

ブートストラップ法による
リサンプリング



復元抽出と非復元抽出

- ▶ 箱の中に赤玉**7**つ，白玉**3**つが入っている

- ▶ **復元抽出**

中から**1**つ玉を出して，
出た色を確認して，元に戻す。
また，中から**1**つ玉を出して，
出た色を確認して，元に戻す・・・

- ✓ 同じものが繰り返し選ばれることもある。
- ✓ 玉を出すときの状態（赤玉**7**つ，白玉**3**つ）は変化しない。
⇒ 出現する確率に変化がない。



- ▶ **非復元抽出**

中から**1**つ玉を出して，
出た色を確認して，戻さず。
また，中から**1**つ玉を出して，
出た色を確認して，戻さず・・・

- ✓ 同じものが繰り返し選ばれることはない。
⇒ 最後は空になる。
- ✓ 玉を出すときの状態は，その前の状況で変化します。
⇒ 出現する確率は変化する。



ブートストラップ法 (1)

例 $n = 20$ (人) の成人男性の身長データ

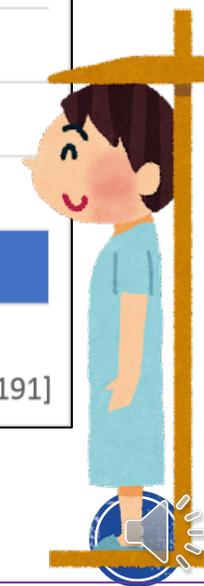
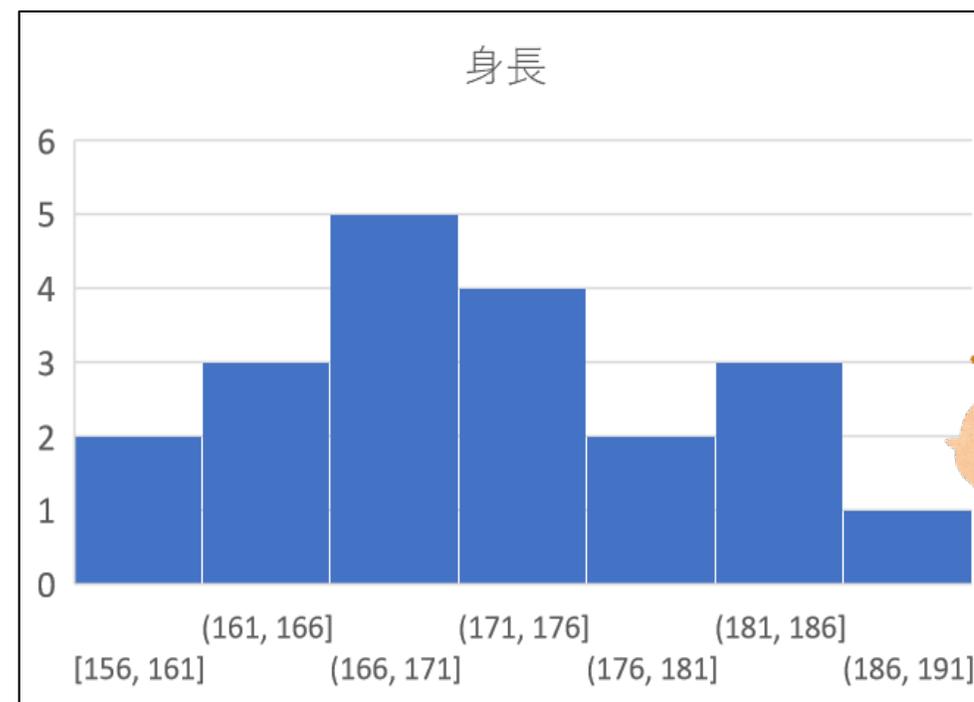
- ✓ 母集団：成人男性の身長全体
- ✓ 母集団から小サンプルのデータが得られた

172.5	161.3	166.3	166.3	165.4
173.4	179.4	175.5	170.9	167.2
176.1	166.5	156.1	184.7	182.8
158.9	186.8	184.6	178.9	165.3

単位：cm

- ✓ 標本平均 171.95
- ✓ 標本分散 8.76
- ✓ 標本標準偏差 2.96

- ✓ ヒストグラムを描く



ブートストラップ法 (2)

▶ ブートストラップ法

(**Bootstrap method**)

- ✓ 母集団：成人男性の身長全体
- ✓ 小サンプル (20人) のデータがえられた
- ✓ **再標本化 (リサンプリング)**



サンプルサイズ**20**のデータを元のデータから抽出する ⇒ **20個の観測値を復元抽出**により選ぶ。

これを繰り返す

1回目： x_1, x_2, \dots, x_{20} ⇒ 重複は可

2回目： x_1, x_2, \dots, x_{20} ⇒ 重複は可

3回目： x_1, x_2, \dots, x_{20} ⇒ 重複は可

- ✓ 毎回の

標本平均 標本分散 標本標準偏差
などを計算してこれらの分布
(ヒストグラム) を作成する。

- ▶ 具体例：**1000回**繰り返したヒストグラム

[シミュレーションブートストラップ法.xlsx](#)

