5. 母平均や母比率も，標本から求められる！

“**確率分布と統計的推測**”の講義も，今回で最後になる。最後のテーマは，“**統計的推測**”だ。統計とは，ある国の国民の身長や体重など…，複数の数値データの集まりを，表やグラフにしたり，その平均や分散を求めたりすることなんだね。

ある集団（**母集団**）について，ある変量（身長や体重など）の統計調査を行うとき，母集団すべてを調べる“**全数調査**”と，母集団の $\mathbf{1}$ 部を**標本**（**サンプル**）として抽出して調べ，その結果を基に元の母集団の状態を推測する“**標本調査**”とがある。

ここでは，標本調査による統計的な推測の仕方について詳しく教えよう。それでは，講義を始めよう！

● 推測統計のイメージを頭に入れよう！

“**統計**”とは，数値データ，すなわち“**変量**”を処理して，表やグラフを作ったり，平均や分散を求める手法のことなんだ。そして，調査の対象全体を“**母集団**”という。そして，

（ⅰ）母集団の要素の個数が小さいときは，母集団全体を全数調査することができる。これを**記述統計**という。これに対して，

（ⅱ）**100** 万個とか **1000** 万個とか，母集団の要素の個数が膨大であるときは，手間と費用の面から全数調査は難しいので，母集団から無作為に適当な数の**標本**（**サンプル**）を抽出し，これを基にして，母集団の分布の特徴を推測する。これを**標本調査**といい，このような統計手法を**推測統計**というんだね。

ここで，母集団から標本を無作為に抽出する手法として，

（ⅰ）要素を **1** 個取り出しては元に戻し，また新たに **1** 個を取り出すことを繰り返す“**復元抽出**”と，

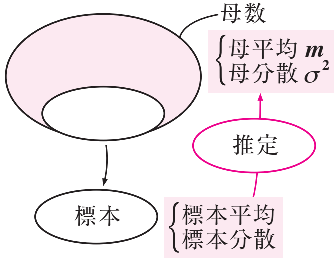
（ⅱ）取り出した要素を元に戻すことなしに，次々と要素を取り出す“**非復元抽出**”の，**2** 通りがある。

でも，母集団の大きさ N が標本の大きさ n に対して，十分に大きければ，

たとえば，$N=100$ 万個の母集団から，$n=1000$ 個程度の標本を取り出すような場合ならば，非復元抽出であっても，復元抽出とみなしても構わない。何故なら，非復元抽出で **1** 個ずつ標本を取り出していっても，母集団の大きさ N が大きければ，その都度母集団の性質が変化することはほとんどないからだ。よって，この講義ではすべて復元抽出と考えることにしよう。

ここで，母集団の特徴を表す平均と分散を特に**母平均**，**母分散**と呼び，それぞれ m と σ^2 で表す。そして，これらを母集団を特徴づける数値としてまとめて**母数**という。これに対して，復元抽出された標本の平均と分散はそれぞれ**標本平均**，**標本分散**と呼ぶことにする。そして，図 **1** に示すように，この標本平均や標本分散を使って，母数を推測するのが推測統計なんだ。ここでは，この推測統計の中の**区間推定**について解説することにしよう。

図 1　母平均・母分散と標本平均・標本分散



● 母集団と標本の関係を具体的に調べよう！

ある地域の **500** 万世帯を母集団として，各世帯が所有する自転車の台数を調べたところ，**0** 台，**1** 台，**2** 台の世帯数が順に **100** 万，**200** 万，**200** 万世帯であったとする。ここで，変量 X を自転車の保有台数とすると，$X=0,\ 1,\ 2$ であり，それぞれの確率が，

$$P(X=0)=\frac{100\text{万}}{500\text{万}}=\frac{1}{5}\ ,\ \text{同様に}\quad P(X=1)=\frac{2}{5}\ ,\ P(X=2)=\frac{2}{5}$$

となるので，母集団の確率分布，すなわち**母集団分布**は表 **1** のようになる。これから，母平均 $m=E(X)$ と母分散 $\sigma^2=V(X)=E(X^2)-\{E(X)\}^2$ を次のように求めることができる。

