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

サイコロの出目の分散と標準偏差

- 分散 $= \frac{(1-3.5)^2}{6} + \frac{(2-3.5)^2}{6} + \frac{(3-3.5)^2}{6} + \frac{(4-3.5)^2}{6} + \frac{(5-3.5)^2}{6} + \frac{(6-3.5)^2}{6}$

  $= 2.92$

- 標準偏差 $= \sqrt{2.92}$

  $= 1.71$

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

足し算と定数倍のための公式

- 独立な確率変数X と Y の和の期待値と分散には, 以下のような公式が成り立つ
  - $\mathrm{E}(X+Y)=\mathrm{E}(X)+\mathrm{E}(Y)$
  - $\mathrm{V}(X+Y)=\mathrm{V}(X)+\mathrm{V}(Y)$
- 確率変数X の定数cを掛けると, 以下のように期待値はc倍,分散はc^2倍になる
  - $\mathrm{E}(cX)=c\mathrm{E}(X)$
  - $\mathrm{V}(cX)=c^2\mathrm{V}(X)$

指標と単位の関係の復習（定数倍の公式の意味）

- 容疑者のブーツの足跡の最大長（単位はcm）から, 平均・分散・標準偏差を計算した
  - 平均: 24.66
  - 分散: 0.11
  - 標準偏差: 0.33
- 単位がmmだったら平均, 分散, 標準偏差はどうなる?
- 答え
  - 平均: 24.66×10
  - 分散: 0.11×10^2
  - 標準偏差: 0.33×10

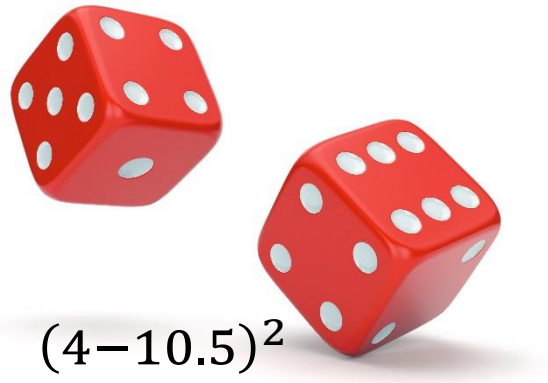
サイコロの振り方で バラツキは違うか

- 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?
  - 一つ振ったときの分散は2.92, 標準偏差は1.71
- 三つ振った合計の分散は
  - $2.92 + 2.92 + 2.92 = 8.76$
- 合計の分散を3で割ったときの分散は
  - $\frac{8.76}{3^2} = 0.972$
- 三つ振った平均の標準偏差は
  - $\sqrt{0.972} = 0.986$
- つまり, 三つ振って平均をとる方がバラツキは小さい

サイコロの振り方で バラツキは違うか

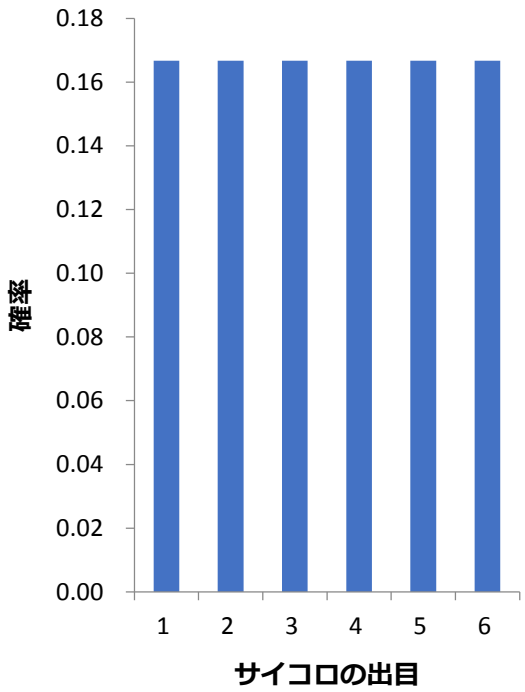
- サイコロを三つ振る
- 出た三つの目の合計を計算したときの期待値は?
  - $3\times3.5=10.5$
- 20面ダイスの期待値も10.5
- それでは, 標準偏差も同じだろうか?
- 三つ振った合計の分散は
  - $2.92 + 2.92 + 2.92 = 8.76$
- 三つ振った合計の標準偏差は
  - $\sqrt{8.76} = 2.96$

