第 IX 部　統計的検定

4　信頼区間

- 信頼区間
- 標準正規分布と一般的な正規分布
- 1.96 という値

4.1　はじめに

- 観測されるデータの分布には正規分布が多く見られます。
- 観測したデータの分布が正規分布だとみなせるならば正規分布の性質を利用できます。
- 全ての正規分布は標準化すれば標準正規分布になります。
- 標準正規分布の特徴を知れば全ての正規分布の特徴が把握できそうです。

4.1.1　正規曲線

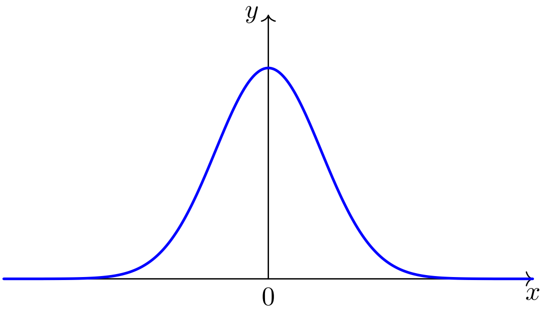
正規曲線とは、

$$y = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \tag{1}$$

をいう。

- 正規曲線は y 軸に対して線対称であり、定義域は $(-\infty,\ +\infty)$ であって、y 軸上で最大値をとる単峰の正値関数である。
- 正規曲線と x 軸に挟まれる範囲の面積は 1 である。

図 1　正規曲線



4.2 標準正規分布

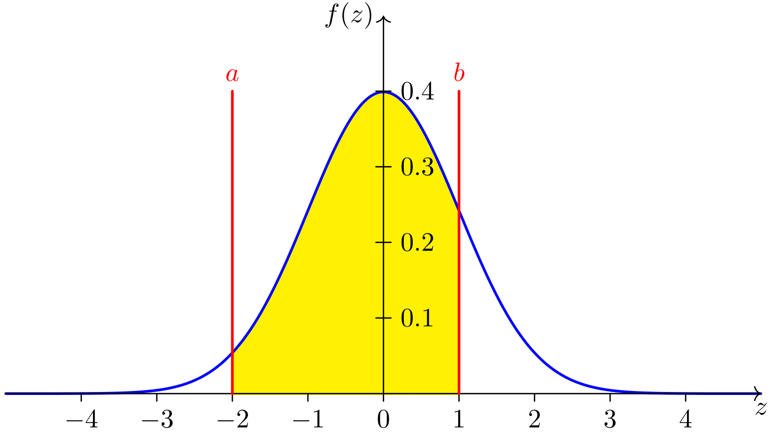
定義

確率変数 Z が 区間 $[a,\ b]$ に属する確率 $\mathrm{Pr}(a \leqq Z \leqq b)$ が、$[a,\ b]$ における正規曲線と z 軸の間の範囲の面積として与えられるとき、確率変数 Z の分布は『標準正規分布』であるといい

$$N(0,1) \tag{2}$$

であらわす。

図 2　標準正規分布



4.2.1　標準正規分布の特徴

標準正規分布において以下の特徴がある。

- 平均は 0 であって、標準偏差は 1 である。
- $\Pr(Z \geqq 1) = 0.158\,655\,254$ である。
- $\Pr(Z \leqq -1) = 0.158\,655\,254$ である。
- $\Pr(-1 \leqq Z \leqq 1) = 0.682\,689\,492$ である。
- $\Pr(Z \geqq 2) = 0.022\,750\,132$ である。
- $\Pr(Z \leqq -2) = 0.022\,750\,132$ である。
- $\Pr(-2 \leqq Z \leqq 2) = 0.954\,499\,736$ である。

4.3　一般的な正規分布

定義

z が標準正規分布 $N(0,\ 1)$ に従うとする。ここで、任意の実数 μ と正数 σ を使い、

$$x = \sigma z + \mu \tag{3}$$

とすると、x は平均 μ, 標準偏差 σ の正規分布に従う。
x の分布を『平均 μ, 標準偏差 σ の正規分布』と呼び、$N(\mu,\ \sigma^2)$ とあらわす。

