統計的な推測(4)

ここでは,

- 母集団分布
- 標本平均の期待値，標準偏差
- 標本比率の期待値，標準偏差

について理解していきましょう。

基本１　母集団分布と標本平均の期待値，分散

(1)　母集団分布

(i)　全数調査，標本調査

- 全数調査　$\cdots$　対象となる集団全部のものを調査する方法。
- 標本調査　$\cdots$　対象となる集団の一部のみを調査し，全体の様子を推測する方法。

(ii)　母集団，標本，大きさ

- 母集団　$\cdots$　標本調査を行うときの対象となる集団全体。
- 標本　$\cdots$　選び出された要素の集合。
- 大きさ　$\cdots$　母集団の要素の個数を母集団の大きさ，標本の個数を標本の大きさという。

(iii)　母集団分布

大きさ N の母集団において，変量 X のとる異なる値を x_1，x_2，$\cdots$，x_n とし，それぞれの値をとる度数を f_1，f_2，$\cdots$，f_n (この合計が N) とする。

このとき，変量 X は次の表のような確率分布に従う確率変数となる。

X	x_1	x_2	$\cdots$	x_n	計
P	$\dfrac{f_1}{N}$	$\dfrac{f_2}{N}$	$\cdots$	$\dfrac{f_n}{N}$	1

この確率分布を母集団分布といい，この分布の平均，分散，標準偏差を，それぞれ，母平均，母分散，母標準偏差という。

(2)　抽出

(i)　復元抽出，非復元抽出

- 復元抽出　$\cdots$　毎回もとに戻しながら次のものを 1 個ずつ取り出すこと
- 非復元抽出　$\cdots$　取り出したものをもとに戻さずに続けて抽出すること

(ii)　注意点

母集団から抽出された大きさ n の無作為標本 X_1，X_2，$\cdots$，X_n は，

- 復元抽出によって得られた　$\rightarrow$　X_1，X_2，$\cdots$，X_n は独立
- 非復元抽出によって得られた　$\rightarrow$　X_1，X_2，$\cdots$，X_n は独立ではない

ただし，母集団の要素の数がきわめて大きいときには，非復元抽出でも X_1，X_2，$\cdots$，X_n は独立であるとみなしてよい。

つまり，それぞれが母集団分布に従う確率変数であるとしてよい。

(3) 標本平均の期待値，標準偏差

(i) 標本平均

母集団から大きさ n の無作為標本 X_1，X_2，$\cdots$，X_n を抽出する。

この資料 (データ) の平均，分散，標準偏差を，それぞれ，標本平均，標本分散，標本標準偏差という。

- 標本平均 $\overline{X}$ $\cdots$ $\overline{X}=\dfrac{1}{n}(X_1+X_2+\cdots+X_n)$
- 標本標準偏差 S $\cdots$ $S=\sqrt{\dfrac{1}{n}\sum_{k=1}^{n}(X_k-\overline{X})^2}$

$\overline{X}$，S もまた，抽出の結果により値が定まる確率変数となる。

(ii) 標本平均の期待値，標準偏差

母平均 m，母標準偏差 σ の母集団から大きさ n の無作為標本を抽出する。

このとき，標本平均 $\overline{X}$ の平均 $E(\overline{X})$，標準偏差 $\sigma(\overline{X})$ は，

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

証　母平均 m，母標準偏差 σ の母集団から大きさ n の無作為標本 X_1，X_2，$\cdots$，X_n を抽出する。各 X_k はそれぞれ母集団分布に従うから，

$$E(X_1)=E(X_2)=\cdots=E(X_n)=m$$

$$V(X_1)=V(X_2)=\cdots=V(X_n)=\sigma^2$$

したがって，

$$\begin{aligned}E(\overline{X})&=E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)\\&=\frac{1}{n}\{E(X_1)+E(X_2)+\cdots+E(X_n)\}\\&=m\end{aligned}$$

X_1，X_2，$\cdots$，X_n は独立であるから，

$$\begin{aligned}V(\overline{X})&=V\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)\\&=\frac{1}{n^2}\{V(X_1)+V(X_2)+\cdots+V(X_n)\}\\&=\frac{\sigma^2}{n}\end{aligned}$$

基本 2　標本平均の分布

(1)　標本平均の分布

母平均 m，母標準偏差 σ の母集団から大きさ n の無作為標本を抽出するとき，標本平均 $\overline{X}$ は，n が大きいとき，近似的に正規分布 $N\left(m,\ \dfrac{\sigma^2}{n}\right)$ に従う。

母集団分布が正規分布のときには，$\overline{X}$ は，常に正規分布 $N\left(m,\ \dfrac{\sigma^2}{n}\right)$ に従う。

(2)　大数の法則

母平均 m の母集団から大きさ n の無作為標本を抽出するとき，その標本平均 $\overline{X}$ は，n が大きくなるにつれて，母平均 m に近づく。

基本 3　標本比率の期待値，標準偏差

(1)　母比率，標本比率

- 母比率 p　$\cdots$　母集団において，ある特性 A をもつ要素の割合
- 標本比率 R　$\cdots$　標本の中で，ある特性 A をもつ要素の割合

(2)　標本比率の期待値，標準偏差

母比率 p の母集団から大きさ n の無作為標本を抽出する。

このとき，標本比率 R の平均 $E(R)$，標準偏差 $\sigma(R)$ は，

$$E(R) = p,\ \sigma(R) = \sqrt{\frac{p(1-p)}{n}}$$

証　(i)　(個数を全体の個数で割る考え方)

