

大学生のための 統計学入門

—テーマ3. 確率分布—

京都大学大学院医学研究科

臨床統計学/臨床統計家育成コース 田中司朗



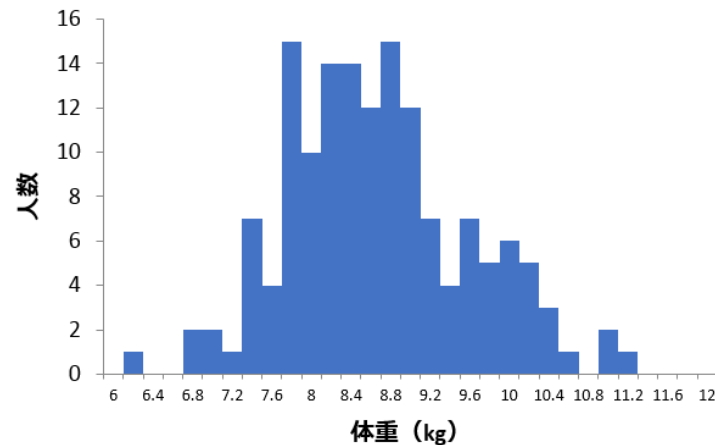
テーマ3. 確率分布

- **確率変数と確率分布**
 - 期待値・分散・標準偏差の計算
 - 二項分布
- サイコロの振り方でバラツキは違うか
 - 正規分布と中心極限定理
- エレベータのブザーは鳴るか
 - 正規分布の利用

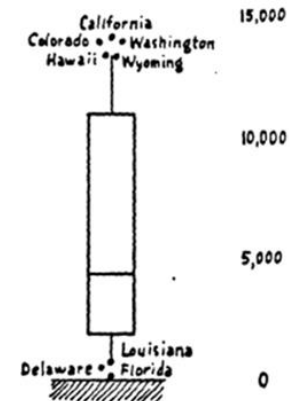


テーマ1, 2の復習

- 位置の指標
 - 平均値, 中央値, 最頻値
- バラツキの指標
 - 分散, 標準偏差, パーセント点, 四分位偏差
- データの分布と代表的なグラフ



ヒストグラム



ボックスプロット

サイコロの振り方で バラツキは違うか



- サイコロを何回も振ったときの, サイコロの出目の分布を考えてみよう
- サイコロの出目は, 平均3.5を中心にばらつく
- サイコロ投げのように確率的試行を繰り返したときにも分布ができる. そのときの分布を確率分布といい, 平均を期待値という

サイコロの振り方で バラツキは違うか



- サイコロを三つ振る
- 出た三つの目の平均を計算したときの期待値は?
 - 一つ振ったときの期待値は3.5
 - 三つ振って, 平均をとっても3.5
- 三つの目の平均の標準偏差は, サイコロを一つ振ったときに比べてどうなる?
 - 三つ振ったほうが大きくなる?
 - 三つ振ったほうが小さくなる?
 - 同じ?

サイコロの振り方で バラツキは違うか



- サイコロを三つ振る
- 出た三つの目の合計を計算したときの期待値は?
 - $3 \times 3.5 = 10.5$
- 20面ダイスの期待値も10.5
- それでは、標準偏差も同じだろうか?
 - 三つ振ったほうが大きくなる?
 - 三つ振ったほうが小さくなる?
 - 同じ?