表 1　母集団分布

X	0	1	2
P	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{2}{5}$

母平均 $m=E(X)=\cancel{0\times\frac{1}{5}}+1\times\frac{2}{5}+2\times\frac{2}{5}=\frac{2+4}{5}=\frac{6}{5}$

母分散 $\sigma^2 = V(X) = \underbrace{\cancel{0^2 \times \frac{1}{5}} + 1^2 \times \frac{2}{5} + 2^2 \times \frac{2}{5}}_{E(X^2)} - \underbrace{\left(\frac{6}{5}\right)^2}_{\{E(X)\}^2} = \frac{10}{5} - \frac{36}{25} = \frac{14}{25}$

どう？このように母集団分布にまとめて，母平均 m と母分散 σ^2 を求めてみると，これまでに確率分布で計算したものとまったく同様であることに気付いたはずだ。つまり，自転車の保有台数 X の母集団分布は，0 と 1 と 2 の数の書かれたカードがそれぞれ 1 枚，2 枚，2 枚あって，これら 5 枚のカードから，1 枚のカードを無作為に選び出したとき，カードに書かれた数を変量 X とおいた確率分布とまったく同じなんだね。

母集団分布

X	0	1	2
P	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{2}{5}$

そして，この母集団分布から，大きさ n の標本を復元抽出する操作は，先程の 5 枚のカードから，1 枚を取り出しては元に戻し，新たに 1 枚取り出す操作を n 回繰り返すことと同じなんだね。ここで，k 回目に取り出したカードに書かれている数を変量 X_k $(k = 1, 2, \cdots, n)$ とおくと，X_1，X_2，…，X_n はすべて，母集団分布と同じ確率分布に従う。よって，次式が成り立つ。

$$\begin{cases} E(X_1) = E(X_2) = \cdots = E(X_n) = m(\text{母平均}) & \cdots\cdots① \\ V(X_1) = V(X_2) = \cdots = V(X_n) = \sigma^2(\text{母分散}) & \cdots\cdots② \end{cases}$$ となる。

ここで，さらに，大きさ n の**標本平均**を $\overline{X}$ とおくと，

$$\overline{X} = \frac{X_1 + X_2 + \cdots + X_n}{n} \quad \cdots\cdots③$$

となるのはいいね。では，この $\overline{X}$ の平均 $E(\overline{X})$ と分散 $V(\overline{X})$ を求めてみることにしよう。

ここで，復元抽出なので，n 個の確率変数 X_1，X_2，…，X_n は互いにすべて独立な確率変数と考えていいんだね。よって，公式：

$$\begin{cases} E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n) \\ V(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1{}^2V(X_1) + a_2{}^2V(X_2) + \cdots + a_n{}^2V(X_n) \end{cases}$$

$(a_1, a_2, \cdots, a_n : \text{実数定数})$

を用いることができるので，

$$\cdot\ E(\overline{X})=E\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)=E\left(\underbrace{\frac{1}{n}}_{a_1}X_1+\underbrace{\frac{1}{n}}_{a_2}X_2+\cdots+\underbrace{\frac{1}{n}}_{a_n}X_n\right)$$

$$=\frac{1}{n}\underbrace{E(X_1)}_{m}+\frac{1}{n}\underbrace{E(X_2)}_{m}+\cdots+\frac{1}{n}\underbrace{E(X_n)}_{m\ (①より)}$$

$$=\underbrace{\frac{m}{n}+\frac{m}{n}+\cdots+\frac{m}{n}}_{n\ 項の和}=\cancel{n}\times\frac{m}{\cancel{n}}=m(\text{母平均})$$ となるし，

$$\cdot\ V(\overline{X})=V\left(\frac{X_1+X_2+\cdots+X_n}{n}\right)=V\left(\underbrace{\frac{1}{n}}_{a_1}X_1+\underbrace{\frac{1}{n}}_{a_2}X_2+\cdots+\underbrace{\frac{1}{n}}_{a_n}X_n\right)$$

