

連続型確率分布

数学と医療・第 5 回

数学・橋本貴宏

2024 年 10 月 2 日 1 限

本講義資料の無断使用・無断転載・SNS 等への投稿を固く禁じます

本日の授業では

- 前回のおさらい
 - (1) 2変量の確率分布
 - (2) 離散型確率分布
- 正規分布
 - (1) 正規分布の性質
 - (2) 正規分布の確率計算
- 二項分布の正規近似

前回のおさらい (2 変量の確率分布)

2 つの離散型確率変数 X, Y に対して

$P(X = x_i, Y = y_j) \cdots$ 同時分布.

$P(X = x_i), P(Y = y_j) \cdots$ 周辺分布.

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j)$$

の関係をみたしているならば, X と Y は独立であるという.

X, Y が独立のとき,

(i) X と Y の共分散 $\text{Cov}(X, Y) = 0$.

(ii) $V(aX + bY) = a^2V(X) + b^2V(Y)$.

前回のおさらい (離散型確率分布)

[1] 二項分布

確率変数 X が $0, 1, 2, \dots, n$ の値をとるとき,

$$P(X = x) = {}_n C_x p^x (1 - p)^{n-x}$$

で与えられる確率分布を二項分布 といい, $B(n, p)$ で表す.

期待値 $E(X) = np$, 分散 $V(X) = np(1 - p)$.

[2] ポアソン分布

確率変数 X が $0, 1, 2, \dots$ の値をとるとき, $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (\lambda > 0)$ で与えられる確率分布をポアソン分布といい, $Po(\lambda)$ で表す. 期待値・分散は, $E(X) = V(X) = \lambda$ となる.

前回の演習問題

下の分布表で与えられる同時確率分布について、次の問いに答えよ。

$X \backslash Y$	1	2	4	X の周辺分布
2	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{2}{3}$
3	$\frac{1}{6}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{3}$
Y の周辺分布	$\frac{5}{12}$	$\frac{1}{3}$	$\frac{1}{4}$	1

(1) 期待値 $E(XY)$ を求めよ。

(2) 共分散 $\text{Cov}(X, Y)$ を求めよ。

(解) (1) $x_i y_j$ の表は

$X \backslash Y$	1	2	4
2	2	4	8
3	3	6	12

であるから、

問題解答

XY の期待値は

$$\begin{aligned} E(XY) &= 2 \cdot \frac{1}{4} + 4 \cdot \frac{1}{4} + 8 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 6 \cdot \frac{1}{12} + 12 \cdot \frac{1}{12} \\ &= \frac{6 + 12 + 16 + 6 + 6 + 12}{12} = \frac{29}{6}. \end{aligned}$$

(2) 周辺分布の期待値はそれぞれ $E(X) = 2 \cdot \frac{2}{3} + 3 \cdot \frac{1}{3} = \frac{7}{3},$

$$E(Y) = 1 \cdot \frac{5}{12} + 2 \cdot \frac{4}{12} + 4 \cdot \frac{3}{12} = \frac{5 + 8 + 12}{12} = \frac{25}{12}$$

となる. ゆえに共分散は

$$\mathbf{Cov}(X, Y) = E(XY) - E(X)E(Y) = \frac{29}{6} - \frac{7}{3} \cdot \frac{25}{12} = -\frac{1}{36}.$$

みなさんからのコメント

- 今回の授業のスライドの例題についてなんですが、両方とも X の周辺分布になっていたのがよくわからなかった
誤植でした。スライドは修正しました
- 高校時代には習わなかったことを使えるので難しいけど楽しい
高校の時は、結構な「しばり」があったのです
- 共分散の定義に二重シグマがあり、計算は煩雑になることが多いが、
 $\text{Cov}(X, Y) = E(XY) - E(X)E(Y)$ という公式により計算できる
この計算法の有用性がわかってもらえて嬉しいです
- 独立の時と独立じゃない時というのがあってどちらから証明？ 確認？
するのが難しかった
基本は「独立性を仮定しない」で考える。独立性は特別なのです

箱の中に戻さない場合

問題

箱の中に当りくじ 3 本と外れくじ 9 本が入っている。この中から 1 本ずつ 3 回くじを引いて当りくじが出た回数を確率変数 X とするとき、次の問いに答えよ。ただし、1 回ごとにくじを箱の中に戻さないとする。

