

2章 母集団と標本

2.1. 母集団と標本

- 母集団: 統計の対象となる集まり

例: 日本の20歳の男子全体, 工場で一定期間に製造された製品の全体

- 個体: 母集団を構成する個々のもの
- 母集団分布 (母分布): 母集団が従う確率分布 ← 推測統計学では母集団分布は仮定する

例: 正規母集団, 二項母集団, ……

- 母数 (パラメタ): 母分布を特徴付ける数値

例: 母平均, 母分散, 母比率, ……

- 統計学における調査

- 全数調査: 母集団に含まれる個体全部について調べる
- 標本調査: 全体的な特徴が母集団と“よく似た”一部分を母集団から取り出す

全数調査は

- 短時間に分析したい (時間的制約)
- 費用を少なくしたい (経済的制約)
- 無限に大きい集団や単に仮想的に考えられた母集団 (原理的制約)

などにより, 通常は行われない.

- 記述統計学: たくさんのデータを度数分布に整理したり, 平均値や分散を求めることで, データ自体の特徴を知ることが目的とする統計学 (pp.53~60の2.3節, 2.4節を参照)
- 推測統計学: 確率論の考えを用いて, 小数の標本から母集団の統計的特徴を推測することを目的とする統計学

2.2. 標本変量と統計量

母集団に“よく似た”統計的特徴をもつ標本の抽出

↓

母集団から何の作為もなく, でたらめに個体を抜き出して, その特性を調べたあと, その個体をもとの集団に戻す (復元抽出). この操作を何回か繰り返す

↓

i 番目に取り出される個体の特性を X_i ($i = 1, 2, \dots, n$) とすると, X_1, X_2, \dots, X_n は, 同じ母分布に従う独立な確率変数. 各 X_i のことを**標本変量**, (X_1, X_2, \dots, X_n) を大きさ (サイズ) n の**標本**という.

また, 実際に抽出して得られた値の組 (x_1, x_2, \dots, x_n) を標本 (X_1, X_2, \dots, X_n) の**実現値**あるいは**標本値**という.

- 統計量: 標本変量 X_1, X_2, \dots, X_n から計算される量

3章 推定法

3.3. 区間推定法-母平均の推定 I

3.3.2. 母分散が既知の正規母集団の母平均の区間推定

[理論] 母分布が $N(\mu, \sigma^2)$ で σ^2 は既知とする. 標本 (X_1, \dots, X_n) において

$$\bar{X} := \frac{1}{n}(X_1 + \dots + X_n): \text{標本平均} \quad \leftarrow \text{母平均 } \mu \text{ に対する統計量}$$

$$\therefore \bar{X} \sim N(\mu, \sigma^2/n) \quad \leftarrow \text{定理 1.8.4}$$

$$\therefore Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \quad \leftarrow \text{定理 1.8.2}$$

- 与えられた $0 < \alpha < 1/2$ に対して, $\Phi(z) = 0.5 - \alpha$ となる点 z , すなわち, 下図の斜線部分の面積が α となる z の値を $z(\alpha)$ とかき, 標準正規分布の α 点という.

各 α に対する $z(\alpha)$ の値は正規分布表 II を用いて求めることができる. 次の値がよく用いられる: $z(0.05) = 1.6449$, $z(0.025) = 1.96$, $z(0.005) = 2.5758$.

さて, $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ なので, 正規分布表 II より $z(\alpha/2)$ の値を求めると

$$P(|Z| < z(\alpha/2)) = 1 - \alpha$$

が成り立つ. それゆえ

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

上式は“標本を抽出してその平均 \bar{X} を得たとき、未知の母平均 μ が

$$\bar{X} - \frac{\sigma}{\sqrt{n}} z(\alpha/2) < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z(\alpha/2)$$

となる確率が $1-\alpha$ である”ことを示している。このとき、 $1-\alpha$ のことを信頼係数 (信頼度) といい、通常は $100(1-\alpha)\%$ と表す。また、上式で定まる μ の区間を母平均 μ の $100(1-\alpha)\%$ 信頼区間という。

[公式: 母平均の区間推定 (母分散既知)] 母分散 σ^2 が既知の正規母集団 $N(\mu, \sigma^2)$ から抽出した大きさ n の標本 (X_1, \dots, X_n) の標本平均を \bar{X} とすると、母平均 μ の $100(1-\alpha)\%$ 信頼区間は

$$\bar{X} - \frac{\sigma}{\sqrt{n}} z(\alpha/2) < \mu < \bar{X} + \frac{\sigma}{\sqrt{n}} z(\alpha/2)$$

例 1. (p.91) 某大学某学年の学生の身長は正規分布 $N(\mu, (5.54)^2)$ に従うとする。この母集団からの大きさ 50 の標本の平均値は 169.8cm であった。この正規母集団の平均身長 μ の 95% 信頼区間を求めよ。