$$=\frac{1}{n^2}\underbrace{V(X_1)}_{\sigma^2}+\frac{1}{n^2}\underbrace{V(X_2)}_{\sigma^2}+\cdots+\frac{1}{n^2}\underbrace{V(X_n)}_{\sigma^2\ (②より)}$$

$$=\underbrace{\frac{\sigma^2}{n^2}+\frac{\sigma^2}{n^2}+\cdots+\frac{\sigma^2}{n^2}}_{n\ 項の和}=n\times\frac{\sigma^2}{n^2}=\frac{\sigma^2}{n}\left(=\frac{母分散}{n}\right)$$ となる。

もちろん，$\overline{X}$ の標準偏差は，$\sigma(\overline{X})=D(\overline{X})=\sqrt{\dfrac{\sigma^2}{n}}=\dfrac{\sigma}{\sqrt{n}}$ となる。

標本平均 $\overline{X}$ の期待値，分散

母平均 m，母分散 σ^2 の大きさ N の母集団から，大きさ n の標本 X_1，X_2，…，X_n を無作為に抽出したとき，

標本平均 $\overline{X}=\dfrac{X_1+X_2+\cdots+X_n}{n}$ の平均 $E(\overline{X})$，分散 $V(\overline{X})$，標準偏差 $D(\overline{X})$ をそれぞれ，$m(\overline{X})$，$\sigma^2(\overline{X})$，$\sigma(\overline{X})$ とおくと，

$m(\overline{X})=E(\overline{X})=m$ ……(＊1)　$\sigma^2(\overline{X})=V(\overline{X})=\dfrac{\sigma^2}{n}$ ……(＊2)

$\sigma(\overline{X})=D(\overline{X})=\dfrac{\sigma}{\sqrt{n}}$ …(＊3)　(ただし，$N\gg n$ とする。)

◆例題 14◆

母平均 $m = 50$，母分散 $\sigma^2 = 25$ の母集団から，大きさ $n = 100$ の標本を抽出したとき，標本平均 $\overline{X}$ の平均 $m(\overline{X})$ と分散 $\sigma^2(\overline{X})$ を求めよ。
(ただし，母集団の大きさを N とおくと，$N \gg n$ とする。)

標本平均 $\overline{X}$ の平均 $m(\overline{X}) = m$，分散 $\sigma^2(\overline{X}) = \dfrac{\sigma^2}{n}$ より

平均 $m(\overline{X}) = m = 50$　分散 $\sigma^2(\overline{X}) = \dfrac{\sigma^2}{n} = \dfrac{25}{100} = \dfrac{1}{4} = 0.25$　となる。

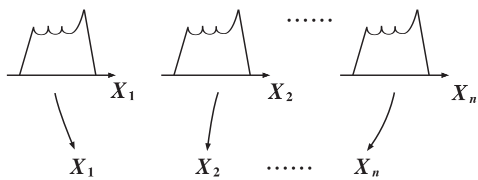
● n が十分に大きければ，中心極限定理が使える！

母平均 m，母分散 σ^2 をもつ母集団から，大きさ n の標本を抽出すると，その標本平均 $\overline{X}$ の平均 $m(\overline{X})$ は m に，また分散 $\sigma^2(\overline{X})$ は $\dfrac{\sigma^2}{n}$ になることは教えたわけだけれど，ここで，n を $n = 50$，100，…と十分に大きくすると，P165 で解説した“中心極限定理(ちゅうしんきょくげんていり)”により，図 2 に示すように，標本平均 $\overline{X}$ の従う確率分布は，平均 m，分散 $\dfrac{\sigma^2}{n}$ の正規分布 $N\left(m,\ \dfrac{\sigma^2}{n}\right)$ に近づくことが分かっている。

ということは，変数 $\overline{X}$ から平均 m を引いて，その標準偏差 $\dfrac{\sigma}{\sqrt{n}}$ で割って，新たな標準化変数を Z，すなわち，