- (1) X の確率分布を求め確率分布表をかけ。
- (2) 期待値 $E(X)$ 及び分散 $V(X)$ を求めよ。

解説

(1) 3 回引いて x 回当りが出る確率を求める。1 回目に引いたときに当りである確率は $\frac{3}{12}$ なので、

解説 (つづき)

$$P(X = 3) = \frac{3}{12} \cdot \frac{2}{11} \cdot \frac{1}{10} = \frac{1}{220}, \quad P(X = 0) = \frac{9}{12} \cdot \frac{8}{11} \cdot \frac{7}{10} = \frac{21}{55}.$$

また, 当りが1回出るのは, 「あはは」「はあは」「ははあ」の3通りあるので

$$P(X = 1) = 3 \cdot \frac{3}{12} \cdot \frac{9}{11} \cdot \frac{8}{10} = \frac{27}{55}. \quad \text{同様に当りが2回出るのは}$$

$$P(X = 2) = 3 \cdot \frac{3}{12} \cdot \frac{2}{11} \cdot \frac{9}{10} = \frac{27}{220}. \quad \text{ゆえに確率分布表は次のようになる.}$$

確率分布表:

X	0	1	2	3
P	21/55	27/55	27/220	1/220

$$\bullet \quad P(X = x) = \frac{{}_3C_x \cdot {}_9C_{3-x}}{{}_{12}C_3} \quad \text{である.}$$

解説 (おまけ)

(2) 確率分布表より

$$E(X) = \sum_{x=0}^3 x p_x = \frac{108 + 54 + 3}{220} = \frac{3}{4},$$

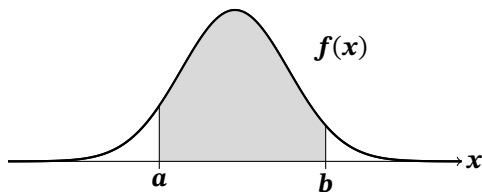
$$\begin{aligned} V(X) &= E(X^2) - E(X)^2 = \sum_{i=0}^3 x^2 p_x - \left(\frac{3}{4}\right)^2 \\ &= \frac{108 + 108 + 9}{220} - \frac{9}{16} = \frac{81}{176}. \end{aligned}$$

※ この分布を超幾何分布といい、くじが N 本, 当たりくじが M 本で引く本数を n とすると,

$$E(X) = n \frac{M}{N}, \quad V(X) = n \frac{M}{N} \left(1 - \frac{M}{N}\right) \cdot \frac{N-n}{N-1}.$$

「確率密度関数」 $f(x)$ を用いてグラフで表す.

$$\begin{cases} f(x) \geq 0, \\ \int_{-\infty}^{\infty} f(x) dx = 1 \end{cases}$$



$a < b$ として $a \leq X \leq b$ である確率:

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

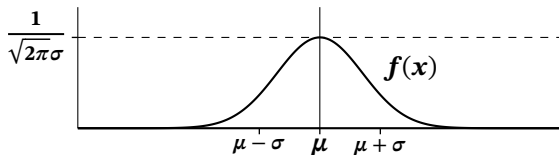
確率はグラフの面積である.

確率密度関数

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

(μ は実数, σ は正数) で表される連続型分布を, 平均 μ , 分散 σ^2 の正規分布 (normal distribution) といい $N(\mu, \sigma^2)$ で表す. (平均と分散は後で示す)

$f'(\mu) = 0$, $f''(\mu \pm \sigma) = 0$ より, 確率密度関数のグラフは次のようになる.