(解) 題意より、 $n = 50$, $\sigma = 5.54$, $\bar{X} = 169.8$, $\alpha = 0.05$. 正規分布表 II より $z(\alpha/2) = z(0.025) = 1.96$. これらを信頼区間に代入:

$$169.8 - \frac{5.54}{\sqrt{50}} \times 1.96 < \mu < 169.8 + \frac{5.54}{\sqrt{50}} \times 1.96 \quad \therefore \quad 168.3 < \mu < 171.3$$

□

95%信頼区間の意味: 上の例 1 の場合で説明すると、母集団から大きさ 50 の標本を抽出し、その 95% 信頼区間を計算する作業を仮に 100 回繰り返して、100 個の信頼区間を求めると、その中の 95 個の信頼区間は母平均 μ を含んでいるが、残りの 5 個の信頼区間は μ を含んでいない可能性があることを意味している。

3.4. 母平均の区間推定 II—母集団または母分散が未知の場合

3.4.1. 大きい標本による母平均の区間推定

[理論] 大きい標本 ($n \geq 50$) の場合には、3.3.2 の公式で

- 母分布は必ずしも正規分布でなくてもよい ← 中心極限定理 (定理 1.8.5)
- 母分散 σ^2 が未知でも、それを標本分散 S^2 の実現値で近似してよい ← 標本分散の一致性 (以下の説明を参照)

標本 (X_1, \dots, X_n) において、母分散 σ^2 および母標準偏差 σ に対する統計量を

- 標本分散: $S^2 := \frac{1}{n-1} \{(X_1 - \bar{X})^2 + \dots + (X_n - \bar{X})^2\}$
- 標本標準偏差: $S := \sqrt{S^2}$

で定める。

定理 2.2.1. (標本分散計算の簡便公式) 以下の公式が成り立つ。

$$S^2 = \frac{1}{n-1} \{(X_1^2 + \dots + X_n^2) - n\bar{X}^2\}$$

(証明)

$$\begin{aligned}(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2 &= X_1^2 + \cdots + X_n^2 - 2(X_1 + \cdots + X_n)\bar{X} + n\bar{X}^2 \\ &= X_1^2 + \cdots + X_n^2 - 2n\bar{X}^2 + n\bar{X}^2 \\ &= X_1^2 + \cdots + X_n^2 - n\bar{X}^2\end{aligned}$$

□

定理 2.5.5. (標本分散の一致性) 母分散が σ^2 の母集団から抽出した大きさ n の標本の標本分散 S^2 の期待値は $E(S^2) = \sigma^2$.

(証明) (X_1, \dots, X_n) を大きさ n の標本とする. 母平均を μ とすると, $E(X_i) = \mu$ ($i = 1, 2, \dots, n$). 仮定より $V(X_i) = \sigma^2$ ($i = 1, 2, \dots, n$).

$$\therefore E(X_i^2) = V(X_i) + \{E(X_i)\}^2 = \sigma^2 + \mu^2 \quad (1)$$

一方, $\bar{X} = (X_1 + \cdots + X_n)/n$ で, X_1, \dots, X_n は互いに独立なので

$$\begin{aligned}E(\bar{X}) &= \frac{1}{n}\{E(X_1) + \cdots + E(X_n)\} = \frac{1}{n}(n\mu) = \mu \\ V(\bar{X}) &= \frac{1}{n^2}\{V(X_1) + \cdots + V(X_n)\} = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n}\end{aligned}$$

よって

$$E(\bar{X}^2) = V(\bar{X}) + \{E(\bar{X})\}^2 = \frac{\sigma^2}{n} + \mu^2 \quad (2)$$

さらに定理 2.2.1 より

$$E(S^2) = \frac{1}{n-1}\{E(X_1^2) + \cdots + E(X_n^2) - nE(\bar{X}^2)\} \quad (3)$$

(1), (2) を (3) に代入すると

$$E(S^2) = \frac{1}{n-1}\left\{n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)\right\} = \frac{1}{n-1}(n-1)\sigma^2 = \sigma^2$$

□

定理 2.5.5 の証明を見ると, 標本分散の定義式で, n ではなく $n-1$ で割った理由が理解できる. また, この定理により, 標本分散は母分散を推定する際の統計量として利用できることがわかる.