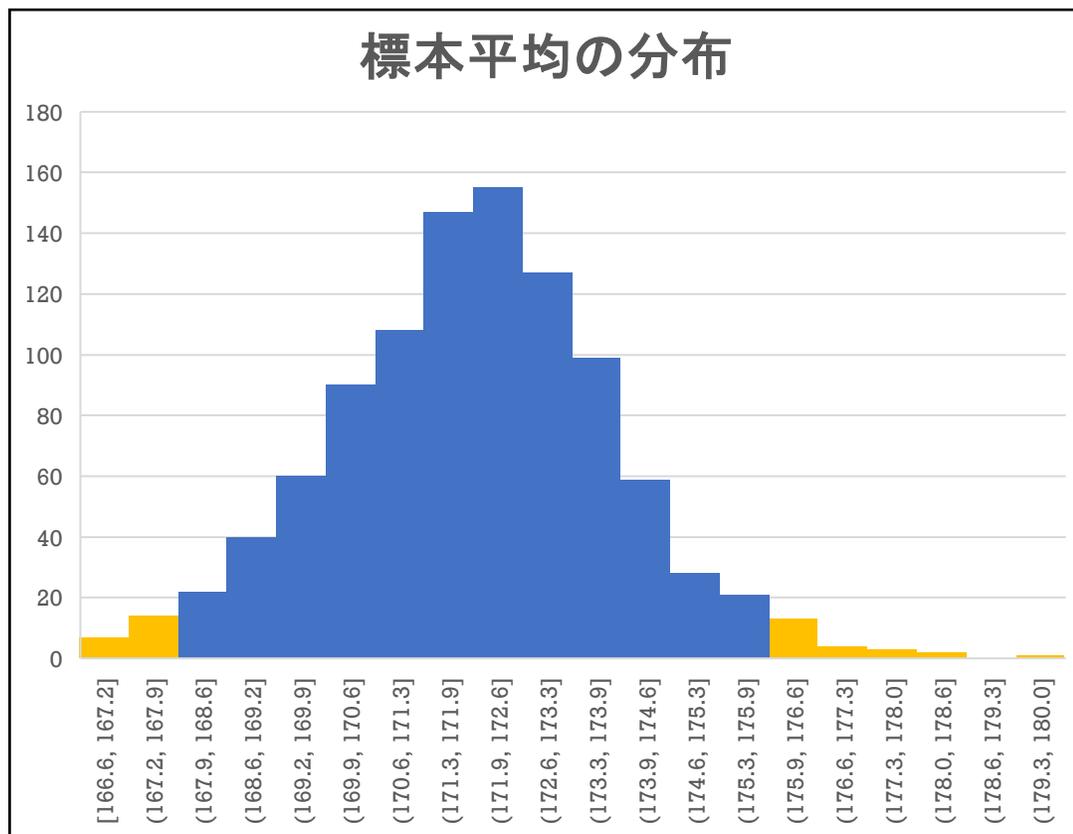


リサンプリングで推定する (1)

▶ 標本平均の分布 (1000回繰返し)

166~180 (全体), **168~176** (裾をなくす)

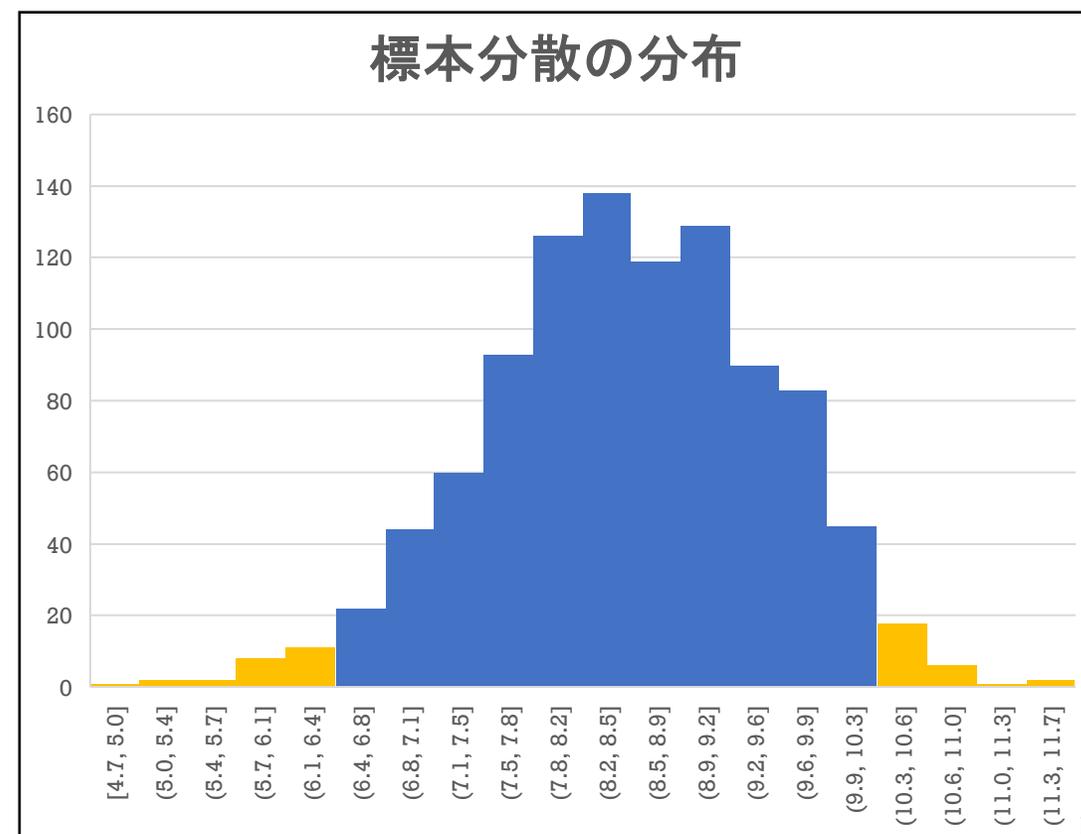
(参考: 標本平均 171.95)



