P.86〜

基本事項

- 調査対象全体からデータを集めて調べる方法を全数調査という．
- 調査の対象全体からその一部を抜き出して調べる方法を____________という．
- 標本調査において，調査の対象全体を____________，母集団に属する個々の対象を____________，個体の総数を____________という．
- **標本**とは，調査のために________から抜き出された__________のことで，母集団から標本を抜き出すことを________という．
- **標本の大きさ**とは，____________________のことである．
- 標本調査の目的は，抽出された標本から母集団の持つ性質を正しく推測することである．

標本調査の様子

母集団 $\xrightarrow{\text{抽出}}$ 標本

↓ 調査

母集団の性質 $\xleftarrow{\text{推測}}$ 標本の性質

- 母集団の各個体を等しい確率で抽出する方法を______________といい，無作為抽出によって選ばれた標本を______________という．
(以降断りがない限り全て無作為抽出とする．)
- 標本の抽出において，毎回戻しながら個体を 1 個ずつ抽出することを______________，個体をもとに戻さないで抽出することを______________という．
(母集団が十分大きいとき，非復元抽出を復元抽出とみなすことがある．)
- 母集団の変量 x と母集団における x の相対度数の一覧 (確率分布) を______________といい，母集団の平均を母平均，標準偏差を母標準偏差という．

X	x_1	x_2	$\cdots$	x_r	計
P	$\frac{f_1}{N}$	$\frac{f_2}{N}$	$\cdots$	$\frac{f_r}{N}$	1

練習 27

下の表は，40 枚の札に書かれた数字とその枚数である．40 枚を母集団，札の数字を変量とするとき，母集団分布を求めよ．また，母平均，母標準偏差を求めよ．

数字	1	2	3	4	5	計
枚数	2	6	24	6	2	40

解答

全体課題 pre/post

ある大都市の全世帯における 1 日の電力消費量 X（母集団）は，(母) 平均 $m = 50\text{kWh}$，(母) 標準偏差 $\sigma = 20\text{kWh}$ であることがわかっている．
調査員がこの都市から無作為に $n = 100$ 世帯を抽出し，その平均消費量（標本平均 $\overline{X}$）を計算した．この標本平均 $\overline{X}$ が，54kWh を超える確率 $P(\overline{X} > 54)$ を求めよ．

解答

♠ エキスパート A「標本平均とは何か.」

目標 A

標本平均が確率変数であることを理解し，母平均との違いを説明できる．また，大きさ n を十分大きくしていったときの標本平均と母平均の関係を説明できる．

定義 (標本平均)

母集団から大きさ n の無作為標本を抽出し，それらの値を $X_1, \cdots, X_n$ とする．このとき，

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

を標本平均という．

説明

- 調査員 A が $n = 100$ 世帯を調査して $\overline{X_A}$ を計算した．
- 翌日，調査員 B が別の $n = 100$ 世帯を調査して $\overline{X_B}$ を計算した．
- このとき，$\overline{X_A}$ と $\overline{X_B}$ の値は【 常に一致する / 一致するとは限らない 】．
- 調査 (抽出) のたびに，抽出された標本の値に応じて，$\overline{X}$ の値も変化する．
  $\rightarrow$ $\overline{X}$ は確率変数であると言える．
- (確率変数だから確率分布が考えられるんだよ!!)
- $\overline{X}$ が固定された値ではないのに対して，母平均 m は，【 1 つしか存在しない固定された値 / 複数の値が存在する 】．
- ある日，調査員 S がマンパワーを発揮し，大都市 (母集団) 全ての世帯を調査し標本平均 $\overline{X_s}$ 計算した．この標本平均 $\overline{X_s}$ は母平均 m に【 一致する / 一致するとは限らない 】．ただし，各世帯の電力消費量は日に依存しないものとする．

まとめ

標本平均は，取り出す標本に応じて値が変わる確率変数であり，ただ一つの値をとる母平均と一致するとは限らない．

♠ エキスパート B「標本平均の期待値 (平均) と標準偏差」

目標 B

母平均 m, 母標準偏差 σ の標本について，標本平均の期待値 $E(\overline{X})$ と標本平均の標準偏差 $\sigma(\overline{X})$ が

$$E(\overline{X}) = m, \quad \sigma(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$

で与えられることを示し，その性質を説明するできる．(標本平均の定義は A の資料をみよ．)