[公式: 母平均の区間推定 (母集団または母分散未知・大標本)] (必ずしも正規とは限らない) 母集団からの大標本 n (≥ 50) の標本平均を \bar{X} , 標本分散を S^2 とすると, 母平均 μ の $100(1-\alpha)\%$ 信頼区間は

$$\bar{X} - \frac{S}{\sqrt{n}} \cdot z(\alpha/2) < \mu < \bar{X} + \frac{S}{\sqrt{n}} \cdot z(\alpha/2)$$

例 2. (p.94) 林りんご園では自家で収穫したりんご“ふじ”を 120 個無作為に抽出し, その重量を調べたところ, 平均 348g, 標準偏差 23g であった. この農園でとれる“ふじ”の平均

重量を 99% の信頼度で区間推定せよ。

(解) 母分散が未知であるが大標本 ($n = 120$)。題意より, $\bar{X} = 348$, $S = 23$, $\alpha = 0.01$ 。また正規分布表 II より $z(\alpha/2) = z(0.005) = 2.5758$ 。これらを信頼区間に代入:

$$348 - \frac{23}{\sqrt{120}} \times 2.5758 < \mu < 348 + \frac{23}{\sqrt{120}} \times 2.5758 \quad \therefore \quad 342.6 < \mu < 353.4$$

□

3.5 母分散の区間推定

[理論: χ^2 分布]

2.6.1. χ^2 分布の基本的性質

$X_1, X_2, \dots, X_n \sim N(0, 1)$ は互いに独立とする。このとき, 確率変数

$$\chi^2 := X_1^2 + X_2^2 + \dots + X_n^2$$

が従う分布を自由度 n の χ^2 分布といい, その確率密度関数は

$$p_{\chi_n^2}(x) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2} & (x > 0) \\ 0 & (x \leq 0) \end{cases}$$

で与えられることが知られている。ただし

$$\Gamma(s) := \int_0^\infty e^{-x} x^{s-1} dx \quad (s > 0)$$

はガンマ関数である。

- χ^2 分布の確率密度関数の形は, その自由度 n とともに変わる (p.64 の図 2.6.1 参照)。
- 与えられた数値 $0 < \alpha < 1$ に対して

$$P(\chi^2 \geq x) = \alpha$$

となる x を $\chi_n^2(\alpha)$ とかき, χ^2 分布の α 点という。 α 点の値は χ^2 分布表 (付表 4, p.194 参照) を用いて求める。

例 1. (χ^2 分布の α 点の求め方) χ^2 分布表を用いて次の a, b の値を求めよ.

(1) χ^2 が自由度 10 の χ^2 分布に従うとき, $P(\chi^2 \geq a) = 0.1$.

(2) χ^2 が自由度 15 の χ^2 分布に従うとき, $P(\chi^2 < 25.0) = b$.

(解) (1) χ^2 分布表より $a = \chi_{10}^2(0.1) = 15.99$. (2) $\chi_{15}^2(\alpha) = 25.0$ となる α を χ^2 分布表から見つけると $\alpha = 0.05$. よって上図より, $b = 1 - \alpha = 1 - 0.05 = 0.95$ \square

定理 2.6.1. (χ^2 分布の期待値と分散) 自由度 n の χ^2 分布に従う確率変数 χ^2 の期待値と分散は $E(\chi^2) = n$, $V(\chi^2) = 2n$.

(証明) 期待値の定義より

$$E(\chi^2) = \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^\infty x \cdot x^{n/2-1} e^{-x/2} dx$$

ここで

$$\begin{aligned} \int_0^\infty x \cdot x^{n/2-1} e^{-x/2} dx &= \int_0^\infty (2y)^{n/2} e^{-y} (2dy) \quad (y = x/2 \text{ とおく}) \\ &= 2^{n/2+1} \int_0^\infty y^{(n+2)/2-1} e^{-y} dy = 2^{n/2+1} \Gamma\left(\frac{n+2}{2}\right) \\ &= 2^{n/2} \cdot 2 \cdot \frac{n}{2} \Gamma\left(\frac{n}{2}\right) \end{aligned}$$

$$\therefore E(\chi^2) = \frac{1}{2^{n/2}\Gamma(n/2)} \cdot 2^{n/2} \cdot n\Gamma(n/2) = n$$

次に

$$E((\chi^2)^2) = \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^\infty x^2 \cdot x^{n/2-1} e^{-x/2} dx$$

ここで

$$\int_0^\infty x^2 \cdot x^{n/2-1} e^{-x/2} dx = \int_0^\infty (2y)^{n/2+1} e^{-y} (2dy) \quad (y = x/2 \text{ とおく})$$

$$\begin{aligned}
&= 2^{n/2+2} \int_0^\infty y^{n/2+1} e^{-y} dy = 2^{n/2+2} \int_0^\infty y^{(n+4)/2-1} e^{-y} dy \\
&= 2^{n/2} \cdot 4\Gamma\left(\frac{n+4}{2}\right) = \frac{n+2}{2} \Gamma\left(\frac{n+2}{2}\right) = \frac{n+2}{2} \cdot \frac{n}{2} \Gamma\left(\frac{n}{2}\right) \\
&= 2^{n/2} n(n+2) \Gamma\left(\frac{n}{2}\right) \\
\therefore E((\chi^2)^2) &= \frac{1}{2^{n/2} \Gamma(n/2)} \cdot 2^{n/2} \cdot n(n+2) \Gamma\left(\frac{n}{2}\right) = n(n+2) \\
\therefore V(\chi^2) &= E((\chi^2)^2) - \{E(\chi^2)\}^2 = n^2 + 2n - n^2 = 2n
\end{aligned}$$