図 2　中心極限定理のイメージ

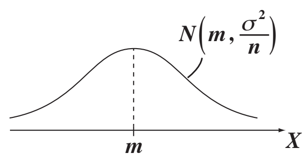
母平均 m，母分散 σ^2 をもつ同一の母集団
(正規分布でなくてもいい)



$\overline{X} = \dfrac{X_1 + X_2 + \cdots\cdots + X_n}{n}$ とおき，

n を十分大きくすると，$\overline{X}$ は

正規分布 $N\left(m, \dfrac{\sigma^2}{n}\right)$ に従う。



$Z=\dfrac{\overline{X}-m}{\dfrac{\sigma}{\sqrt{n}}}$ とおくと，Z は平均 0，分散 1 の標準正規分布 $N(0,\ 1)$ に従うことになるんだね。

例題 14 では，$n=100$ は十分に大きいと考えられるので，$\overline{X}$ は正規分布 $N\left(\underbrace{50}_{m},\ \underbrace{\dfrac{1}{4}}_{\frac{\sigma^2}{n}}\right)$ に従う。よって，この標準化変数 $Z=\dfrac{\overline{X}-50}{\sqrt{\dfrac{1}{4}}}=\dfrac{\overline{X}-50}{\dfrac{1}{2}}$

$=2(\overline{X}-50)$ は，標準正規分布 $N(0,\ 1)$ に従うことになるんだね。大丈夫？

この確率密度 $f_s(z)=\dfrac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$

ここで，標準正規分布 $N(0,\ 1)$ の確率表として，表 2 に示した $u=1.96$ と $u=2.58$ の値がとても重要となる。それは，図 3(ⅰ) に示すように $Z\geqq1.96$ となる確率が 2.5% いうことは，標準正規分布の対称性から，$-1.96\leqq Z\leqq1.96$ となる確率が 95% となることを示しているからであり，同様に，図 3(ⅱ) から，$Z\geqq2.58$ となる確率が 0.5% ということは，$-2.58\leqq Z\leqq2.58$ となる確率が 99% になることを示しているからなんだね。

これを基に，母平均 m の“95% **信頼区間**(しんらいくかん)”や“99% **信頼区間**(しんらいくかん)”などを導くことができるので，±1.96 と ±2.58 の値はシッカリ頭に入れておこう。

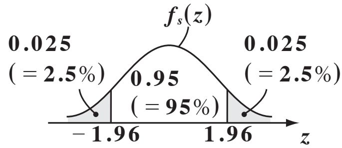
表 2 標準正規分布の確率表

$$\alpha=\int_u^{\infty}f_s(z)\,dz$$

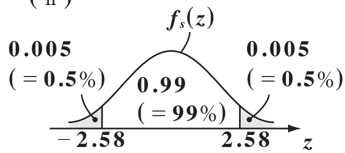
u	α
1.96	0.025
2.58	0.005

図 3 $Z=1.96$ と $Z=2.58$ の意味

(ⅰ)



(ⅱ)



● 母平均 m の区間推定をしてみよう！

ここで，母集団の母分散 σ^2（または，標準偏差 σ）は既知（きち）で，母平均 m

"分かっている" の意味

が未知（みち）の場合，大きさ n の標本から求めた標本平均 $\overline{X}=\left(\dfrac{X_1+X_2+\cdots+X_n}{n}\right)$

"分かっていない" の意味

を基に，母平均 m の（Ⅰ）95% "**信頼区間**（しんらいくかん）" と，（Ⅱ）99% "**信頼区間**" を求めてみよう。

（Ⅰ）母平均 m の 95% 信頼区間について，

$\overline{X}$ の標準化変数 Z は，$Z=\dfrac{\overline{X}-m}{\dfrac{\sigma}{\sqrt{n}}}$ ……①であり，Z は標準正規分布に従うので，Z が 95% の確率で存在し得る範囲は，前述した通り $-1.96\leqq Z\leqq 1.96$ となる。つまり，次式のようになる。