まず, 確率分布の条件を示す. ガウス積分 $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$ を用いることに
より, $x = \mu + \sqrt{2\sigma}y$ と変数変換すれば

$$\int_{-\infty}^{\infty} f(x)dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} e^{-y^2} \sqrt{2\sigma} dy = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} e^{-y^2} dy = 1.$$

平均は

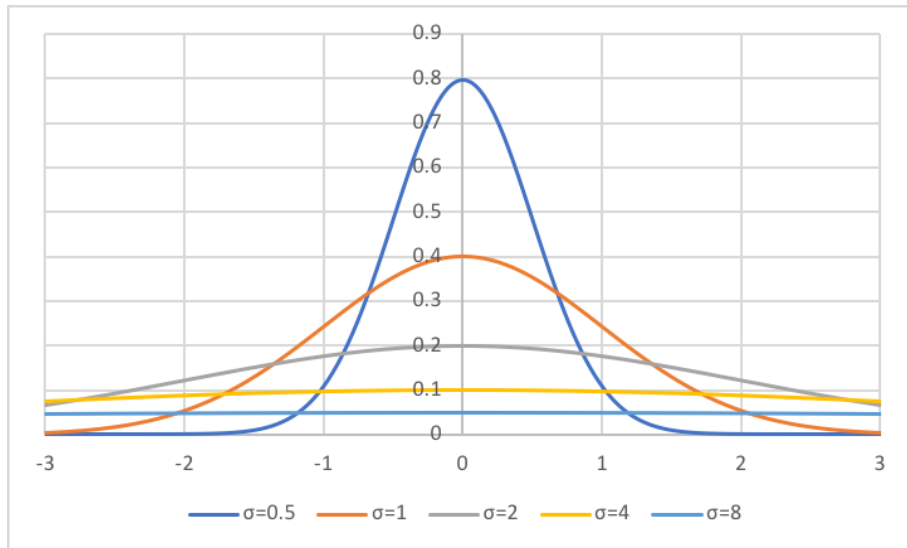
$$\begin{aligned} E(X) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (\mu + \sqrt{2\sigma}y) \sqrt{2\sigma} e^{-y^2} dy \\ &= \frac{1}{\sqrt{\pi}} \left(\mu \int_{-\infty}^{\infty} e^{-y^2} dy + \sqrt{2\sigma} \int_{-\infty}^{\infty} y e^{-y^2} dy \right) \end{aligned}$$

$$= \frac{1}{\sqrt{\pi}} \mu \cdot \sqrt{\pi} + \sqrt{2}\sigma \left[-\frac{1}{2} e^{-y^2} \right]_{-\infty}^{\infty} = \mu + 0 = \mu.$$

分散は

$$\begin{aligned} V(X) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} (x - \mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} 2\sigma^2 y^2 \cdot e^{-y^2} \sqrt{2}\sigma dy = \frac{2\sigma^2}{\sqrt{\pi}} \int_{-\infty}^{\infty} y^2 e^{-y^2} dy \\ &= \frac{2\sigma^2}{\sqrt{\pi}} \left(\left[-\frac{y}{2} e^{-y^2} \right]_{-\infty}^{\infty} + \frac{1}{2} \int_{-\infty}^{\infty} e^{-y^2} dy \right) = 0 + \frac{\sigma^2}{\sqrt{\pi}} \cdot \sqrt{\pi} = \sigma^2. \end{aligned}$$

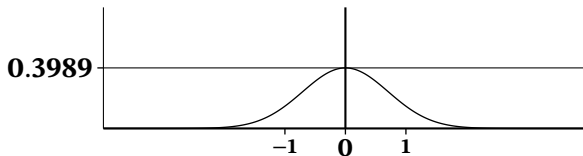
正規分布 ... σ の値を変えると



確率密度関数が

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

である連続型確率分布を、**標準正規分布** (standard normal distribution) という.
平均 $\mu = 0$, 分散 $\sigma^2 = 1$ である. 確率密度関数のグラフは次のようになる.



- 標準正規分布は特別な分布なので, 確率変数 Z で表すことにする.

Z が標準正規分布に従っているとき、確率 $P(a \leq Z \leq b)$ は p.201 正規分布表 I を用いて求める.