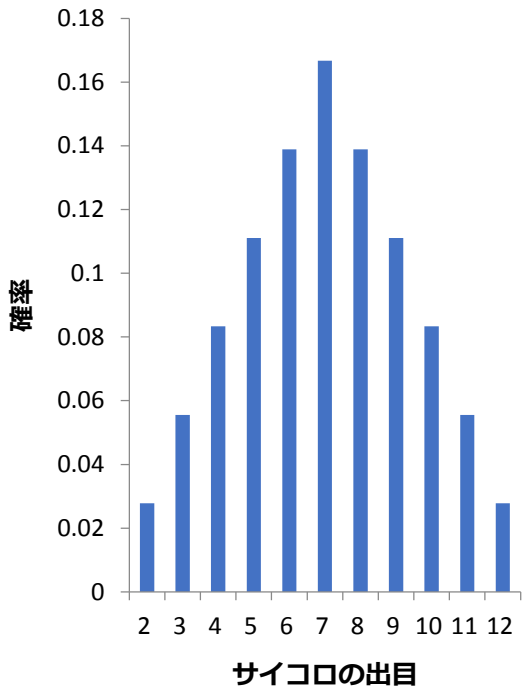
20面ダイスの出目の分散と標準偏差

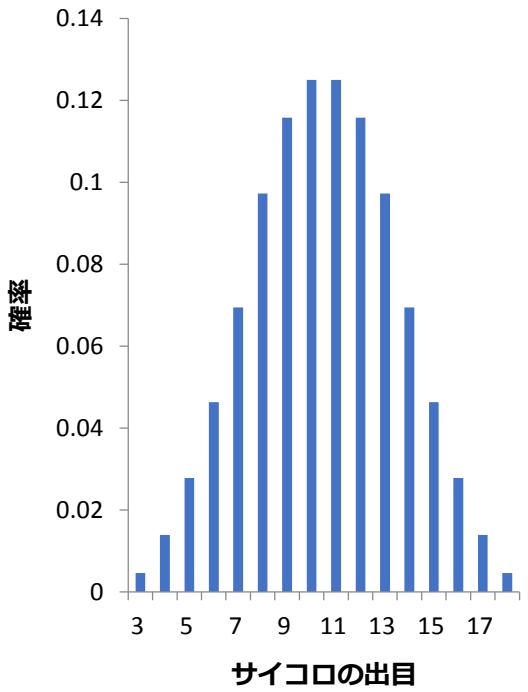


- 分散 $= \frac{(1-10.5)^2}{20} + \frac{(2-10.5)^2}{20} + \frac{(3-10.5)^2}{20} + \frac{(4-10.5)^2}{20}$
  $+ \frac{(5-10.5)^2}{20} + \cdots + \frac{(19-10.5)^2}{20} + \frac{(20-10.5)^2}{20}$
  $= 33.25$
- 標準偏差 $= \sqrt{33.25}$
  $= 5.77$
- つまり, 三つ振って合計する方がバラツキは小さい

サイコロを振る数を増やすと出目は中心に集中する

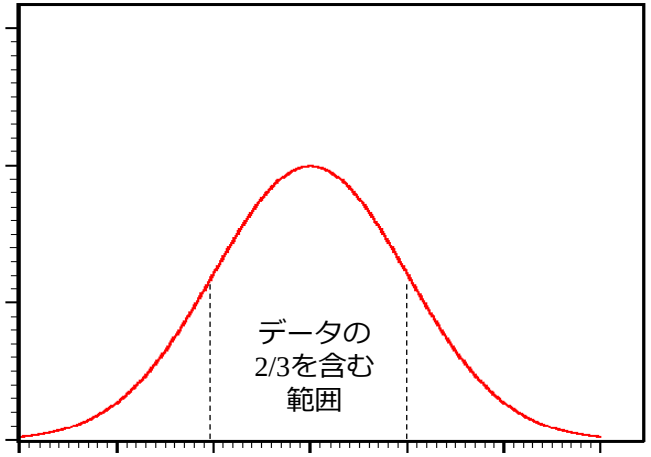


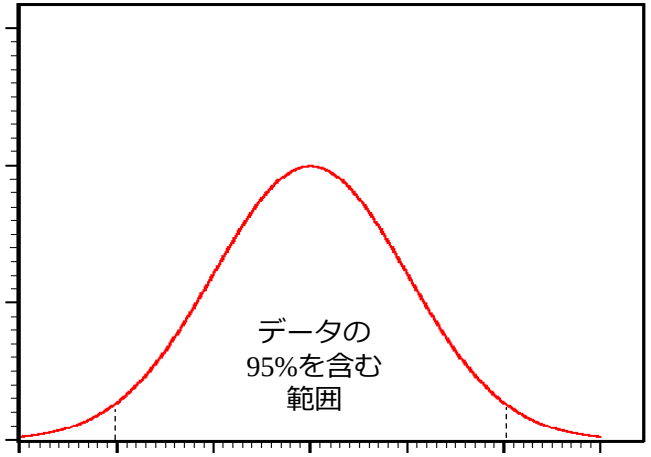


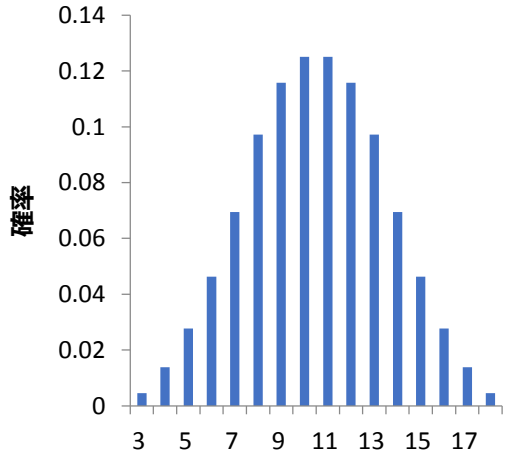


正規分布
もっともよく用いられる確率分布

- 特徴
  - 左右対称
  - 平均±標準偏差に, 全面積の 2/3 が含まれる
  - 平均±2×標準偏差に, 全面積の 95%が含まれる
- サイコロの数を増やすと, 合計の確率分布は正規分布に近づく（中心極限定理）







正規分布
もっともよく用いられる確率分布

ドイツ10マルク紙幣の数学者Gauss （1777～1855年）と彼が発見した正規分布