$P(-1.96\leqq Z\leqq 1.96)=0.95$ …② となる。よって，（ ）内を変形すると，

各辺に $\dfrac{\sigma}{\sqrt{n}}$ をかけて

$-1.96\leqq\dfrac{\overline{X}-m}{\dfrac{\sigma}{\sqrt{n}}}\leqq 1.96$ より，$\underbrace{-1.96\dfrac{\sigma}{\sqrt{n}}\leqq\overline{X}-m}_{(ア)}\leqq 1.96\dfrac{\sigma}{\sqrt{n}}$ （$\overline{X}-m\leqq 1.96\dfrac{\sigma}{\sqrt{n}}$：（イ））

（ア）より，$m\leqq\overline{X}+1.96\dfrac{\sigma}{\sqrt{n}}$　（イ）より，$\overline{X}-1.96\dfrac{\sigma}{\sqrt{n}}\leqq m$

②は，$P\left(\overline{X}-1.96\dfrac{\sigma}{\sqrt{n}}\leqq m\leqq\overline{X}+1.96\dfrac{\sigma}{\sqrt{n}}\right)=0.95$ ……③ となる。

$\overline{X}$ と σ と n は既知だから，母平均 m が 95% の確率で存在する範囲，すなわち "95% **信頼区間**（しんらいくかん）" が，

$$\overline{X}-1.96\frac{\sigma}{\sqrt{n}}\leqq m\leqq\overline{X}+1.96\frac{\sigma}{\sqrt{n}}\quad\cdots\cdots(*1)$$

と導かれるんだね。

（Ⅱ）母平均 m の 99% 信頼区間について，

同様に，Z が 99% の確率で存在し得る範囲は，$-2.58\leqq Z\leqq 2.58$ より，

$P(-2.58\leqq Z\leqq 2.58)=0.99$ ……④

④の Z に①を代入して，( )内を同様に変形すれば，

$$P\left(\overline{X}-2.58\frac{\sigma}{\sqrt{n}}\leqq m\leqq \overline{X}+2.58\frac{\sigma}{\sqrt{n}}\right)=0.99 \quad\cdots\cdots⑤$$ となるので，

母平均 m が 99% の確率で存在する範囲，すなわち“**99% 信頼区間**”が，

$$\overline{X}-2.58\frac{\sigma}{\sqrt{n}}\leqq m\leqq \overline{X}+2.58\frac{\sigma}{\sqrt{n}} \quad\cdots\cdots(*2)$$ と導かれるんだね。

このように，母平均 m などの 95% や 99% の信頼区間を求めることを，まとめて**区間推定**(くかんすいてい)と呼び，この 95% や 99% の確率を**信頼度**(しんらいど)という。

でも，m だけ未知で，σ (または，σ^2) が既知というのは，変な設定だね。ここで，m も σ (または，σ^2) も共に未知の場合でも，標本の大きさ n が十分大きければ，近似的に $(*1)$ や $(*2)$ の母標準偏差 σ の代わりに**標本標準偏差** S で代用できることが分かっているんだね。
以上をまとめて示そう。

母平均 m の区間推定

(Ⅰ) 母標準偏差 σ が既知のとき，

(ⅰ) 母平均 m の 95% 信頼区間は，次のようになる。

$$\overline{X}-1.96\frac{\sigma}{\sqrt{n}}\leqq m\leqq \overline{X}+1.96\frac{\sigma}{\sqrt{n}} \quad\cdots\cdots(*1)$$

(ⅱ) 母平均 m の 99% 信頼区間は，次のようになる。

$$\overline{X}-2.58\frac{\sigma}{\sqrt{n}}\leqq m\leqq \overline{X}+2.58\frac{\sigma}{\sqrt{n}} \quad\cdots\cdots(*2)$$

(Ⅱ) 母標準偏差 σ が未知のとき，

(ⅰ) 母平均の 95% 信頼区間は，次のように近似できる。

$$\overline{X}-1.96\frac{S}{\sqrt{n}}\leqq m\leqq \overline{X}+1.96\frac{S}{\sqrt{n}} \quad\cdots\cdots(*1)'$$

