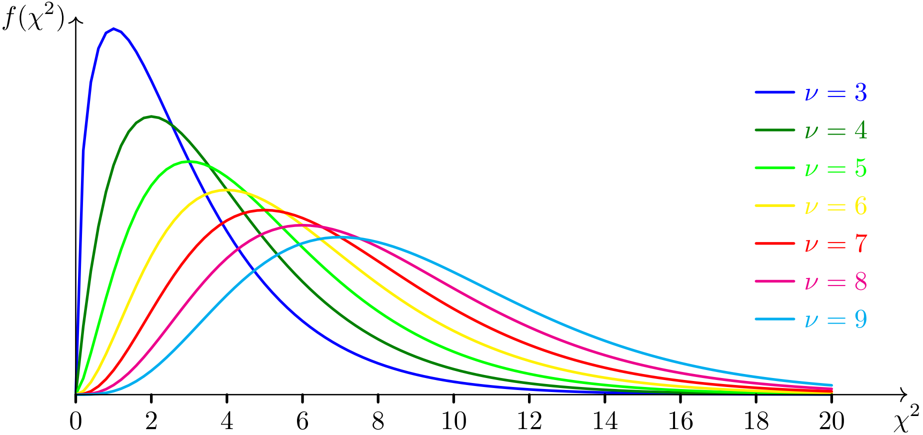


表 1　χ^2 分布の 95%信頼区間の臨界値

ν	$\chi^2_{0.025}$	$\chi^2_{0.975}$	ν	$\chi^2_{0.025}$	$\chi^2_{0.975}$	ν	$\chi^2_{0.025}$	$\chi^2_{0.975}$
1	0.001	5.024	8	2.180	17.535	60	40.482	83.298
2	0.051	7.378	9	2.700	19.023	70	48.758	95.023
3	0.216	9.348	10	3.247	20.483	80	57.153	106.629
4	0.484	11.143	20	9.591	34.170	90	65.647	118.136
5	0.831	12.833	30	16.791	46.979	100	74.222	129.561
6	1.237	14.449	40	24.433	59.342	200	162.728	241.058
7	1.690	16.013	50	32.357	71.420	500	439.936	563.852

10.1.7　χ^2統計量

正規分布 $N(\mu,\ \sigma^2)$ に従う確率変数を標準化した

$$z = \frac{x_i - \mu}{\sigma} \tag{10}$$

は標準正規分布 $N(0,\ 1)$ に従うのだから、$N(\mu,\ \sigma^2)$ から大きさ n の標本 A についての $\chi^2_{(\mu)}$ 統計量

$$\chi^2_{(\mu)} = \sum_{x_i \in A}^{n} \left(\frac{x_i - \mu}{\sigma} \right)^2 \tag{11}$$

は自由度 $\nu = n$ の χ^2 分布に従う。

μ の代わりに標本平均 $\bar{X}$ を使った統計量

$$\chi^2_{(\bar{X})} = \sum_{x_i \in A}^{n} \left(\frac{x_i - \bar{X}}{\sigma} \right)^2 \tag{12}$$

は自由度 $\nu = n - 1$ の χ^2 分布に従う。

ここで

$$
\begin{aligned}
\chi^2_{(\bar{X})} &= \frac{1}{\sigma^2}\sum_{x_i \in A}^{n}\left(x_i - \bar{X}\right)^2 && (13) \\
&= \frac{n}{\sigma^2}\frac{1}{n}\sum_{x_i \in A}^{n}\left(x_i - \bar{X}\right)^2 && (14) \\
&= \frac{nV_{(\bar{X})}}{\sigma^2} && (15)
\end{aligned}
$$

10.2　t 分布

標準正規分布 $N(0,\ 1)$ に従う確率変数 z と、これとは独立に自由度 $\nu = n$ の χ^2 に従う確率変数 $\chi^2_{(z)}$ を用いて統計量を作り $t_{(z)}$ であらわす。この統計量

$$t_{(z)} = \frac{z}{\sqrt{\frac{\chi^2_{(z)}}{\nu}}} \tag{16}$$

が従う分布を『**t 分布**』という。このとき、$\nu = n$ を t 分布の『**自由度**』という。

- t 分布は $t_{(z)} = 0$ を中心に左右対称の釣り鐘型の形状をしている。
- t 分布の形状は自由度 $\nu = n$ により異なり、自由度 $\nu = n$ が大きいときには標準正規分布によく近似する。

