統計学

6月29日（火）
第10回

兵庫県立大学 社会情報科学部

山本 岳洋

t.yamamoto@sis.u-hyogo.ac.jp

2021年度前期・火曜3限　神戸商科キャンパス 全学共通科目

本日の内容

- 前回の課題の解説・コメント返信
- 母集団と標本の残り
- 今週の課題（締切:7月2日）について

- 今日はミニ演習は<u>ありません</u>
- 中間レポートの解説配布と
  簡単な解説は来週する予定です

問1 a)

- **神戸市の高校生の平日睡眠時間は400分，標準偏差は80分である．いま，神戸市の高校生を母集団とし，その睡眠時間が正規分布 $N(400, 80^2)$に従うと仮定しよう．このとき，以下の問いa., b.に答えよ**

a. 神戸市の高校生の95%はどのくらい睡眠時間をとっているかを求めよ．すなわち，母集団におけるある学生の睡眠時間を Y とするとき，$P(a \leq Y \leq b) = 0.95$ となるa, bをそれぞれ求めよ．

問1 a)

a. 神戸市の高校生の95%はどのくらい睡眠時間をとっているかを求めよ．すなわち，母集団におけるある学生の睡眠時間を Y とするとき，$P(a \leq Y \leq b) = 0.95$ となるa, b をそれぞれ求めよ．

$P(-1.96 \leq Z \leq 1.96) = 0.95$ より

$P(400 - 1.96 \cdot 80 \leq Y \leq 400 - 1.96 \cdot 80) = 0.95$

$P(243.6 \leq Y \leq 596.4) = 0.95$

答え: **243.2** 分以上 **556.8** 分以下

問1 b)

b. この母集団からサンプルサイズ100の標本を無作為復元抽出し，標本平均 $\bar{X}$を求める．このとき，次の1. – 4.に答えよ．

- 標本平均 $\bar{X}$ の期待値 $E(\bar{X})$ を答えよ．
- 標本平均 $\bar{X}$ の標準偏差 $\sqrt{V(X)}$ を答えよ．

標本平均の性質その1

- 母平均 μ，母分散 σ^2 の母集団からサンプルサイズ n の標本 $X_1, \ldots, X_n$ を無作為復元抽出する．
このとき，標本平均 $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$
について以下が成り立つ

- 標本平均の期待値 $E(\bar{X}) = \mu$
- 標本平均の分散 $V(\bar{X}) = \dfrac{\sigma^2}{n}$
- 標本平均の標準偏差 $D(\bar{X}) = \sqrt{V(X)} = \dfrac{\sigma}{\sqrt{n}}$

問1 b)

b. この母集団からサンプルサイズ100の標本を無作為復元抽出し，標本平均 $\bar{X}$を求める． このとき，次の1. – 4.に答えよ．

- 標本平均 $\bar{X}$ の期待値 $E(\bar{X})$ を答えよ．
- 標本平均 $\bar{X}$ の標準偏差 $\sqrt{V(X)}$ を答えよ．

答え

- **期待値** $E(\bar{X}) = \mu =$ **400** **分**
- **標準偏差** $\sqrt{V(X)} = \frac{\sigma}{\sqrt{n}} = \frac{80}{\sqrt{100}} =$ 8 **分**

標本平均の性質その2

- 母平均 μ，母分散 σ^2 の母集団からサンプルサイズ n の標本 $X_1, \ldots, X_n$ を無作為復元抽出する．
  このとき，標本平均 $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$
  について以下が成り立つ

- 母集団分布が正規分布 $N(\mu, \sigma^2)$に従うとき，
  標本平均 $\bar{X}$ は 正規分布 $N\left(\mu, \frac{\sigma^2}{n}\right)$に従う
- 母集団分布が正規分布に従ってなくとも，
  n が十分大きければ $\bar{X}$ は $N(\mu, \frac{\sigma^2}{n})$ に従う

問1 b)

b. この母集団からサンプルサイズ100の標本を無作為復元抽出し，標本平均 $\bar{X}$を求める．このとき，
$P(c \leq \bar{X} \leq d) = 0.95$ となる c , d　をそれぞれ求めよ．