確率変数と確率分布

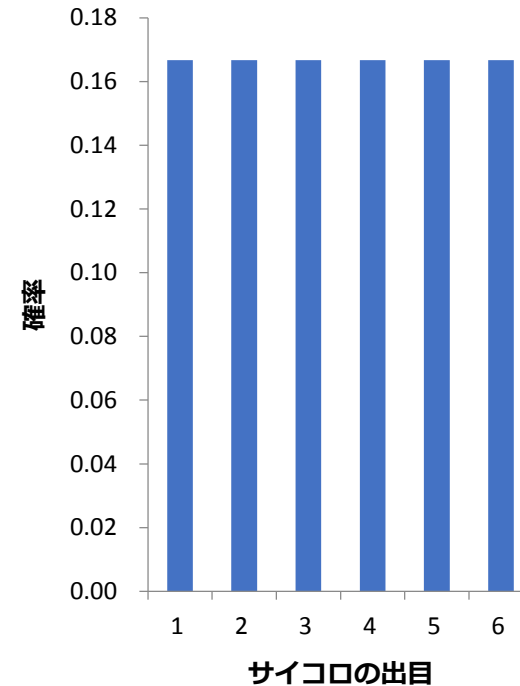
- 確率変数
 - サイコロの出目のように, 実現値に確率が割り当てられているような変数のこと
- 確率分布
 - 実現値に対する確率の散らばり具合のこと
 - 特徴は, 期待値 (平均) , 分散, 標準偏差で表される

サイコロの出目の期待値



• 期待値 = $\frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5 + \frac{1}{6} \times 6$
= 3.5

サイコロの出目	確率
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6



数学記号によるデータの表現

- 合計を表す記号 $\sum_{i=1}^I p_i x_i$
 - 第一項 $p_1 x_1$ から最後の項 $p_I x_I$ までの和
- $$\begin{aligned}\sum_{i=1}^6 p_i x_i &= p_1 x_1 + p_2 x_2 + p_3 x_3 + p_4 x_4 + p_5 x_5 + p_6 x_6 \\ &= \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \frac{1}{6} \times 3 + \frac{1}{6} \times 4 + \frac{1}{6} \times 5 + \frac{1}{6} \times 6\end{aligned}$$

確率変数の期待値の公式

- 確率変数 X が, I 通りの実現値 $x_1, x_2, x_3, \dots, x_I$ をとり, 確率 $p_1, p_2, p_3, \dots, p_I$ が割り当てられている
- そのとき, 確率変数 X の期待値 $E(X)$ は, 以下のよう
に定義される
- $$E(X) = p_1x_1 + p_2x_2 + \dots + p_Ix_I$$
$$= \sum_{i=1}^I p_i x_i$$

確率変数の実現値	確率
x_1	p_1
x_2	p_2
x_3	p_3
x_4	p_4
\vdots	\vdots
x_I	p_I

確率変数の 分散と標準偏差の公式

- 期待値 m を持つ確率変数 X のバラツキは, 分散と標準偏差によって, 以下のように表される
- $$V(X) = p_1(x_1 - m)^2 + p_2(x_2 - m)^2 + \cdots + p_I(x_I - m)^2$$
$$= \sum_{i=1}^I p_i(x_i - m)^2$$
- $$S(X) = \sqrt{\sum_{i=1}^I p_i(x_i - m)^2}$$

確率変数の実現値	確率
x_1	p_1
x_2	p_2
x_3	p_3
x_4	p_4
\vdots	\vdots
x_I	p_I

二項分布

- コインを n 枚投げたときのように, 結果が成功か失敗 (表か裏) のいずれかである独立な試行を n 回行った
- このときの成功数で表される確率分布を二項分布と呼ぶ
- ただし, 各試行における成功確率 p は一定とする

コインを1枚投げたときの 二項分布の期待値

- 公式に $p = 0.5, l = 2$ を代入
 - $E(X) = \sum_{i=1}^l p_i x_i$
- 期待値 $= (1 - p) \times 0 + p \times 1$
 $= (1 - 0.5) \times 0 + 0.5 \times 1$
 $= 0.5$

成功数の実現値	確率
0	0.5
1	0.5

コインを1枚投げたときの 二項分布の分散と標準偏差

- 公式に $p = 0.5, l = 2, m = 0.5$ を代入
 - $V(X) = \sum_{i=1}^l p_i (x_i - m)^2$
- 分散 $= (1 - p) \times (0 - m)^2 + p \times (1 - m)^2$
 $= (1 - 0.5) \times (0 - 0.5)^2 + 0.5 \times (1 - 0.5)^2$
 $= 0.25$
- 標準偏差 $= \sqrt{0.25}$
 $= 0.5$

成功数の実現値	確率
0	0.5
1	0.5