(ⅱ) 母平均の 99% 信頼区間は，次のようになる。

$$\overline{X}-2.58\frac{S}{\sqrt{n}}\leqq m\leqq \overline{X}+2.58\frac{S}{\sqrt{n}} \quad\cdots\cdots(*2)'$$

(ただし，$\overline{X}$：標本平均，S：標本標準偏差)

● 母比率の推定も押さえておこう！

たとえば，大量生産された工業製品の不良品の割合や，ある国の全有権者の X 政党への支持率のように，ある性質をもつものの全体に対する割合を，母集団の場合は**母比率**(ほひりつ)と呼び p で，また，大きさ n の標本の場合は**標本比率**(ひょうほんひりつ)と呼び $\overline{p}$ で表すことにしよう。そして，n と $\overline{p}$ を用いて，母比率 p の "**95% 信頼区間**" や "**99% 信頼区間**" を区間推定してみよう。

不良品や政党支持率など，ある性質 A に対して母比率 p をもつ母集団から，大きさ n の標本を無作為に抽出する場合，1 つ 1 つの標本を n 回抽出

これは，母集団の大きさ N が十分大きいものとして，非復元でも復元と考えていい。

すると考えると，これは事象 A が n 回中 r 回起こる反復試行の確率 P_r を求めることと同様なことに気付くはずだ。

つまり，1 回の試行 (抽出) で事象 A の起こる確率が母比率 p，起こらない確率が $q(=1-p)$ であり，n 回中 r 回だけ事象 A の起こる確率と同様に，n 個の標本中 r 個だけ A の性質をもつ確率 P_r は，

$P_r = {}_n\mathrm{C}_r\, p^r q^{n-r}$ $(r=0,\ 1,\ 2,\ \cdots,\ n)$ となるんだね。

よって，確率変数 X を $X=r$ $(r=0,\ 1,\ 2,\ \cdots,\ n)$ とおくと，X は二項分布 $B(n,\ p)$ に従い，さらに n が十分に大きければ，これは近似的に平

この平均は np，分散は $npq=np(1-p)$ だね。

均 np，分散 $np(1-p)$ の正規分布 $N(np,\ np(1-p))$ に従うことになるんだね。さらに，n が十分に大きいときは，分散 $np(1-p)$ の p を近似的に標本比率 $\overline{p}$ でおきかえてもよいことが分かっているので，X は結局，正規分布 $N(np,\ n\overline{p}(1-\overline{p}))$ に従うと言える。

であれば，X から平均 np を引いて，標準偏差 $\sqrt{n\overline{p}(1-\overline{p})}$ で割った標準化変数 $Z\left(=\dfrac{X-np}{\sqrt{n\overline{p}(1-\overline{p})}}\right)$ は，標準正規分布 $N(0,\ 1)$ に従う。よって，Z が 95% 存在する範囲や 99% 存在する範囲から，母比率 p の "**95% 信頼区**

$-1.96 \leqq Z \leqq 1.96$　　$-2.58 \leqq Z \leqq 2.58$

間" や "**99% 信頼区間**" を次のように求めることができる。

(Ⅰ) 母比率 p の 95% 信頼区間について，

$P(\underline{-1.96 \leqq Z \leqq 1.96}) = 0.95$ …① より，左辺の ( ) 内を変形して，

$$-1.96 \leqq \frac{X-np}{\sqrt{n\overline{p}(1-\overline{p})}} \leqq 1.96 \text{ より，} \underbrace{-1.96\sqrt{n\overline{p}(1-\overline{p})} \leqq X - np}_{(ア)} \leqq 1.96\sqrt{n\overline{p}(1-\overline{p})}$$

$$\underbrace{X - np \leqq 1.96\sqrt{n\overline{p}(1-\overline{p})}}_{(イ)}$$

(ア) より，$np \leqq X + 1.96\sqrt{n\overline{p}(1-\overline{p})}$　両辺を n で割って，$\dfrac{X}{n} = \overline{p}$ より，

$$p \leqq \frac{X}{n} + 1.96\frac{\sqrt{n\overline{p}(1-\overline{p})}}{n} \quad \therefore\ p \leqq \overline{p} + 1.96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}$$