答え

標本平均 $\bar{X}$ は$N\left(\mu, \frac{\sigma^2}{n}\right) = N(400, 8^2)$に従うので

$P(-1.96 \leq Z \leq 1.96) = 0.95$ より

$$P\left(\mu - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$P(400 - 1.96 \cdot 8 \leq \bar{X} \leq 400 + 1.96 \cdot 8) = 0.95$$

$$P(384.32 \leq \bar{X} \leq 415.68) = 0.95$$

答え: **384.3** 分 以上 **415.7** 分 以下

問1 b)

b. この母集団からサンプルサイズ**10,000**の標本を無作為復元抽出し，標本平均 $\bar{X}$を求める．このとき，
$P(c \leq \bar{X} \leq d) = 0.95$ となる c , d　をそれぞれ求めよ．

答え

標本平均 $\bar{X}$ は$N\left(\mu, \frac{\sigma^2}{n}\right) = N(400, 0.8^2)$に従うので

$P(-1.96 \leq Z \leq 1.96) = 0.95$ より

$$P\left(\mu - 1.96 \cdot \frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 1.96 \cdot \frac{\sigma}{\sqrt{n}}\right) = 0.95$$

$$P(400 - 1.96 \cdot 0.8 \leq \bar{X} \leq 400 + 1.96 \cdot 0.8) = 0.95$$

$$P(398.432 \leq \bar{X} \leq 401.568) = 0.95$$

答え: **398.4** 分 以上 **401.6** 分 以下

問1 補足

- 95%となるようなcとdの幅 $d-c$ について考えてみると,

$$c = \mu - 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$d = \mu + 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

$$d - c = 2 \cdot 1.96 \cdot \frac{\sigma}{\sqrt{n}}$$

- つまり，サンプルサイズnが100倍 になれば，幅は $\frac{1}{10}$ になる.

問2

- **確率 $\frac{1}{4}$ で当選し $\frac{3}{4}$ ではずれるくじを復元抽出で n 回引くことを考える．これは，母集団が二項分布 $B(1, 0.25)$ に従うとき，この母集団からサンプルサイズ n の標本 $X_1, \cdots, X_n$ を無作為復元抽出していると考えることができる．このとき，$X_i\ (1 \leq i \leq n)$ は引いたくじが当たりであれば1，そうでなければ 0となるような確率変数である．また，標本中の当選回数をサンプルサイズ n で割った変数，すなわち $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n)$ を定義するとこれは標本平均の定義そのものである．**

- **出題意図**
  - 母集団と標本の関係が分かっているか？
  - 標本平均の期待値と分散は，いろいろな考え方から導ける

a.

- 母平均 μ を答えよ.

- **答え**
  - $B(n,p)$の期待値はnpである．いま，母集団は $B(1,0.25)$に従っているので，母平均 $\mu = \frac{1}{4}$

b.

- $X_i\ (1 \leq i \leq n)$ の期待値 $E(X_i)$と分散 $V(X_i)$ について答えよ.

- **答え**
  - 無作為復元抽出なので， X_iは母集団と同じ分布 $B\left(1,\frac{1}{4}\right)$ にしたがう．よって,
  - 期待値 $E(X_i) = \mu = \frac{1}{4}$
  - 分散 $V(X_i) = 1 \cdot \frac{1}{4} \cdot \left(1 - \frac{1}{4}\right) = \frac{3}{16}$

C.

- $X_1 + \cdots + X_n$ の期待値 $E(X_1 + \cdots + X_n)$ と分散 $V(X_1 + \cdots + X_n)$ を答えよ.

- 考え方1: 二項分布として考える
- 考え方2: 標本平均の性質を用いる
  - どちらの考え方で考えても同じ答え

考え方その1

- $X_1 + \cdots + X_n$ は結局のところ，くじを n回引いて当選する回数を表すので，

  $X_1 + \cdots + X_n$は $B\left(n, \frac{1}{4}\right)$ に従う

- **答え**
  - 期待値 $E(X_1 + \cdots + X_n) = n \cdot \frac{1}{4} = \frac{n}{4}$
  - 分散 $V(X_1 + \cdots + X_n) = n \cdot \frac{1}{4} \cdot \left(1 - \frac{1}{4}\right) = \frac{3}{16}n$

考え方その2

- 標本平均の性質より，（母集団が二項分布かどうかに関わらず）

  期待値$E(\bar{X}) = \mu = \frac{1}{4}$，分散$V(\bar{X}) = \frac{\sigma^2}{n} = \frac{1}{n} \cdot \frac{3}{16}$

