2　区間推定

2.1　母平均の区間推定（母分散が既知の場合）

上側 $\boldsymbol{\alpha}$ 点　X は連続型確率変数，$0 < \alpha < 1$ とする。次の式を満たす Q_α のことを上側 α 点または上側 100α パーセント点という。

$$P\left(X < Q_\alpha\right) = 1 - \alpha, \qquad P\left(X > Q_\alpha\right) = \alpha$$

標準正規分布の上側 α 点を z_α と表す。その値は標準正規分布表から見つけることができ，上側 2.5 パーセント点は 1.96，上側 0.5 パーセント点は 2.58 である。

$$\begin{aligned} P\left(X > 1.96\right) &= 0.025, \qquad z_{0.025} = 1.96 \\ P\left(X > 2.58\right) &= 0.005, \qquad z_{0.005} = 2.58 \end{aligned} \tag{1}$$

	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	**.4750**	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	**.4951**	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974

正規母集団の標本平均　母集団分布が正規分布に従うとき，その母集団のことを正規母集団という。正規母集団において，母平均を μ，母分散を σ^2 とし，この母集団から取り出した大きさ n の標本の標本平均を $\bar{X}$ とする。1.2 節の結果から，

$$E\left[\bar{X}\right] = \mu, \qquad V\left[\bar{X}\right] = \frac{\sigma^2}{n}, \qquad \sigma\left[\bar{X}\right] = \frac{\sigma}{\sqrt{n}}$$

$\bar{X}$ の標準化を Z とすると，Z は標準正規分布に従う。

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0, 1)$$

正規母集団でない場合においても，n が十分大きいならば，中心極限定理により，Z は近似的に標準正規分布に従う。

母平均の区間推定（母分散が既知の場合）　標準正規分布に従う Z について

$$P\left(-z_{\alpha/2} \leqq Z \leqq z_{\alpha/2}\right) = 1 - \alpha$$

が成り立ち，特に $\alpha = 0.05$ のとき，$z_{\alpha/2} = z_{0.05/2} = 1.96$，$\alpha = 0.01$ のとき，$z_{\alpha/2} = z_{0.01/2} = 2.58$ であるから，次のようになる。

$$P\left(-1.96 \leqq Z \leqq 1.96\right) = 0.95, \qquad P\left(-2.58 \leqq Z \leqq 2.58\right) = 0.99$$

Z を $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ に置き換え，

$$P\left(-z_{\alpha/2} \leqq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leqq z_{\alpha/2}\right) = 1 - \alpha$$

さらに母平均 μ について解くと，次式が得られる。

$$P\left(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \leqq \mu \leqq \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha \tag{2}$$

$1 - \alpha$ を信頼度または信頼係数といい，次の区間のことを信頼度 $1 - \alpha$ の**信頼区間**
*i （confidence interval）あるいは簡単に $100(1 - \alpha)\%$ 信頼区間という。[*i]

$$\left[\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \quad \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right] \tag{3}$$

特に 95% 信頼区間，99% 信頼区間は各々

$$\left[\bar{x} - 1.96\frac{\sigma}{\sqrt{n}}, \quad \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right], \qquad \left[\bar{x} - 2.58\frac{\sigma}{\sqrt{n}}, \quad \bar{x} + 2.58\frac{\sigma}{\sqrt{n}}\right]$$

となる。この公式を用いる場合は，母分散 σ^2 は既知でなければならない。

2.2　母平均の区間推定（母分散が未知の場合）

***t* 分布**　$Z_1, Z_2, \ldots, Z_n$ が互いに独立で，標準正規分布に従うとき，Y を

$$Y = Z_1^2 + Z_2^2 + \cdots + Z_n^2$$

と定める。Y が従う確率分布のことを自由度 n のカイ 2 乗分布（χ^2-distribution）という。次に，Z と Y は互いに独立で，Z が標準正規分布，Y が自由度 n のカイ 2 乗分布に従うとき，T を

$$T = \frac{Z}{\sqrt{Y/n}}$$

と定める。T が従う確率分布のことを自由度 n の t 分布（Student's t-distribution）
*ii という [*ii]。$n \to \infty$ のとき，自由度 n の t 分布は標準正規分布に限りなく近づく。

上側 $\boldsymbol{\alpha}$ 点と両側 $\boldsymbol{\alpha}$ 点　原点について対称な分布（標準正規分布，t 分布）に対し，次の式を満たす Q_α のことを両側 α 点または両側 100α パーセント点という。

$$P\left(|X| < Q_\alpha\right) = 1 - \alpha, \qquad P\left(|X| > Q_\alpha\right) = \alpha$$

*iii ここでは t 分布の両側 α 点を $t_{\alpha/2}(n)$ と表すことにする。[*iii]

[*i] 実数の集合 $[a, b] = \{x \mid a \leqq x \leqq b\}$ や $(a, b) = \{x \mid a < x < b\}$ を区間という。
[*ii] 本稿では，χ^2 分布や t 分布の詳細にふれない。
[*iii] 上側 α 点，両側 α 点の表記は，文献により様々である。

t 分布表（両側 α 点）　次表を用いて，自由度 n と α から両側 α 点 $t_{\alpha/2}(n)$ を求めることができる。[*iv]。 [*iv]