□

定理 2.6.3. (自由度 n の χ^2 分布) (X_1, X_2, \dots, X_n) を正規母集団 $N(\mu, \sigma^2)$ から抽出された大きさ n の標本とすると

$$\chi^2 := \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2$$

は自由度 n の χ^2 分布に従う。

定理 2.6.4. (標本分散の分布) (X_1, X_2, \dots, X_n) を正規母集団 $N(\mu, \sigma^2)$ から抽出された大きさ n の標本とし、その標本平均を

$$\bar{X} := \frac{1}{n} \sum_{i=1}^n X_i,$$

標本分散を

$$S^2 := \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

とすると

$$\chi^2 := \frac{(n-1)S^2}{\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2$$

は自由度 $n-1$ の χ^2 分布に従う。すなわち、母平均 μ が未知の場合は、自由度が1つ下がる。

3.5.1 母平均が既知の場合

[理論] (X_1, X_2, \dots, X_n) を正規母集団 $N(\mu, \sigma^2)$ (μ は既知) から抽出した大きさ n の標本とすると

$$\chi^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \text{自由度 } n \text{ の } \chi^2 \text{ 分布} \quad \leftarrow \text{定理 2.6.3}$$

χ^2 分布表より $\chi_n^2(\alpha/2)$ と $\chi_n^2(1-\alpha/2)$ の値を求めると

$$P(\chi_n^2(1-\alpha/2) < \chi^2 < \chi_n^2(\alpha/2)) = 1 - \alpha$$

が成り立つ。よって

$$P\left(\chi_n^2(1-\alpha/2) < \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 < \chi_n^2(\alpha/2)\right) = 1 - \alpha$$

上式は“標本を抽出して $\sum_{i=1}^n (X_i - \mu)^2$ を得たとき、未知の分散 σ^2 が

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_n^2(\alpha/2)} < \sigma^2 < \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_n^2(1 - \alpha/2)}$$

となる確率が $1 - \alpha$ である”ことを示している。

[公式：母分散の区間推定 (母平均既知)] 正規母集団 $N(\mu, \sigma^2)$ (μ は既知) から抽出した大きさ n の標本を (X_1, \dots, X_n) とすると、母分散 σ^2 の $100(1 - \alpha)\%$ 信頼区間は

$$\frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_n^2(\alpha/2)} < \sigma^2 < \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_n^2(1 - \alpha/2)}$$

例 1. (p.97) 風邪薬 1 錠中のある成分は平均が 2.50mg になるように製造される。この錠剤 10 錠についてその成分の含有量を調べたところ次のデータを得た。この成分含有量の分散の 98% 信頼区間を求めよ。この含有量は正規分布に従うとみなせるとする。

2.46, 2.51, 2.52, 2.48, 2.49, 2.50, 2.54, 2.53, 2.49, 2.52 (mg)

(解) 自由度 $n = 10$, $\alpha = 0.02$ なので χ^2 分布表より $\chi_{10}^2(\alpha/2) = \chi_{10}^2(0.01) = 23.2$, $\chi_{10}^2(1 - \alpha/2) = \chi_{10}^2(0.99) = 2.56$. 与えられたデータと $\mu = 2.50$ より, $\sum_{i=1}^{10} (X_i - \mu)^2 = 0.0056$. これらを信頼区間に代入:

$$\therefore \frac{0.0056}{23.2} < \sigma^2 < \frac{0.0056}{2.56} \quad \therefore \quad 0.000241 < \sigma^2 < 0.00219$$

□

3.5.2. 母平均が未知の場合

[理論] μ が未知なので μ を \bar{X} で近似すると

$$\chi^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n-1}{\sigma^2} S^2 \sim \text{自由度 } n-1 \text{ の } \chi^2 \text{ 分布} \quad \leftarrow \text{定理 2.6.4}$$

[公式: 母分散の区間推定 (母平均未知)] 正規母集団 $N(\mu, \sigma^2)$ (μ は未知) から抽出した大きさ n の標本 (X_1, \dots, X_n) の標本平均を \bar{X} , 標本分散を S^2 とすると, 母分散 σ^2 の $100(1-\alpha)\%$ の信頼区間は

$$\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{n-1}^2(\alpha/2)} < \sigma^2 < \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\chi_{n-1}^2(1-\alpha/2)}$$

または

$$\frac{(n-1)S^2}{\chi_{n-1}^2(\alpha/2)} < \sigma^2 < \frac{(n-1)S^2}{\chi_{n-1}^2(1-\alpha/2)}$$

例 2. (p.98) 例 1 において母平均が 2.5mg であることを知らなかったとすれば, 母分散の 98%信頼区間はどうか.