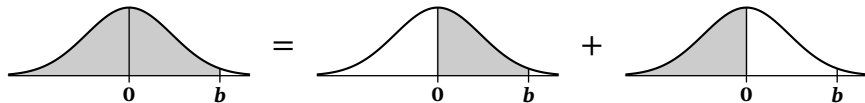
この表には、 $I(z) = P(0 \leq Z \leq z)$ の値が表されている. 例えば、 $I(1.34) = P(0 \leq Z \leq 1.34)$ を求めるには、 $z = 1.34$ のところの値より、 $I(1.34) = 0.4099$ とすれば良い.

z	.00	.01	...	.04
0.0	0.0000	0.0049	...	0.0160
0.1	0.3898	0.0438	...	0.0557
$\vdots$				$\vdots$
1.3	...	...	...	0.4099

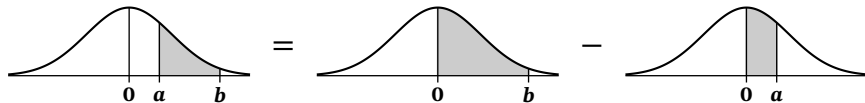
標準正規分布の確率の求め方 (その2)

他の場合も、グラフの y 軸対称性を用いて求めれば良い。

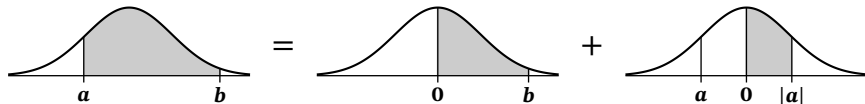
(i) $b > 0$ のとき, $P(Z \leq b) = I(b) + 0.5$



(ii) $0 \leq a < b$ のとき $P(a < Z \leq b) = I(b) - I(a)$

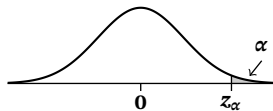


(iii) $a < 0 < b$ のとき $P(a < Z \leq b) = I(b) + I(|a|)$



上側 $100\alpha\%$ 点

確率 $\alpha > 0$ が与えられたとき, $P(Z \geq z) = \alpha$ をみたす z を $N(0,1)$ の上側 $100\alpha\%$ 点といい, z_α で表す. (p.202 正規分布表 II)



$$z_{0.05} = 1.645,$$

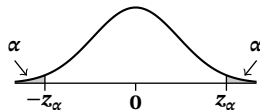
$$z_{0.025} = 1.960,$$

$$z_{0.005} = 2.576$$

はよく使われる. また, 正規分布の対称性より

上側 $100\alpha\%$ 点が z_α

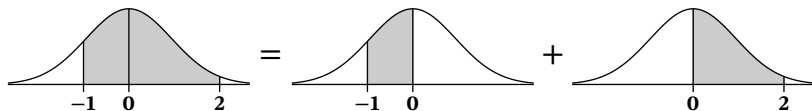
$\iff$ 下側 $100\alpha\%$ 点が $-z_\alpha$.



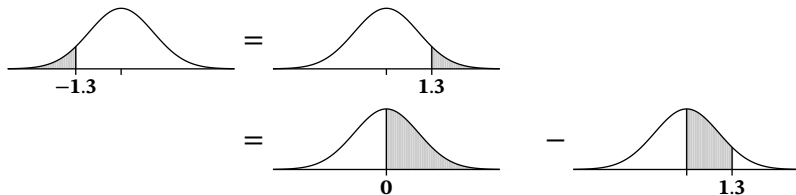
例題 3.4.1 (その1)

(p.68)

$$\begin{aligned}(1) \quad P(-1 < Z < 2.3) &= P(-1 < Z < 0) + P(0 < Z < 2.3) \\ &= P(0 < Z < 1) + P(0 < Z < 2.3) \\ &= I(1) + I(2.3) = 0.3413 + 0.4893 = 0.8306.\end{aligned}$$

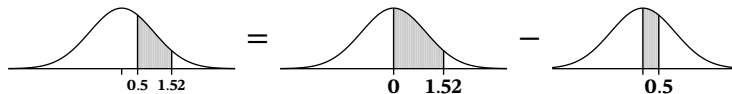


$$(2) \quad P(Z \leq -1.3) = P(1.3 \leq Z) = 0.5 - P(0 \leq Z \leq 1.3) = 0.5 - 0.4032 = 0.0968.$$