4.3.1　標準化

(3) を z について解く。

$$x = \sigma z + \mu \tag{3}$$

$$x - \mu = \sigma z \tag{4}$$

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

(5) の変換を標準化という。標準化された値を z 値という。

正規分布は平均 μ と標準偏差 σ を与えると一意に決まる。

4.4　信頼区間

標準正規分布において、z が $-1 \leqq z \leqq 1$ の範囲に収まる確率は

$$\Pr(-1 \leqq z \leqq 1) = 0.682\,689\,492 \tag{6}$$

であって、$-2 \leqq z \leqq 2$ の範囲に収まる確率は

$$\Pr(-2 \leqq z \leqq 2) = 0.954\,499\,736 \tag{7}$$

である。そして、$-1.96 \leqq z \leqq 1.96$ の範囲に収まる確率は

$$\Pr(-1.96 \leqq z \leqq 1.96) = 0.950\,004\,21 \tag{8}$$

である。この $-1.96 \leqq z \leqq 1.96$ の区間を『95 % 信頼区間』という。

4.4.1 信頼区間の意味

- 観測される z が 95 % 信頼区間の範囲に収まる確率は 0.95 である。
- 範囲に含まれない確率は 0.05 である。
- 信頼区間の幅を広げれば、範囲に含まれる確率は大きくなる。
- 標準正規分布では定義域は $(-\infty,\ +\infty)$ なので、$-1.96 \leqq z_i \leqq 1.96$ は相対的にほんのわずかな区間である。
- 信頼区間の範囲を決めるのは分析者である。
- 同じ確率を持つならば、信頼区間は狭い範囲であること好ましい。
- 標準正規分布は左右対称であって、対称軸で最大値を持つ単峰形をしているので、同じ確率を持つならば $z = 0$ を中心にした対称の区間が最も好ましい。

4.4.2 一般的な正規分布の 95% 信頼区間

x が平均 μ, 標準偏差 σ の正規分布 $N(\mu,\ \sigma)$ に従う場合

$$x = \sigma z + \mu \tag{3}$$

なので

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

である。

従って、x の 95% 信頼区間は

$$\begin{aligned}
-1.96 &\leqq \quad z \quad \leqq +1.96 && (9)\\
-1.96 &\leqq \frac{x-\mu}{\sigma} \leqq +1.96 && (10)\\
-1.96\,\sigma &\leqq x-\mu \leqq +1.96\,\sigma && (11)\\
\mu - 1.96\,\sigma &\leqq \quad x \quad \leqq \mu + 1.96\,\sigma && (12)
\end{aligned}$$

である。

例題 IX−4 − 1

裏が出る確率が $p = 0.5$ であるコインがある。このコインを 100 回投げ裏の出る回数を x とする。x の 95% 信頼区間を求めなさい。

解法 IX−4 − 1

x は $B(100,\ 0.5)$ に従う。x は確率変数なので X とあらわす。二項分布に従う確率変数 X の期待値, 分散, 標準偏差は

$$E(X) = np = 100 \times 0.5 = 50 \tag{13}$$

$$V(X) = np(1 - p) = 100 \times 0.5 \times 0.5 = 25 \tag{14}$$

$$D(X) = \sqrt{V(X)} = \sqrt{25} = 5 \tag{15}$$

である。

n が十分に大きいので、x は $N(50,\ 5^2)$ に従う。従って 95% 信頼区間は、

$$50 - 1.96 \times 5 \ \leqq\ x\ \leqq\ 50 + 1.96 \times 5 \tag{16}$$

$$40.2 \ \leqq\ x\ \leqq\ 59.8 \tag{17}$$

である。

従って、このコインを 100 回投げると、95% の確率で、「40 回以上 60 回以下が裏になる」といえる。

例題 IX−4 − 2

ある生き物の身長を y とする。y は平均 $\mu = 171.4$, 標準偏差 $\sigma = 5.8$ の正規分布に従うことが分かっている。y の 95% 信頼区間を求めなさい。

解法

y は $N(171.4,\ 5.8^2)$ に従うので、95% 信頼区間は、

$$171.4 - 1.96 \times 5.8 \ \leqq\ y \ \leqq\ 171.4 + 1.96 \times 5.8 \tag{18}$$

$$160.032 \ \leqq\ y \ \leqq\ 182.768 \tag{19}$$

である。

問題 IX**−4 − 1**

ある生き物の身長を z とする。z は平均 $\mu = 157.5$、標準偏差 $\sigma = 5.4$ の正規分布に従うことが分かっている。z の 95% 信頼区間を求めなさい。

解例 IX$-\mathbf{4}-\mathbf{1}$

z は $N(157.5,\ 5.4^2)$ に従うので、95% 信頼区間は、

$$157.5 - 1.96 \times 5.4 \leqq z \leqq 157.5 + 1.96 \times 5.4$$
$$146.916 \leqq z \leqq 168.084$$

である。

4.5 まとめ

- 標準正規分布 $N(0,\ 1)$ において、$-1.96 \leqq z \leqq +1.96$ の範囲の確率は 95% である。
- 標準正規分布において、

$$-1.96 \leqq z \leqq +1.96$$

  を『95% 信頼区間』という
- $N(\mu,\ \sigma^2)$ を『平均 μ, 標準偏差 σ の正規分布』という。
- $N(\mu,\ \sigma^2)$ の『95% 信頼区間』は

$$\mu - 1.96\ \sigma \leqq x \leqq \mu + 1.96\ \sigma$$

  である。

4.5.1 参考文献

(1) 「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

(2) 「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