$n \setminus \alpha$	0.05	0.01	$n \setminus \alpha$	0.05	0.01	$n \setminus \alpha$	0.05	0.01
1	12.71	63.66	16	2.120	2.921	31	2.040	2.744
2	4.303	9.925	17	2.110	2.898	32	2.037	2.738
3	3.182	5.841	18	2.101	2.878	33	2.035	2.733
4	2.776	4.604	19	2.093	2.861	34	2.032	2.728
5	2.571	4.032	20	2.086	2.845	35	2.030	2.724
6	2.447	3.707	21	2.080	2.831	40	2.021	2.704
7	2.365	3.499	22	2.074	2.819	50	2.009	2.678
8	2.306	3.355	23	2.069	2.807	60	2.000	2.660
9	2.262	3.250	24	2.064	2.797	70	1.994	2.648
10	2.228	3.169	25	2.060	2.787	80	1.990	2.639
11	2.201	3.106	26	2.056	2.779	90	1.987	2.632
12	2.179	3.055	27	2.052	2.771	100	1.984	2.626
13	2.160	3.012	28	2.048	2.763	110	1.982	2.621
14	2.145	2.977	29	2.045	2.756	120	1.980	2.617
15	2.131	2.947	30	2.042	2.750	∞	1.960	2.576

母平均の区間推定（母分散が未知の場合）　正規母集団において，母平均を μ，母分散を σ^2 とし，母分散は未知とする。この母集団から取り出した大きさ n の標本の標本平均を $\bar{X}$，不偏分散を U^2 とする。

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad U^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

次の式で定める T は自由度 $n-1$ の t 分布に従うことが知られている。

$$T = \frac{\bar{X}-\mu}{\frac{U}{\sqrt{n}}} \sim t(n-1)$$

この T について次が成り立つ。

$$P\left(-t_{\alpha/2}(n-1) \leqq T \leqq t_{\alpha/2}(n-1)\right) = 1-\alpha$$

よって，母分散 σ^2 が未知の場合の母平均 μ の信頼区間は次のようになる。

$$\left[\bar{x} - t_{\alpha/2}(n-1)\frac{u}{\sqrt{n}}, \quad \bar{x} + t_{\alpha/2}(n-1)\frac{u}{\sqrt{n}}\right] \tag{4}$$

2.3　母比率の区間推定

二項母集団　母集団の各要素が，A または $\bar{A}$ のいずれかであるとき，その母集団のことを二項母集団という。ある特性を満たす要素は A，その特性を満たさない要素は $\bar{A}$ と考えればよい。母集団の中の A の割合（特性を満たす要素の割合）のことを**母比率**，標本の中の A の割合のことを**標本比率**という。

[*iv] $n=\infty$ の欄は，標準正規分布のパーセント点と同じである。$t_{\alpha/2}(\infty) = z_{\alpha/2}$

母集団分布　母比率 p の二項母集団から 1 個の要素を取り出し，その要素が A なら $X = 1$，$\bar{A}$ なら $X = 0$ とすると，X は Bernoulli 分布に従う。

$X:$	0	1
$P:$	$1-p$	p

4.1 節の結果から，母平均 μ は p，母分散 σ^2 は $p(1-p)$ である。

母比率の区間推定　母比率 p の二項母集団から大きさ n の標本を取り出し，その要素が A なら $X_i = 1$，$\bar{A}$ なら $X_i = 0$ とすると，$X_1, X_2, \ldots, X_n$ の標本平均 $\bar{X}$ は標本比率に等しい。標本比率を $\hat{P}$ とする。1.2 節の結果から，

$$E\left[\hat{P}\right] = \mu = p, \qquad \sigma\left[\hat{P}\right] = \frac{\sigma}{\sqrt{n}} = \sqrt{\frac{p(1-p)}{n}}$$

n が十分大きいとき，$\hat{P}$ の標準化を Z とすると，Z は近似的に標準正規分布に従う。

$$Z = \frac{\hat{P} - p}{\sqrt{p(1-p)/n}} \sim N(0,1)$$

母平均の区間推定（母分散が既知の場合）と同様に，次式が成り立つ。

$$P\left(-z_{\alpha/2} \leqq \frac{\hat{P} - p}{\sqrt{p(1-p)/n}} \leqq z_{\alpha/2}\right) = 1 - \alpha$$

Z の分子にある p について解き，次のように変形する。

$$P\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}} \leqq p \leqq \hat{P} + z_{\alpha/2}\sqrt{\frac{p(1-p)}{n}}\right) = 1 - \alpha$$

母比率 p を標本比率 $\hat{P}$ に置き換えると，次の結果が得られる。

$$P\left(\hat{P} - z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}} \leqq p \leqq \hat{P} + z_{\alpha/2}\sqrt{\frac{\hat{P}(1-\hat{P})}{n}}\right) = 1 - \alpha$$

標本の大きさ n が十分大きいとき，母比率 p の信頼区間は次のようになる。

$$\left[\hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \quad \hat{p} - z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] \tag{5}$$

参考文献

統計学入門（基礎統計学）
東京大学教養学部統計学教室（編）東京大学出版会 978-4-13-042065-5

統計学
久保川 達也 （著） 東京大学出版会 978-4-13-062921-8

はじめての統計学
道家 暎幸（共著） コロナ社 978-4-339-06113-0

確率統計 新版 （新版数学シリーズ）
岡本 和夫 （ほか著） 実教出版 978-4-407-32171-5

統計学序論 改訂版
山本 義郎 （著） 東海大学出版部 978-4-486-02133-9

確率統計（高専テキストシリーズ）
上野 健爾（監修） 森北出版 978-4-627-05561-2

基本統計学 第４版
宮川 公男 （著） 有斐閣 978-4-641-16455-0

新統計入門
小寺 平治 （著） 裳華房 978-4-7853-1099-8

Schaum's Outline of Introduction to Probability and Statistics
Seymour Lipschutz （著） McGraw-Hill Education 978-0-07-176249-6

A Dictionary of Statistics
Graham Upton （著） Oxford Univ Pr 978-0-19-967918-8

www5e.biglobe.ne.jp/~emm386/statistics/