

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**



エレベータのブザーは鳴るか

- どのエレベータも積載重量が決められていて、それを越えるとブザーが鳴る
- あるエレベータの定員と積載重量

用途	乗用	定員	9名	積載	600kg
FOR PASSENGER		9 PERSONS		LOAD	600kg

- 積載重量に比べて定員が多くはないだろうか?
 - 一人あたりの体重は $600\text{kg}/9\text{人}=67\text{kg}$
 - これを越えるとブザーが鳴ってしまう

エレベータのブザーは鳴るか

- ある集団からランダムに選ばれた9人がエレベータに乗り込んだ
- ブザーは何パーセントの確率で鳴るだろうか?
- 集団の一例
 - ダンスボーカルユニットEのメンバー14人の体重

メンバーA	74kg	メンバーH	74kg
メンバーB	67kg	メンバーI	63kg
メンバーC	63kg	メンバーJ	67kg
メンバーD	69kg	メンバーK	72kg
メンバーE	69kg	メンバーL	63kg
メンバーF	65kg	メンバーM	65kg
メンバーG	58kg	メンバーN	65kg

メンバー14人の体重は どのような分布か

- 平均 = $\frac{\text{一つ一つのデータの合計}}{\text{データの個数}}$
- 分散 = $\frac{(\text{データ}-\text{平均})^2 \text{の合計}}{\text{データの個数}-1}$

計算手順

1. 一つ一つのデータと平均との差をとる
2. 「平均との差の二乗」を求める
3. 「平均との差の二乗」を合計する
4. 合計を「データの個数-1」で割る

- 標準偏差 = $\sqrt{\text{分散}}$

メンバー14人の体重は どのような分布か

- 平均 $= \frac{74+74+67+63+63+67+69+72+69+63+65+65+58+65}{14}$
 $= 66.7$
- 分散 $= \frac{(74-66.7)^2+(74-66.7)^2+\dots+(65-66.7)^2}{14-1}$
 $= 19.3$
- 標準偏差 $= \sqrt{19.3}$
 $= 4.4$

一般成人とメンバー14人の 体重分布の比較

- 14人の平均±標準偏差
 - 66.7 ± 4.4
- 一般成人体重の平均±標準偏差（国民健康・栄養調査）
 - 男性 65.7 ± 11.1
 - 女性 52.9 ± 9.0
- 平均体重は一般成人男性よりわずかに重いものの、体重のバラツキが小さい（極端に重い人や軽い人がいない）ことがわかる
- それではエレベータのブザーが鳴りやすいのはどちらか

メンバー14人のうち9人が乗るとき エレベータのブザーは鳴るか

- メンバー14人のうち, 体重の軽い9人を選ぶと?
 - 合計体重は576kgなのでブザーは鳴らない
- メンバー14人のうち, 9人をランダムに選ぶと?
 - ブザーが鳴る確率

$$= \frac{\text{14人から9人を選び合計体重が600kgを越える場合の数}}{\text{14人から9人を選ぶ場合の数 } ({}_{14}C_9)}$$