- **答え**
  - $E(X_1 + \cdots + X_n) = nE(\bar{X})$より

    $E(X_1 + \cdots + X_n) = \frac{n}{4}$
  - $V(X_1 + \cdots + X_n) = n^2 V(\bar{X})$より，

    $V(X_1 + \cdots + X_n) = \frac{3}{16}n$

コメントへの返信

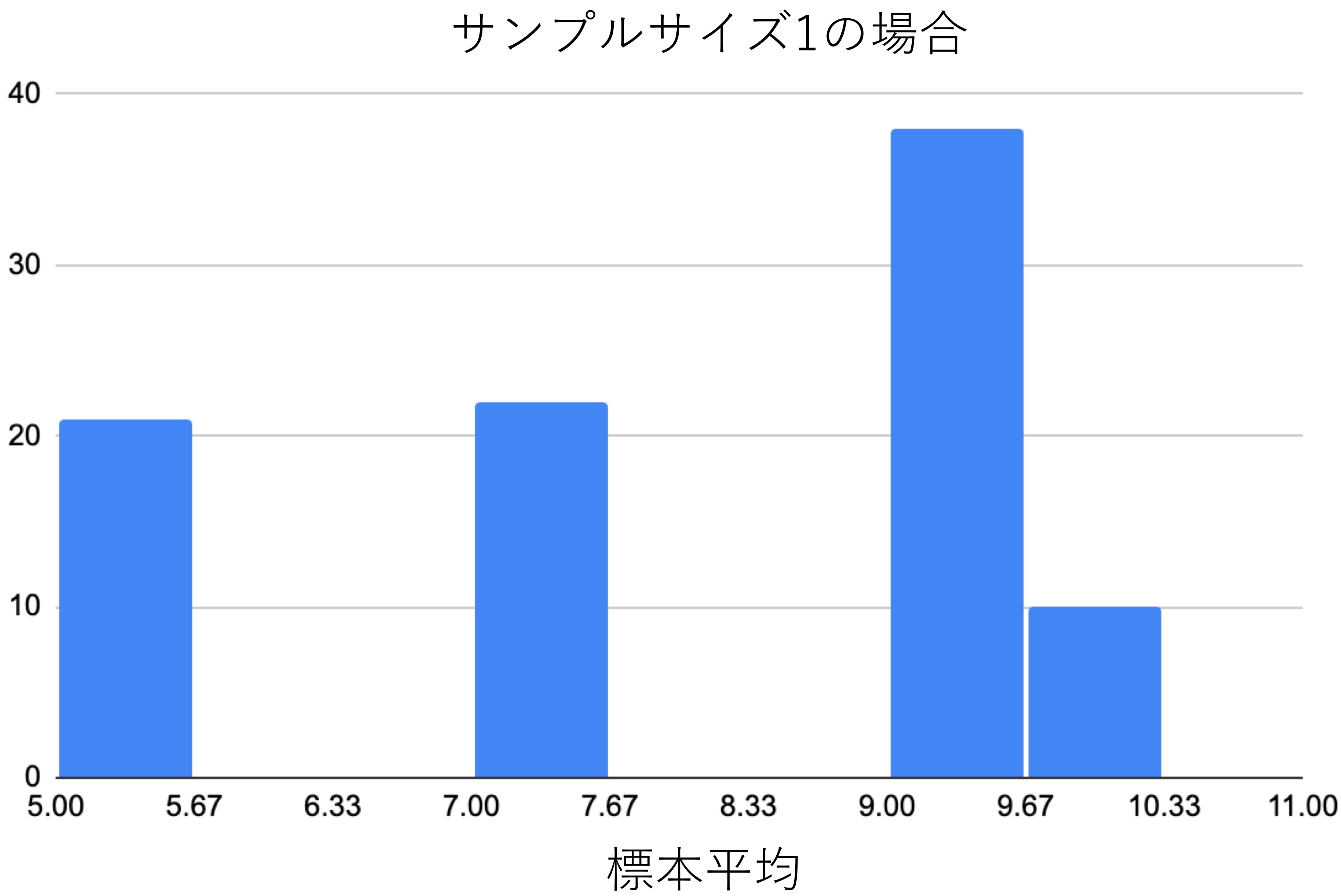
- 問2の考え方が難しい

- この辺から急に記号が増えて難しくなってきた