(解) 自由度は $10 - 1 = 9$, $\alpha = 0.02$ なので, χ^2 分布表より $\chi_9^2(\alpha/2) = \chi_9^2(0.01) = 21.7$, $\chi_9^2(1-\alpha/2) = \chi_9^2(0.99) = 2.09$. また, $\bar{X} = 2.504$, $\sum_{i=1}^{10} (X_i - \bar{X})^2 = 0.00544$. これらを信頼区間に代入:

$$\frac{0.00544}{21.7} < \sigma^2 < \frac{0.00544}{2.09} \quad \therefore \quad 0.00025 < \sigma^2 < 0.00260$$

3.6. 母集団比率の区間推定

[理論: 二項母集団]

2.9.1. 二項母集団

母集団をある条件を満たすクラス C と満たさないクラス C' の 2 つに分類する. C, C' が母集団で占める母比率を p, q とする.

- 標本度数 $N =$ 母集団から抽出した大きさ n の標本の中でクラス C に入る個体数
- 標本比率 $P = N/n \leftarrow$ 母比率 p に対する統計量

定理 2.9.1. 母比率 p の母集団から大きさ n の標本を抽出するとき

(1) 標本度数 N は二項分布 $B(n, p)$ に従う:

$$P(N = k) = {}_n C_k p^k (1-p)^{n-k} \quad (k = 0, 1, \dots, n)$$

(2) n が十分大きく, $np \geq 5$, $nq \geq 5$ ならば, 近似的に

$$Z = \frac{N - np}{\sqrt{npq}} = \frac{P - p}{\sqrt{(pq)/n}} \sim N(0, 1) \quad (\text{ラプラスの定理})$$

(証明) (1) 二項分布の定義より明らか. (2) 定理 1.9.2 を見よ. □

例 1. (p.74) 血液型が A 型の割合は日本人では約 38% である. K 小学校新入生 78 名のうち, A 型の生徒が 39 人以上である確率はいくらか.

(解) 新入生の中から無作為に選んだ 78 名のうちの A 型の生徒の人数を N とおくと, $N \sim B(78, 0.38)$. $n = 78$, $p = 0.38$. よって, $np = 78 \times 0.38 \div 29.6 \geq 5$, $nq = 78 \times 0.62 \div 48.4 \geq 5$. ゆえに, 二項分布の正規近似が使えて, $N \sim N(np, npq)$. $\therefore Z = (N - np) / \sqrt{npq} \sim N(0, 1)$. $npq = 78 \times 0.38 \times 0.62 \div 18.4$ なので

$$\begin{aligned} P(N \geq 39)_{B(78, 0.38)} &= P(N \geq 39 - 0.5)_{N(29.6, 18.4)} \\ &= P\left(\frac{N - 29.6}{\sqrt{18.4}} \geq \frac{38.5 - 29.6}{\sqrt{18.4}}\right) = P(Z \geq 2.07) \\ &= 0.5 - \Phi(2.07) = 0.5 - 0.4808 = 0.0192 \end{aligned}$$

□

3.6.1. 大きい標本の場合

[理論] 母集団をある条件を満たすクラス C と満たさないクラス C' に分類して, クラス C が母集団で占める比率 (母比率) を p とする. この母集団から大きさ n の標本を抽出し,

- N : 標本度数, i.e., 標本の中でクラス C に入る個体数
- $P = N/n$: 標本比率

とすると, n が十分大きく, $np \geq 5$ かつ $nq \geq 5$ であれば, 近似的に

$$Z = \frac{\sqrt{n}(P - p)}{\sqrt{p(1-p)}} \sim N(0, 1) \quad \leftarrow \text{定理 2.9.1}$$

正規分布表 II より $z(\alpha/2)$ の値を求めると

$$P(|Z| < z(\alpha/2)) = 1 - \alpha$$

が成り立つ. よって

$$P\left(\left|\frac{\sqrt{n}(P - p)}{\sqrt{p(1-p)}}\right| < z(\alpha/2)\right) = 1 - \alpha$$

ゆえに、母比率 p の見かけ上の $100(1 - \alpha)\%$ 信頼区間として次を得る.

$$P - \sqrt{\frac{p(1-p)}{n}} \cdot z(\alpha/2) < p < P + \sqrt{\frac{p(1-p)}{n}} \cdot z(\alpha/2)$$

ところが、上式で p は未知なので、 p を標本比率 P で近似して次の公式を得る.