証明を兼ねた説明

- 母集団から大きさ n の無作為標本を抽出し，それらの値を $X_1, X_2, \cdots, X_n$ とする．
- このとき，これら X_1 から X_n は全て母集団分布に従う確率変数である．
  - 実際，例えば X_1 は母集団からランダムに取り出された 1 世帯目のワット数であり，その値が**どんな確率でどんな値を取るか**は母集団の分布の通りだよね．
  - 母集団分布に従うってことは期待値や標準偏差もそれぞれ母集団分布の値に等しくなる．
- ということで，これら X_1 から X_n の期待値と標準偏差はそれぞれ母平均, 母標準偏差に等しくなる．つまり，

$$E(X_1) = E(X_2) = \cdots = E(X_n) = m,$$
$$\sigma(X_1) = \sigma(X_2) = \cdots = \sigma(X_n) = \sigma.$$

- 従って次の結果を得る．

$$E(\overline{X}) = E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \underbrace{\frac{1}{n}}_{\text{定数倍は外へ}} \times \left(\underbrace{E(X_1) + E(X_2) + \cdots + E(X_n)}_{\text{期待値の加法性 } E(X+Y)=E(X)+E(Y)}\right)$$
$$= \frac{1}{n} \cdot n \cdot m = m.$$

$$V(\overline{X}) = V\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right) = \underbrace{\frac{1}{n^2}}_{\text{分散は 2 乗量}} \times \left(\underbrace{V(X_1) + V(X_2) + \cdots + V(X_n)}_{\text{復元抽出だから } X_1 \text{から } X_n \text{は互いに独立}}\right)$$
$$= \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n},$$
$$\sigma(\overline{X}) = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}.$$

- この結果から，標本平均の期待値 (平均) は母平均と【 一致し / 一致せず 】，標本数 n を大きくすればするほど標本平均の標準偏差は【 大きくなる / 小さくなる 】ことがわかる．

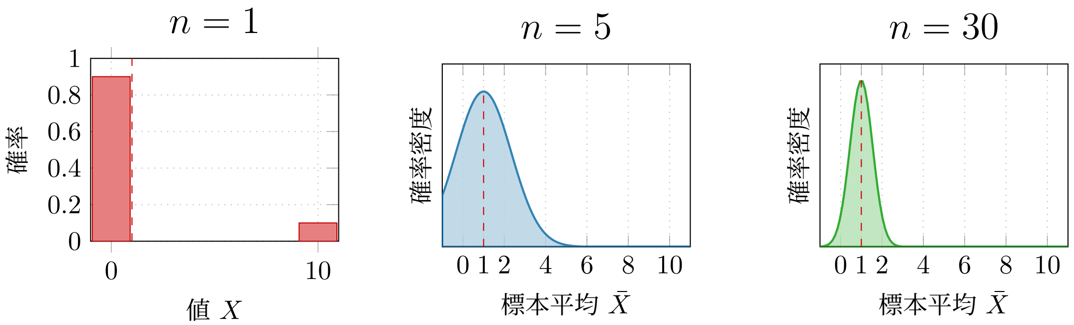
♠ エキスパート C「標本平均が従う分布 (中心極限定理)」

目標 C

母集団がどのような分布であっても，標本が大きければ，標本平均は近似的に正規分布に従うとみなせることを直感的に理解し，説明できる．

説明

- 例えば下の図は「90% の確率でハズレ (0 円)，10% の確率であたり (10 円)」の宝くじを n 回引く様子である．
- この宝くじの期待値 (母平均) は______円である．
- $n = 1$ は元の分布のままで，2 箇所に分かれた歪な分布になっている．
- $n = 5$ のグラフは，宝くじを 1 枚づつ 5 回引いたときの平均賞金 $\overline{X} = (X_1 + X_2 + \cdots + X_5)/5$ の分布を滑らかに繋いだものである．分布は左よりの歪みを残しているものの，先ほどよりは，母平均_____を中心とした山を形成している．
- $n = 30$ のグラフは，平均賞金 $\overline{X} = (X_1 + X_2 + \cdots + X_{30})/30$ の分布で，母平均を中心とした左右対称な釣鐘型の「正規分布」に近づいている．



- 母集団から大きさ n の無作為標本を抽出し，それらの値を $X_1, \cdots, X_n$ とするとき，

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

を標本平均という．
- 上の例では，賞金 X の従う母集団分布は【 正規分布とは言えない / 正規分布である 】が，それらを足して標本数で割った $\overline{X}$ は n が大きければ標本平均が従う分布は正規分布っぽいと【 言える / 言えない 】．
- より詳しく知りたい人はこちらの動画を参照せよ．→→→

まとめ

元の分布がどんなに歪な形でも標本数を大きくしていけば，標本平均の分布は正規分布に近づいていくことが知られている．この性質は**中心曲限定理**と呼ばれている．