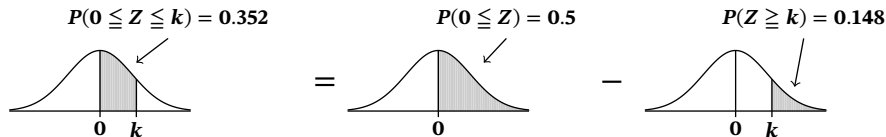


例題 3.4.1 (その2)

$$(3) \quad P(0.5 \leq Z < 1.52) = P(0 \leq Z < 1.52) - P(0 \leq Z \leq 0.5) = 0.4357 - 0.1915 = 0.2442.$$



$$(4) \quad P(0 \leq Z < k) = 0.352 \text{ のとき, } P(Z \geq k) = 0.5 - 0.352 = 0.148. \\ z_{0.148} = 1.045 \text{ であるから, } k = 1.045$$



確率変数 X が $N(\mu, \sigma^2)$ に従うとき, 確率変数 Z を

$$Z = \frac{X - \mu}{\sigma} \quad (\text{平均を引いて標準偏差で割る})$$

と定義すると

$$E(Z) = \frac{1}{\sigma}(E(X) - \mu) = 0, \quad V(Z) = \frac{1}{\sigma^2}V(X) = 1$$

より, Z は $N(0, 1)$ に従っている. この変数変換を標準化という. 次が成り立つ.

$$P(a \leq X \leq b) = P\left(\frac{a - \mu}{\sigma} \leq Z \leq \frac{b - \mu}{\sigma}\right)$$

例題 3.4.2

(p.70)

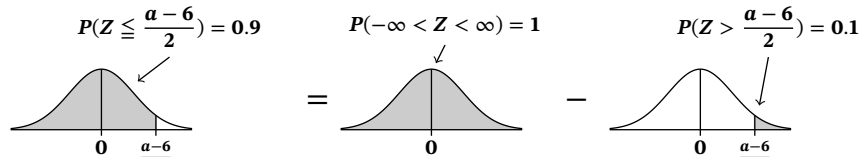
$Z = \frac{X - 6}{2}$ と標準化する.

$$(1) P(X \geq 2) = P(Z \geq \frac{2-6}{2}) = P(Z \geq -2) = P(Z \leq 2) = 0.5 + P(0 \leq Z \leq 2) = 0.9772.$$

$$(2) P(1 < X \leq 5) = P(\frac{1-6}{2} < Z \leq \frac{5-6}{2}) = P(-2.5 < Z \leq -0.5) = P(0.5 \leq Z < 2.5) = 0.4938 - 0.1915 = 0.3023.$$

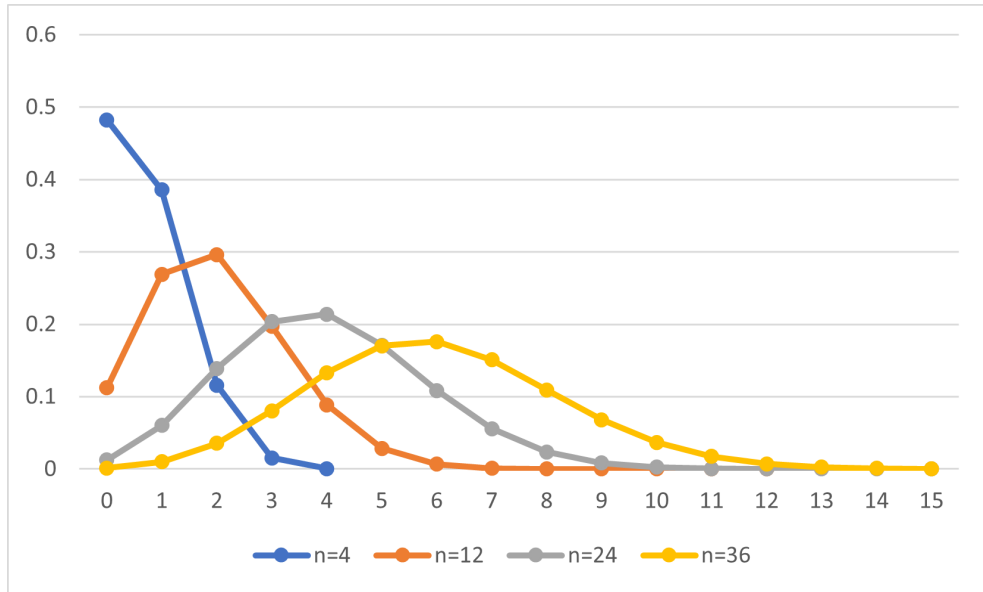
$$(3) P(X \leq a) = P(Z \leq \frac{a-6}{2}) = 0.9 \text{ より, } P(Z > \frac{a-6}{2}) = 0.1. z_{0.1} = 1.282$$

であるから $\frac{a-6}{2} = 1.282$. ゆえに $a = 8.564$.