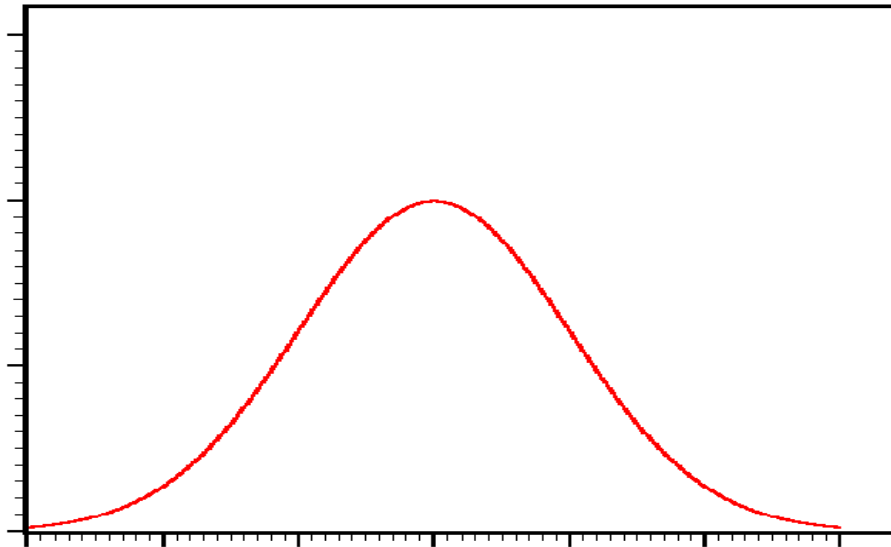
$$= \frac{971}{2002}$$

$$= 0.49$$

- およそ2回に1回はブザーが鳴る

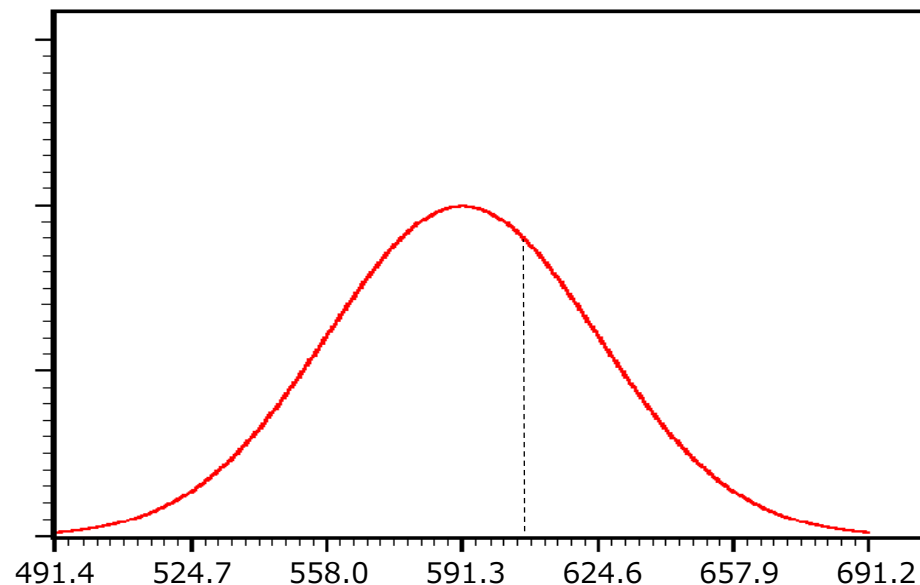
一般成人男性9人が乗るとき エレベータのブザーは鳴るか

- 一般成人の体重分布は不明だが、平均±標準偏差が 65.7 ± 11.1 ということはわかっている
- そこで、正規分布を用いて確率を調べる



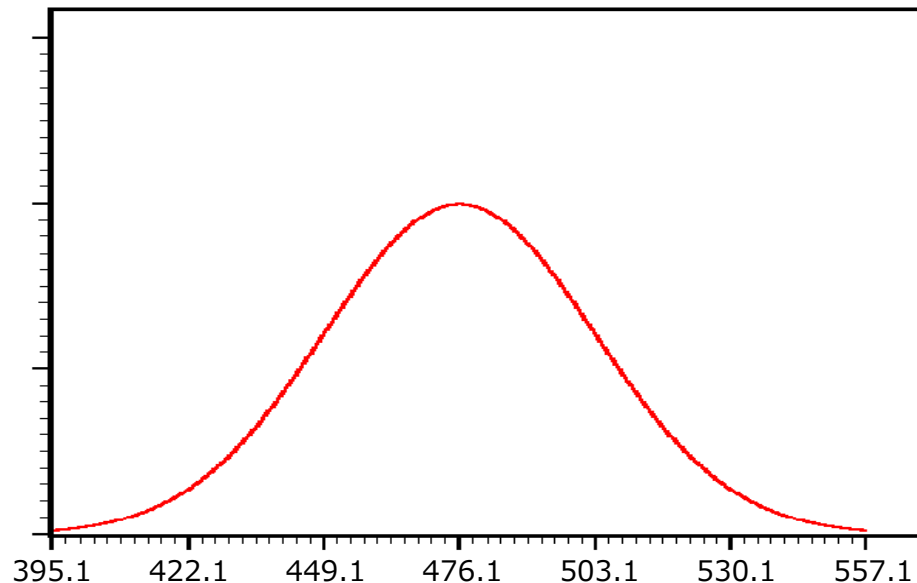
一般成人男性9人が乗るとき エレベータのブザーは鳴るか

- 9人分の体重の平均±標準偏差を求めると
 - 平均 $65.7 \times 9 = 591.3$
 - 標準偏差 $11.1 \times \sqrt{9} = 33.3$
- 平均591.3, 標準偏差33.3の正規分布で600kgより右側の確率は39.7%



一般成人女性9人が乗るとき エレベータのブザーは鳴るか

- 9人分の体重の平均±標準偏差を求めると
 - 平均 $52.9 \times 9 = 476.1$
 - 標準偏差 $9.0 \times \sqrt{9} = 27.0$
- 平均476.1, 標準偏差27.0の正規分布で600kgより右側の確率は0.1%以下



テーマ3のまとめ

- サイコロを三つ振ったり, エレベータに9人乗ったりしたとき, 合計の分布は意外とバラツキが小さい
- このことは, サイコロのように確率分布が分かっているとき, 確率計算すれば確かめることができる
- 実際には, 体重のように分布は分からないことが多い
- それでも, 「データの数が増えると, 合計の確率分布は正規分布に近づく」という中心極限定理を利用すれば確率を計算できる

さらに学びたい人に

- 教科書

- 高校生からの統計入門. 加藤久和著. 筑摩書房2016
- 統計学入門. 東京大学教養学部統計学教室編. 東京大学出版会1991