[公式: 母比率の区間推定 (大標本)] 母集団の中でクラス C に入る個体数の比率を p とする. この母集団から大きさ n の標本を抽出したときの標本比率を P とすると、 n が十分大きく、 $np \geq 5$ かつ $nq \geq 5$ ならば、母比率 p の $100(1 - \alpha)\%$ 信頼区間は次式で与えられる:

$$P - \sqrt{\frac{P(1-P)}{n}} \cdot z(\alpha/2) < p < P + \sqrt{\frac{P(1-P)}{n}} \cdot z(\alpha/2)$$

例 1. (p.101) M 大学生 100 人を選び、ある日の中日・巨人戦のテレビ中継について調査したところ、37 人が視聴していたことがわかった. M 大学生のこの野球中継視聴率 p の 95%信頼区間を求めよ.

(解) $n = 100$, $P = 37/100 = 0.37$. $np \doteq nP = 100 \times 0.37 = 37 \geq 5$, $nq = 100 - 37 = 63 \geq 5$. よって上の公式が使える. $\alpha = 0.05$ なので正規分布表より $z(\alpha/2) = z(0.025) = 1.96$. これらを信頼区間に代入:

$$0.37 - \sqrt{\frac{0.37 \times (1 - 0.37)}{100}} \times 1.96 < p < 0.37 + \sqrt{\frac{0.37 \times (1 - 0.37)}{100}} \times 1.96$$

$$\therefore 0.37 - 0.095 < p < 0.37 + 0.095 \quad \therefore 0.275 < p < 0.465$$

□

4章 統計的仮説検定

4.3. 母平均の検定—母分散が既知の正規母集団／母集団または母分散が未知で大標本の場合

[問題の設定] 正規母集団 $N(\mu, \sigma^2)$ (σ^2 は既知) の未知の母平均 μ を具体的な値 μ_0 と比較して大小関係などを判定したい。

[検定の手順]

(1) 帰無仮説 $H_0 : \mu = \mu_0$

(2) 対立仮説 H_1 として次の3つのうちのどれか1つを選ぶ。

(a) $\mu > \mu_0$

(b) $\mu < \mu_0$

(c) $\mu \neq \mu_0$

対立仮説として, (a), (b), (c) を選んだとき, それぞれ, **右側検定**, **左側検定**, **両側検定**という。右側検定, 左側検定のことをまとめて**片側検定**という。

(3) **検定統計量**: 大きさ n の標本の標本平均 \bar{X} を検定統計量に設定。仮説 H_0 のもとでは

$$\bar{X} \sim N(\mu_0, \sigma^2/n).$$

$$\therefore Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1) \quad (*)$$

(4) **棄却域の設定**: 有意水準 (=危険率) $100\alpha\%$ の**棄却域**を対立仮説 H_1 に応じて定める。すなわち, 対立仮説が

(a) $\mu > \mu_0$ ならば $P(Z > z(\alpha)) = \alpha$ となる領域 $Z > z(\alpha)$ (**右側検定**)

(b) $\mu < \mu_0$ ならば $P(Z < -z(\alpha)) = \alpha$ となる領域 $Z < -z(\alpha)$ (**左側検定**)

(c) $\mu \neq \mu_0$ ならば $P(|Z| > z(\alpha/2)) = \alpha$ となる領域 $|Z| > z(\alpha/2)$ (**両側検定**)

さて, 実際に標本を抽出して, Z の値を計算したとき, Z の値が棄却域 $Z > z(\alpha)$, $Z < -z(\alpha)$, $|Z| > z(\alpha/2)$ に入ったとする。このとき, 次の2つの考え方がある:

- ① 確率 α でしか起きない珍しいことがたまたま起きた.
- ② 帰無仮説 $H_0: \mu = \mu_0$ がそもそも間違っている.

統計的仮説検定では、常に②の考え方をとる.

- (5) **実現値の計算:** 実際の標本から \bar{X} の実現値を計算して (*) に代入し, Z の値を求める.
- (6) **結論:** 上で求めた Z の値が (4) で設定した棄却域に入れば, “仮説 H_0 は有意水準 $100\alpha\%$ で棄却される”. したがって, “仮説 H_1 が採択される”. もし, Z の値が棄却域に入らなければ, “仮説 H_0 は有意水準 $100\alpha\%$ で棄却されない”. したがって, “仮説 H_0 が採択される”.

(注意) 上で述べた母平均の検定において, 大標本 ($n \geq 50$) の場合には

- (i) 母分布の型が正規分布でなくても (\leftarrow 中心極限定理),
- (ii) 母分散 σ^2 が未知でも, それを標本分散 S^2 で近似する (\leftarrow 標本分散の一致性)

ことにより, 上の検定形式は近似的に有効.