図 3　t 分布の確率密度関数

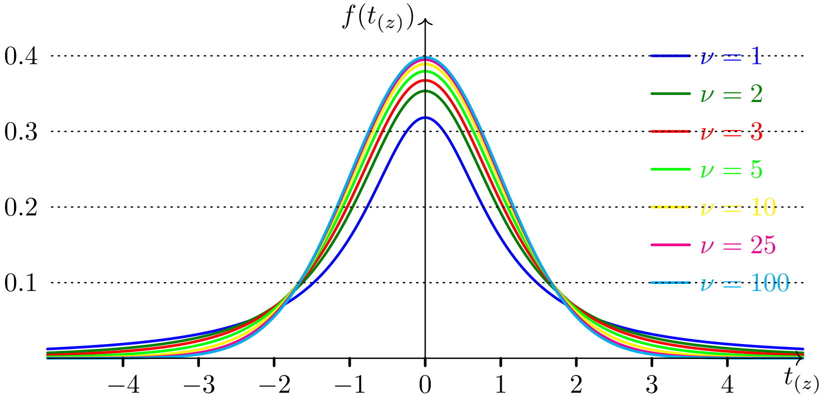


表 2 t 分布の 95 % 臨界値

ν	臨界値	ν	臨界値	ν	臨界値	ν	臨界値
5	2.570 582	12	2.178 813	19	2.093 024	30	2.042 272
6	2.446 912	13	2.160 369	20	2.085 963	40	2.021 075
7	2.364 624	14	2.144 787	21	2.079 614	50	2.008 559
8	2.306 004	15	2.131 450	22	2.073 873	100	1.983 972
9	2.262 157	16	2.119 905	23	2.068 658	500	1.964 720
10	2.228 139	17	2.109 816	24	2.063 899	1000	1.962 339
11	2.200 985	18	2.100 922	25	2.059 539	5000	1.960 439

10.3 標本分散を使った t 分布

標準化において μ に代えて標本平均 $\bar{X}$ を用いた統計量

$$\chi^2_{(\bar{X})} = \frac{nV_{(\bar{X})}}{\sigma^2} \tag{15}$$

の両辺の平方根をとり、$D_{(\bar{X})} = \sqrt{V_{(\bar{X})}}$ とすると

$$\sqrt{\chi^2_{(\bar{X})}} = D_{(\bar{X})}\frac{\sqrt{n}}{\sigma} \tag{17}$$

である。

この $\chi^2_{(\bar{X})}$ の自由度は $\nu = n - 1$ である。

標本平均を標準化した確率変数は

$$z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \tag{6}$$

$$= \left(\bar{X} - \mu\right)\frac{\sqrt{n}}{\sigma} \tag{18}$$

である。

これらを (16) に代入すると

$$t_{(\bar{X})} = \frac{z}{\sqrt{\frac{\chi^2_{(\bar{X})}}{\nu}}} \tag{16}$$

$$= z\frac{\sqrt{\nu}}{\sqrt{\chi^2_{(\bar{X})}}} \tag{19}$$

$$= \left(\bar{X} - \mu\right)\frac{\cancel{\sqrt{n}}}{\cancel{\sigma}}\frac{\sqrt{\nu}}{D_{(\bar{X})}\cancel{\frac{\sqrt{n}}{\sigma}}} \tag{20}$$

$$= \frac{\left(\bar{X} - \mu\right)\sqrt{\nu}}{D_{(\bar{X})}} \tag{21}$$

ここで $\chi^2_{(\bar{X})}$ の自由度は $\nu = n - 1$ なので

$$= \frac{\left(\bar{X} - \mu\right)\sqrt{n-1}}{D_{(\bar{X})}} \tag{22}$$

$$= \frac{\bar{X} - \mu}{\left(\frac{D_{(\bar{X})}}{\sqrt{n-1}}\right)} \tag{23}$$

は自由度 $\nu = n - 1$ の t 分布に従う。ここで

$$D_{(\bar{X})} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \bar{X}\right)^2} \tag{8}$$

である。

10.4 数値例

$\bar{X}$,$D_{(\bar{X})}$,n は、得られたデータより算出可能な統計量であり、自由度 ν の t 分布の 95 % 信頼区間の臨界値は表 2 などで得られている。

従って、正規分布に従う n 個の標本から、未知の母平均 μ の区間推定が可能なのである。

10.4.1 例題 1

正規母集団から $n = 6$ の標本を抽出したところ

$$\{76,\ 85,\ 82,\ 83,\ 76,\ 78\} \tag{24}$$

であった。母平均を 95 % の信頼区間で区間推定しなさい。

解法

$$\bar{X} = \frac{76 + 85 + 82 + 83 + 76 + 78}{6} = 80 \tag{25}$$

$$V_{(\bar{X})} = \frac{(-4)^2 + (5)^2 + (2)^2 + (3)^2 + (-4)^2 + (-2)^2}{6} = 12.33 \tag{26}$$

$$D_{(\bar{X})} = \sqrt{12.33} = 3.51 \tag{27}$$

$$\nu = 6 - 1 = 5 \tag{28}$$

表 2 より自由度 $\nu = 5$ の 95%信頼区間は、

$$-2.570582 \leqq t_{(\bar{X})} \leqq 2.570582 \tag{29}$$

である。

従って、$t_{(\bar{X})}$ の 95%信頼区間は、

$$-2.570582 \leqq \frac{(80-\mu)\sqrt{5}}{3.51} \leqq 2.570582 \tag{30}$$

なので、母平均 μ の 95%信頼区間は

$$75.964 \leqq \mu \leqq 84.036 \tag{31}$$

である。

問題 VI−**10 − 1**

正規母集団から以下の標本を抽出した。

$$\{116.5,\ 122.6,\ 128.5,\ 133.5,\ 138.6,\ 145.0,\ 116.2,\ 121.7,\ 127.6,\ 133.3,\ 140.4,\ 146.7\}$$