二項分布 $B(n, 1/6)$ で n を大きくすると

(p.74)

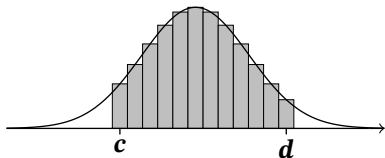


二項分布の正規近似 (その1)

(p.72)

二項分布を思い出そう. p.74 のグラフをしてみる. 二項分布においては n が大きくなるとグラフの山が右に移りながら, 左右対称の正規分布に近づいてくる.

- (1) $B(n, p)$ は, $np \geq 5$ かつ $n(1-p) \geq 5$ のとき, 正規分布 $N(np, np(1-p))$ で近似することができる.
- (2) 二項分布は離散型確率分布であるが, 正規分布は連続型確率分布であるので, **連続補正** (半整数補正) を施す。 $c < d$ のとき $P(c \leq X \leq d)$ を求めるのであれば 0.5 ずつ広げて $P(c - 0.5 \leq X \leq d + 0.5)$ を求めるようにする.



両側に 0.5 ずつ広げる

例題 3.5.1 硬貨を 30 枚投げるとき, 表の出る枚数が 10 枚以上 20 枚以下である確率を正規分布による近似を用いて求めよ。

(解) 表の出る枚数を X とすれば, X は二項分布 $B(30, 0.5)$ に従う。

$np = 15$, $np(1 - p) = 7.5$ より, これを正規分布 $N(\mu, \sigma^2) = N(15, 7.5)$ で近似する。

連続性補正により, $P(9.5 \leq X \leq 20.5)$ を求めることになる。標準化を行うと,

$$\begin{aligned} P(9.5 \leq X \leq 20.5) &= P\left(\frac{9.5 - 15}{\sqrt{7.5}} \leq Z \leq \frac{20.5 - 15}{\sqrt{7.5}}\right) \\ &= P(-2.01 \leq Z \leq 2.01) \\ &= 2 \cdot 0.4778 = 0.9556. \end{aligned}$$

まとめ

- 正規分布

確率密度関数は $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$N(\mu, \sigma^2)$ と表す. 平均は μ , 分散は σ^2 .

- 標準正規分布

$\mu = 0, \sigma = 1$ としたもの. 区間が与えられていて確率を求めるときは p.201「正規分布表Ⅰ」で, 上型確率 α が与えられていて上側 $100\alpha\%$ 点 z_α を求めるときは p.202「正規分布表Ⅱ」

- 正規分布の確率の求め方

$Z = \frac{X - \mu}{\sigma}$ と標準化して標準正規分布に帰着させる

10月2日の演習問題

問題

赤ちゃん（女児）の出生時身長は、平均 **48.5** (cm)、標準偏差 **2.5** (cm) の正規分布に従うことが知られている。このとき次を求めよ。ただし、確率 (%) は整数値で、身長は小数第1位までで答えよ。

- (1) 身長が **47.0** (cm) ~ **53.5** (cm) であるのは全体の何 % か。
- (2) 身長が **51.0** (cm) 以上であるのは全体の何 % か。
- (3) 身長が高い方から **2.5** % の範囲内に入るのは、何 (cm) 以上の赤ちゃんであるか。

次回

「母集団と標本」をやります. 数学 B の「統計的な推測」でやる内容です.

課題提出

AIDLE-K にアクセスし, 「1002 課題提出」から提出する. 提出期限は **10/5 23:59** とする. フィードバック (今日中) も忘れずに.

注意

- ファイル形式に注意し, 拡張子をつける (jpg, pdf 等)
- 解答用紙にも名前を書くこと
- **過去のプリントで出していないものがある場合は, 超過でもかまわないので出すこと**