▶ 標本分散の分布 (1000回繰返し)

4.7~11.7 (全体), **6.4~10.3** (裾をなくす)

(参考: 標本分散 8.76)

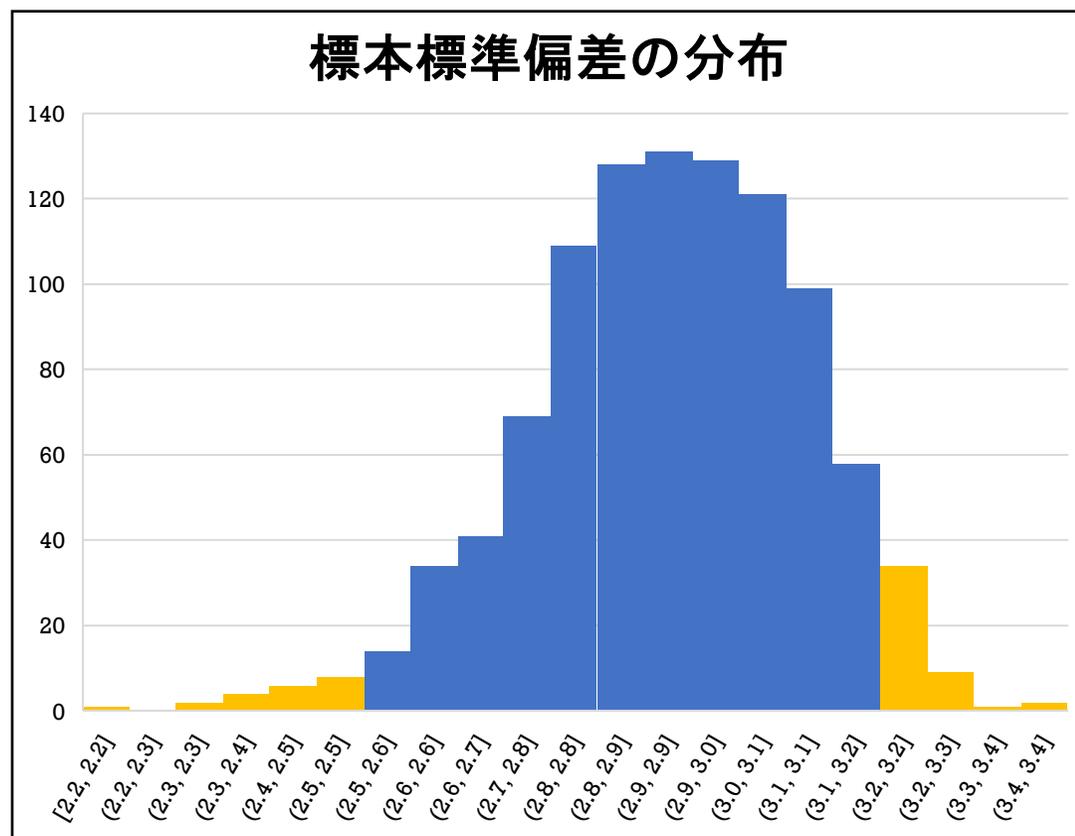


リサンプリングで推定する (2)

▶ 標本標準偏差の分布 (1000回繰返し)

2.2~3.4 (全体), 2.5~3.2 (裾をなくす)

(参考: 標本標準偏差 2.96)



▶ **ブートストラップ信頼区間**

後述の信頼区間の考え方に近い.

ここでは**95%**信頼区間をイメージした.