母平均を 95 % の信頼区間で区間推定しなさい。なお、有効桁数は各自の判断によるものとする。

解例 VI−**10 − 1**

$$\bar{X} = 130.9$$

$$V_{(\bar{X})} = 100.345$$

$$D_{(\bar{X})} = 10.017$$

$$\nu = 12 - 1 = 11$$

表 2 より自由度 $\nu = 11$ の 95%信頼区間は、

$$-2.200985 \leqq t_{(\bar{X})} \leqq 2.200985$$

なので

$$-2.200985 \leqq \frac{(130.9 - \mu) \times \sqrt{11}}{10.017} \leqq 2.200985$$

従って、母平均 μ の 95%信頼区間は

$$124.2 \leqq \mu \leqq 137.5$$

問題 VI$-$**10 − 2**

正規母集団から以下の標本を抽出した。

$$\{21.3,\ 24.2,\ 27.7,\ 31.1,\ 34.4,\ 39.2,\ 21.2,\ 23.7,\ 27.0,\ 30.4,\ 35.0,\ 40.0\}$$

母平均を 95 % の信頼区間で区間推定しなさい。なお、有効桁数は各自の判断によるものとする。

解例 VI$-\mathbf{10-2}$

$$\bar{X} = 29.6$$

$$V_{(\bar{X})} = 39.100$$

$$D_{(\bar{X})} = 6.253$$

$$\nu = 12 - 1 = 11$$

$t_{(\bar{X})}$ の 95%信頼区間は、

$$-2.200985 \leqq \frac{(29.6 - \mu) \times \sqrt{11}}{6.253} \leqq 2.200985$$

従って、母平均 μ の 95%信頼区間は

$$25.5 \leqq \mu \leqq 33.7$$

問題 VI**−10 − 3**

正規母集団から以下の標本を抽出した。

$$\{11.6,\ 10.7,\ 10.0,\ 9.7,\ 9.4,\ 9.0,\ 11.8,\ 10.9,\ 10.4,\ 10.0,\ 9.6,\ 9.3\}$$

母平均を 95 % の信頼区間で区間推定しなさい。なお、有効桁数は各自の判断によるものとする。

解例 VI**−10 − 3**

$$\bar{X} = 10.2$$

$$V_{(\bar{X})} = 0.740$$

$$D_{(\bar{X})} = 0.860$$

$$\nu = 12 - 1 = 11$$

$t_{(\bar{X})}$ の 95%信頼区間は、

$$-2.200985 \leqq \frac{(10.2 - \mu) \times \sqrt{11}}{0.860} \leqq 2.200985$$

従って、母平均 μ の 95%信頼区間は

$$9.6 \leqq \mu \leqq 10.8$$

10.5　まとめ

- 標準正規分布 $N(0,\ 1)$ に従う確率変数 z と、独立な自由度 $\nu = n$ の $\chi^2_{(z)}$ を用いた統計量

$$t_{(z)} = \frac{z}{\sqrt{\frac{\chi^2_{(z)}}{\nu}}}$$

  が従う分布が『t 分布』であ る。$\nu = n$ を「t 分布の『自由度』」という。
- t 分布の確率密度関数の形状は自由度 ν によって異なる。従って臨界値は自由度ごとに異なる。
- t 分布の確率密度関数の形状は自由度 ν が十分大きいとき標準正規分布に近似する。
- t 分布を利用することで標本から母平均の区間推定を行うことが可能である。

参考文献

- 『統計学入門』東京大学教養学部統計学教室 篇「東京大学出版会」
- 「行動科学における統計解析法」芝祐順　南風原朝和　著 『東京大学出版会』1990 年 3 月 20 日　初版
- 「完全独習　統計学入門」小島　寛之　著 『ダイヤモンド社』2006 年 9 月 28 日　初版
- 『すぐわかる　確率・統計』石村園子・畑宏明著「東京都書」
- 『データ解析のための数理統計入門』久保川達也著「共立出版」
- 『技術者のための高等数学 7 確率と統計（原書代 8 版）』E・クライツィグ著　田栗正章訳「培風館」
- 『経済数学教室　各巻』小山昭雄著 「岩波書店」

- 『新体系　大学数学　入門の教科書　上下』吉沢光雄著「講談社」