この母集団から大きさ n の無作為標本 X_1，X_2，$\cdots$，X_n を抽出するとき，性質 A をもつ個数を X とすると，X は二項分布 $B(n,\ p)$ に従うので，

$$E(X) = np,\ \sigma(X) = \sqrt{np(1-p)}$$

標本比率 R は，$R = \dfrac{X}{n}$ と表されるので，

$$\begin{aligned} E(R) &= E\left(\frac{X}{n}\right) \\ &= \frac{1}{n}E(X) \\ &= p \end{aligned}$$

$$\begin{aligned} \sigma(R) &= \sigma\left(\frac{X}{n}\right) \\ &= \left|\frac{1}{n}\right|\sigma(X) \\ &= \sqrt{\frac{p(1-p)}{n}} \end{aligned}$$

証 (ii) (標本比率を標本平均と捉える考え方)

性質 A をもつとき 1 を，もたないとき 0 を値にとる変量 X を考える。

母比率 p の母集団から，変量 X に関する大きさ n の無作為標本

X_1，X_2，$\cdots$，X_n を抽出するとき，

$$\frac{X_1+X_2+\cdots+X_n}{n}$$

の値は R であるから，R は X についての標本平均である。

ここで，X についての母集団分布は，次の通りである。

X	0	1	計
P	$1-p$	p	1

よって，

$$E(X)=0\cdot(1-p)+1\cdot p$$
$$=p$$
$$V(X)=(-p)^2\cdot(1-p)+(1-p)^2\cdot p$$
$$=p(1-p)$$
$$\sigma(X)=\sqrt{p(1-p)}$$

母平均が p，母標準偏差が $\sqrt{p(1-p)}$，標本の大きさが n であるから，

$$E(R)=p,\ \sigma(R)=\sqrt{\frac{p(1-p)}{n}}$$

1・A

2 と書かれたカードが 2 枚，4 と書かれたカードが 3 枚ある。これを母集団とし，札の数字を変量 X とする。

(1) X の母集団分布，母平均 m，母標準偏差 σ を求めよ。

(2) この母集団から，復元抽出により大きさ 2 の標本 X_1，X_2 を抽出する。この標本平均 $\overline{X} = \dfrac{X_1 + X_2}{2}$ について，期待値 $E(\overline{X})$，標準偏差 $\sigma(\overline{X})$ を求めよ。

2・A

母平均 120，母標準偏差 30 をもつ母集団から，大きさ 100 の無作為標本を抽出するとき，その標本平均を $\overline{X}$ とする。

(1) 標本平均 $\overline{X}$ の期待値 $E(\overline{X})$，標準偏差 $\sigma(\overline{X})$ を求めよ。

(2) $P(\overline{X} > 123)$ を求めよ。

3・A

母比率が $\dfrac{1}{3}$ である母集団から，大きさ 200 の無作為標本を抽出するとき，その標本比率を R とする。

(1) 標本比率 R の期待値 $E(R)$，標準偏差 $\sigma(R)$ を求めよ。

(2) $P(0.33 \leqq R \leqq 0.34)$ を求めよ。

4・B

　17 歳の男子の身長は，平均 166 cm，標準偏差 12 cm の正規分布に従うとする。無作為に抽出された 17 歳の男子 100 名の平均身長が 167.5 cm 以上 169 cm 以下である確率を求めよ。

5・B

　A 市の新生児の男子と女子の割合は等しいことがわかっている。ある年の A 市の新生児の中から 100 人を無作為抽出したときの女子の割合を R とする。

(1)　標本比率 R の期待値 $E(R)$ と標準偏差 $\sigma(R)$ を求めよ。

(2)　標本比率が 50 %以上 57 %以下である確率を求めよ。

6・B

　ある国では，その国民の血液型の割合は

　O 型 30 %，A 型 35 %，B 型 25 %，AB 型 10 %

であるといわれている。

　いま，無作為に 400 人を選ぶとき，AB 型の人が 37 人以上 49 人以下となる確率を求めよ。

7・C

ある県において，参議院議員選挙における有権者の A 政党支持率は 30 %であるという。この県の有権者の中から，無作為に n 人を抽出するとき，k 番目に抽出された人が A 政党支持なら 1，不支持なら 0 の値を対応させる確率変数を X_k とする。

(1)　標本平均 $\overline{X} = \dfrac{X_1 + X_2 + \cdots + X_n}{n}$ について，期待値 $E(\overline{X})$ を求めよ。

(2)　標本平均 $\overline{X}$ の標準偏差 $\sigma(\overline{X})$ を 0.02 以下にするためには，抽出される標本の大きさは，少なくとも何人以上必要であるか。

8・C

適した問題が見つかり次第，掲載します。
気長にお待ちください。

9・C

確率変数 Z の期待値を $E(Z)$，分散を $V(Z)$ とする。

(1)　任意の正の数 k に対して，$|\,Z - E(Z)\,| \geqq k$ ならば 1 を，$|\,Z - E(Z)\,| < k$ ならば 0 を割り当てる確率変数 G を考える。G の期待値を $E(G)$ とすると，

$$E(G) = P(|\,Z - E(Z)\,| \geqq k)$$

であることを示せ。

(2)　(1) で定めた G に対して，次の不等式が成り立つことを示せ。

$$k^2 G \leqq |\,Z - E(Z)\,|^2$$

(3)　次の不等式が成り立つことを示せ。

$$P(|\,Z - E(Z)\,| \geqq k) \leqq \frac{V(Z)}{k^2}$$

(4)　母平均 m，母標準偏差 σ をもつ母集団から，大きさ n の標本 X_1，X_2，$\cdots$，X_n を抽出する。その標本平均 $\overline{X}$ と任意の正の実数 k に対して

$$\lim_{n\to\infty} P(|\,\overline{X} - m\,| < k) = 1$$

が成り立つことを示せ。