例 1 (p.121). 直径 5mm のボールベアリングを量産するために, ある製造条件のもとで試作した 100 個の標本について実測し, 平均値 4.998mm, 標準偏差 0.009mm を得た. この条件では規格値より直径が小さくなるを考えるべきか. 有意水準 5% で検定せよ.

(解) 大標本 ($n = 100$) なので, 母分布の型や母分散が未知でも, 未知の母分散 σ^2 を標本分散 S^2 で置き換えれば, 上の検定方式が使える.

[母平均の検定 (大標本)]

- 母平均 μ と比較する値 $\mu_0 = 5$
- 標本の大きさ $n = 100$
- 母標準偏差 (または標本標準偏差) $\sigma = S = 0.009$
- 標本平均の実現値 $\bar{X} = 4.998$

- (1) 帰無仮説 $H_0: \mu = 5$
- (2) 対立仮説 $H_1: \mu < 5$ (左側検定)
- (3) 検定統計量: 標本平均 \bar{X} . 仮説 H_0 のもとでは

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1)$$

- (4) 棄却域の設定: 有意水準 5% $\therefore 100\alpha = 5 \therefore \alpha = 0.05 \therefore z(0.05) = 1.6449$.
よって, 棄却域: $Z < -1.6449$
- (5) 実現値の計算: 母標準偏差 σ は未知なので, それを標本標準偏差 S で代用する.

$$Z = \frac{4.998 - 5}{0.009/\sqrt{100}} \doteq -2.22$$

- (6) 結論: Z の実現値は棄却域に入る. ゆえに仮説 H_0 は有意水準 5% で棄却される. すなわち, 「この条件で製造された製品の直径は規格値より小さくなると考えられる (有意水準 5%)」. □

4.7 正規母集団の分散に関する検定

4.7.1 母分散の検定

[問題の設定] 正規母集団 $N(\mu, \sigma^2)$ の未知の母分散 σ^2 を具体的な値 σ_0^2 と比較して、大小関係などを判定する。

[検定の手順]

- (1) 帰無仮説 $H_0 : \sigma^2 = \sigma_0^2$
- (2) 対立仮説 H_1 として次の3つのうちのどれか1つを選ぶ。

$$(a) \sigma^2 > \sigma_0^2 \quad (b) \sigma^2 < \sigma_0^2 \quad (c) \sigma^2 \neq \sigma_0^2$$

- (3) 検定統計量: 母集団 $N(\mu, \sigma^2)$ から大きさ n の標本を抽出し、その標本分散 S^2 を検定統計量に設定。仮説 H_0 のもとでは

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \text{自由度 } n-1 \text{ の } \chi^2 \text{ 分布} \quad (*)$$

- (4) 棄却域の設定: 有意水準 $100\alpha\%$ の棄却域を、対立仮説 H_1 に応じて定める。すなわち、対立仮説が

- (a) $\sigma^2 > \sigma_0^2$ ならば $P(\chi^2 > \chi_{n-1}^2(\alpha)) = \alpha$ となる領域 $\chi^2 > \chi_{n-1}^2(\alpha)$ (右側検定)
- (b) $\sigma^2 < \sigma_0^2$ ならば $P(\chi^2 < \chi_{n-1}^2(1-\alpha)) = \alpha$ となる領域 $\chi^2 < \chi_{n-1}^2(1-\alpha)$ (左側検定)
- (c) $\sigma^2 \neq \sigma_0^2$ ならば $P(\chi^2 > \chi_{n-1}^2(\alpha/2) \text{ または } \chi^2 < \chi_{n-1}^2(1-\alpha/2)) = \alpha$ となる領域 $\chi^2 < \chi_{n-1}^2(1-\alpha/2) \text{ または } \chi^2 > \chi_{n-1}^2(\alpha/2)$ (両側検定)

- (5) 実現値の計算: 実際の標本から標本分散 S^2 の実現値を計算して (*) に代入し、 χ^2 の値を求める。
- (6) 結論: 上で求めた χ^2 の値が (4) で設定した棄却域に入るか否かを確認し、仮説 H_0 が有意水準 $100\alpha\%$ で棄却されるか否かを結論する。

例 1 (p.134). 正規母集団から大きさ 20 の標本を取ったところ, 標本分散 $S^2 = 13.6$ であった. 有意水準 5% で仮説 $H_0 : \sigma^2 = 14.2$ を対立仮説 $H_1 : \sigma^2 < 14.2$ に対して検定せよ.

(解) 正規母集団の母分散の検定方式が使える.