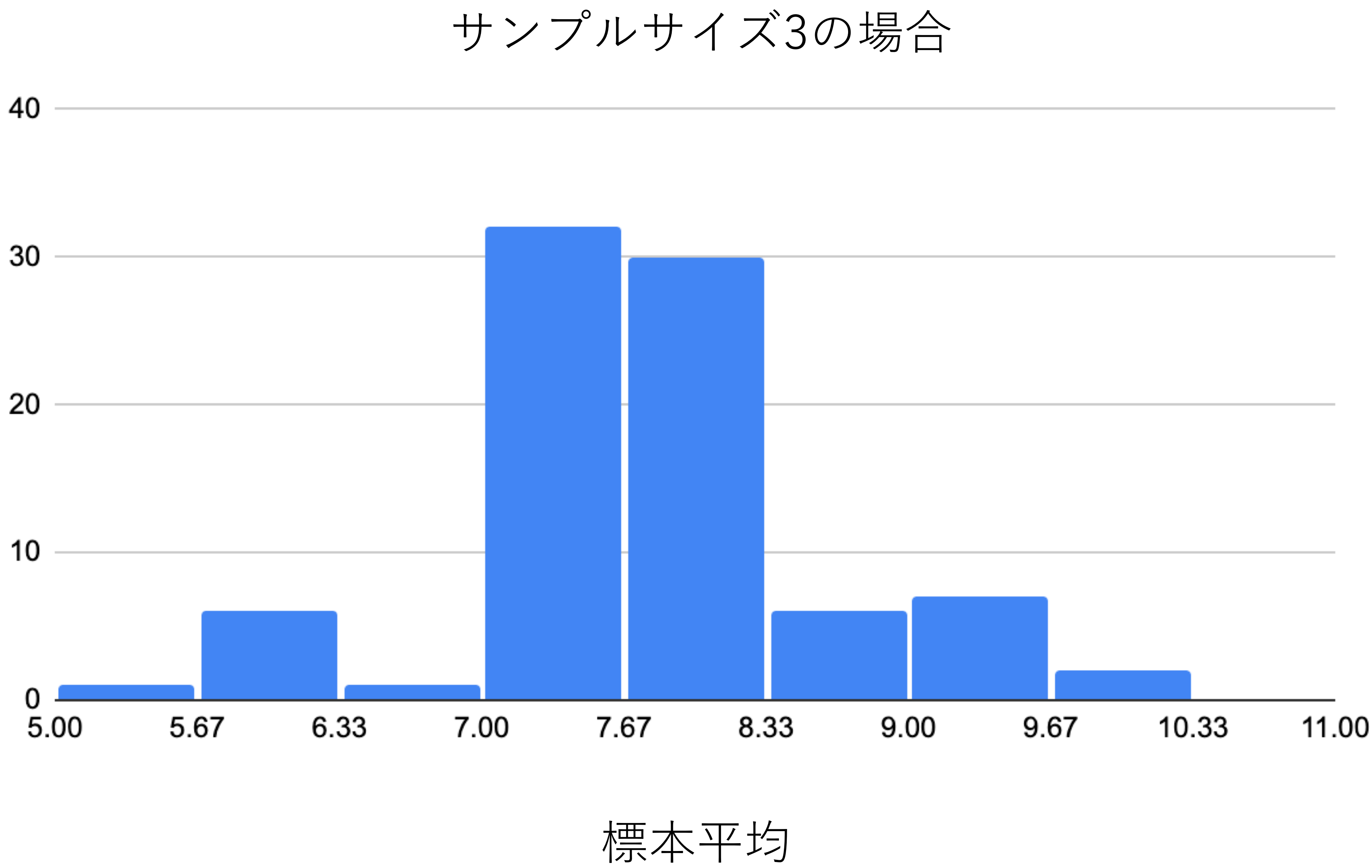
- Wordしんどい

大数の法則と中心極限定理の補足資料
（6月22日講義資料に掲載したものを再掲）

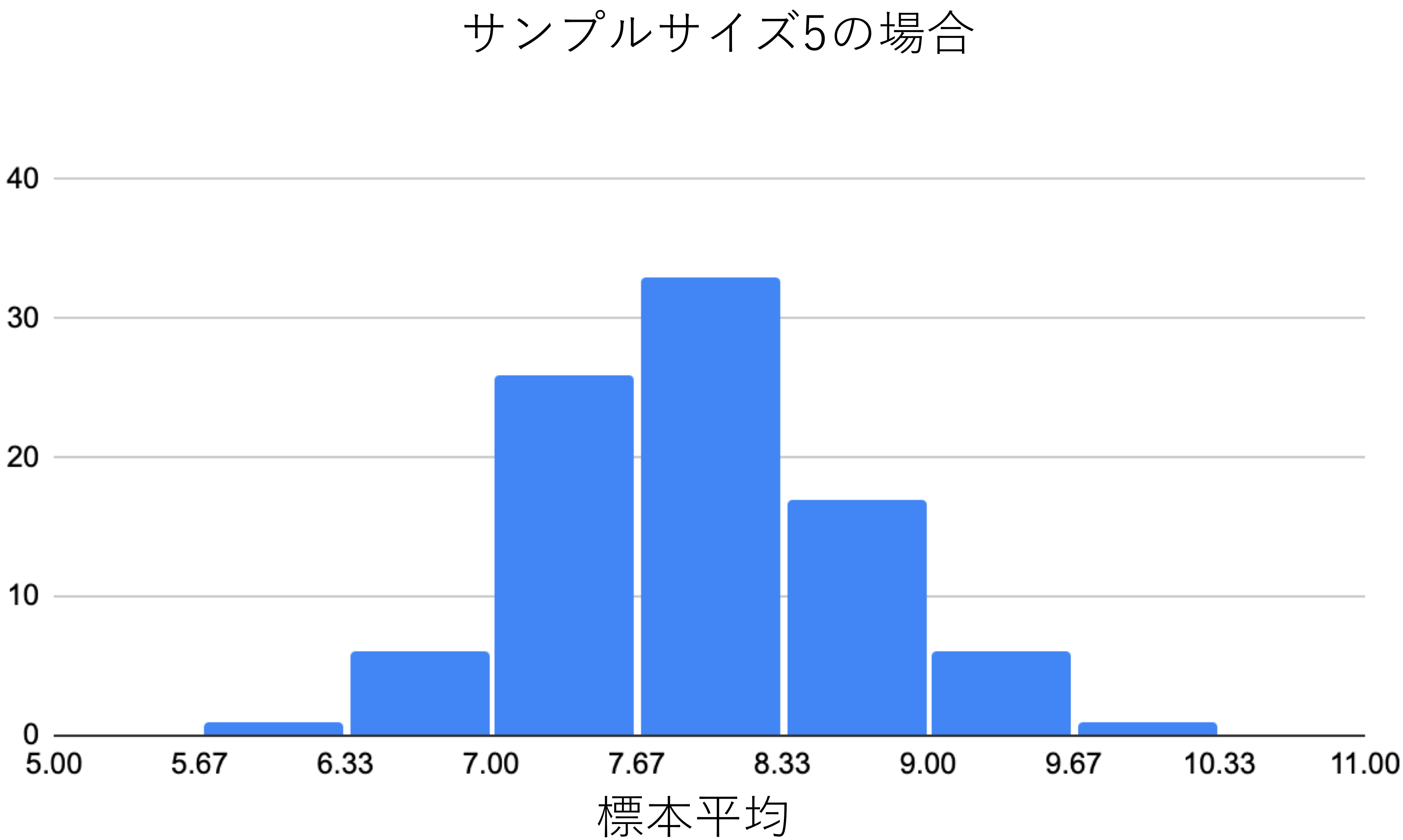
6月15日ミニ演習から 求めた標本平均 $\bar{X}$ のヒストグラム



6月15日ミニ演習から 求めた標本平均 $\bar{X}$ のヒストグラム



6月15日ミニ演習から
求めた標本平均 $\bar{X}$ のヒストグラム



だんだん正規分布に近づいていっていますよね？

見てとれること

- サンプルサイズnを大きくすると,
  標本平均が母平均（μ =8秒）付近となる確率が高くなっていく
  – これが，大数の法則
- サンプルサイズnを大きくすると,
  標本平均の分布は正規分布に近づいていく
  – これが，中心極限定理

今週の課題について

- 内容は講義ページよりPDFを確認してください
- 締切: 7月2日 21:00

- 信頼区間を求める問題がありますが，講義資料中の記号を当てはめて計算するだけでOKです．中身については次週講義します．