[母分散の検定]

- 母分散 σ^2 と比較する値 $\sigma_0^2 = 14.2$
- 標本の大きさ $n = 20$
- 標本分散の実現値 $S^2 = 13.6$

- (1) 帰無仮説 $H_0 : \sigma^2 = 14.2$
- (2) 対立仮説 $H_1 : \sigma^2 < 14.2$ (左側検定)
- (3) 検定統計量: 標本分散 S^2 . 仮説 H_0 のもとでは

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \text{自由度 } n-1 \text{ の } \chi^2 \text{ 分布}$$

- (4) 棄却域の設定: 有意水準 5% $\therefore 100\alpha = 5 \therefore \alpha = 0.05 \therefore \chi_{19}^2(0.95) = 10.12$. よって, 棄却域: $\chi^2 < 10.12$
- (5) 実現値の計算:

$$\chi^2 = \frac{(20-1) \times 13.6}{14.2} = 18.20$$

- (6) 結論: χ^2 の実現値は棄却域に入らない. ゆえに仮説 H_0 は有意水準 5% で棄却されない. すなわち, 「正規母集団の母分散は 14.2 であると考えられる (有意水準 5%)」. \square

4.8 母比率に関する検定—大標本の場合

4.8.1 母比率の検定

[問題の設定] 大標本を取り得る場合に, 母集団の未知の母比率 p を具体的な値 p_0 と比較して, 大小関係などを判定する.

[検定の手順]

- (1) 帰無仮説 $H_0 : p = p_0$
- (2) 対立仮説 H_1 として次の 3 つのうちのどれか 1 つを選ぶ.

$$(a) p > p_0 \quad (b) p < p_0 \quad (c) p \neq p_0$$

- (3) 検定統計量: 大きさ n の標本における標本比率 P を検定統計量に設定. n が十分大きく $np_0 \geq 5$ かつ $nq_0 \geq 5$ ならば, 仮説 H_0 のもとで P は近似的に $N(p_0, p_0(1-p_0)/n)$ に従う.

$$\therefore Z = \frac{P - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0, 1) \quad (*)$$

- (4) 棄却域の設定: 有意水準 $100\alpha\%$ の棄却域を, 対立仮説 H_1 に応じて定める. すなわち, 対立仮説が

- (a) $p > p_0$ ならば $P(Z > z(\alpha)) = \alpha$ となる領域 $Z > z(\alpha)$ (右側検定)
 (b) $p < p_0$ ならば $P(Z < -z(\alpha)) = \alpha$ となる領域 $Z < -z(\alpha)$ (左側検定)
 (c) $p \neq p_0$ ならば $P(|Z| > z(\alpha/2)) = \alpha$ となる領域 $|Z| > z(\alpha/2)$ (両側検定)

- (5) 実現値の計算: 実際の標本から標本比率 P の実現値を計算して (*) に代入し, Z の値を求める.
 (6) 結論: 上で求めた Z の値が棄却域に入るか否かを調べて, 仮説 H_0 が有意水準 $100\alpha\%$ で棄却されるか否かを結論する.

例 1 (p.137). サイコロ賭博で有り金を巻き上げられてしまった男が, サイコロにいかさま細工がしてあるのではと疑って, 実際に 120 回投げてみたら, 1 の目が 28 回出た. 彼はこの結果から, 1 の目の出方が異常に大きいと主張できるか. 有意水準 5% で検定せよ.

(解) 母比率 (大標本の場合) の検定方式が使える.

[母比率の検定 (大標本)]

- 母比率 p と比較する値 $p_0 = 1/6$
- 標本の大きさ $n = 120$
- 標本比率の実現値 $P = 28/120 \doteq 0.2333$

- (1) 帰無仮説 $H_0 : p = 1/6$
 (2) 対立仮説 $H_1 : p > 1/6$ (右側検定)
 (3) 検定統計量: 標本比率 P . ここで $np_0 = 120/6 = 20 > 5$ かつ $nq_0 = 120 - 20 = 100 > 5$. よって仮説 H_0 のもとでは近似的に

$$Z = \frac{P - p_0}{\sqrt{p_0(1 - p_0)/n}} \sim N(0, 1)$$

- (4) 棄却域の設定: 有意水準 5% $\therefore 100\alpha = 5 \therefore \alpha = 0.05 \therefore z(0.05) = 1.6449$. よって, 棄却域: $Z > 1.6449$
 (5) 実現値の計算:

$$Z = \frac{0.2333 - \frac{1}{6}}{\sqrt{\frac{1}{6} \left(1 - \frac{1}{6}\right) / 120}} \doteq 1.9586$$

- (6) 結論: Z の実現値は棄却域に入る. ゆえに仮説 H_0 は有意水準 5% で棄却される. すなわち, 「1 の目の出方は異常に大きいと考えられる (有意水準 5%)」. \square