(イ) より，$X - 1.96\sqrt{n\overline{p}(1-\overline{p})} \leqq np$　同様に，両辺を n で割って，

$$\frac{X}{n} - 1.96\frac{\sqrt{n\overline{p}(1-\overline{p})}}{n} \leqq p \quad \therefore\ \overline{p} - 1.96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \leqq p$$

$$P\left(\overline{p} - 1.96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \leqq p \leqq \overline{p} + 1.96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}}\right) = 0.95 \quad \cdots②$$

となる。

これから，母比率 p の 95% 信頼区間が，

$$\overline{p} - 1.96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \leqq p \leqq \overline{p} + 1.96\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \quad \cdots\cdots(*3)$$

(Ⅱ) 母比率 p の 99% 信頼区間について，

$P(-2.58 \leqq Z \leqq 2.58) = 0.99$ …③ より，(Ⅰ) とまったく同様の変形を行えば，p の 99% 信頼区間が次のように導けることも大丈夫だね。

$$\overline{p} - 2.58\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \leqq p \leqq \overline{p} + 2.58\sqrt{\frac{\overline{p}(1-\overline{p})}{n}} \quad \cdots\cdots(*4)$$

以上で，母平均 m や母比率 p の区間推定の解説は終了です。後は絶対暗記問題で具体的に計算して練習しておこう。問題を解くことにより，複雑な公式にも慣れることができるんだからね。頑張ってくれ！

母平均の区間推定

絶対暗記問題 60	難易度 ★★	CHECK1	CHECK2	CHECK3

ある地域の高校 **2** 年生 **10** 万人に数学のテストを行った。この採点結果を母集団として，これから **400** 人の成績を無作為に標本抽出した結果，標本平均が **65** 点，標本標準偏差が **10** 点であった。このとき，母平均 m の(ⅰ) **95**%信頼区間と(ⅱ) **99**%信頼区間を求めよ。

ヒント！ 標本の大きさ $n=400$ は十分大きいと考えられるので，母標準偏差 σ の代わりに標本標準偏差 S を用いて，母平均 m の(ⅰ) 95% 信頼区間の公式：$\overline{X}-1.96\dfrac{S}{\sqrt{n}} \leqq m \leqq \overline{X}+1.96\dfrac{S}{\sqrt{n}}$ と，(ⅱ) 99% 信頼区間の公式：$\overline{X}-2.58\dfrac{S}{\sqrt{n}} \leqq m \leqq \overline{X}+2.58\dfrac{S}{\sqrt{n}}$ を用いればいいんだね。

解答＆解説

標本の大きさ $n=400$，標本平均 $\overline{X}=65$，標本標準偏差 $S=10$ より，

(ⅰ) 母平均 m の 95% 信頼区間は，

$$\underbrace{65-1.96\frac{10}{\sqrt{400}}}_{65-1.96\times\underbrace{\frac{10}{20}}_{0.98}=64.02} \leqq m \leqq \underbrace{65+1.96\frac{10}{\sqrt{400}}}_{65+0.98=65.98}$$

公式：$\overline{X}-1.96\dfrac{S}{\sqrt{n}} \leqq m \leqq \overline{X}+1.96\dfrac{S}{\sqrt{n}}$

$\therefore\ 64.02 \leqq m \leqq 65.98$ ……………………………………………(答)

(ⅱ) 母平均 m の 99% 信頼区間は，

$$\underbrace{65-2.58\frac{10}{\sqrt{400}}}_{65-2.58\times\underbrace{\frac{10}{20}}_{1.29}=63.71} \leqq m \leqq \underbrace{65+2.58\frac{10}{\sqrt{400}}}_{65+1.29=66.29}$$

公式：$\overline{X}-2.58\dfrac{S}{\sqrt{n}} \leqq m \leqq \overline{X}+2.58\dfrac{S}{\sqrt{n}}$

$\therefore\ 63.71 \leqq m \leqq 66.29$ ……………………………………………(答)

母平均の区間推定の応用

絶対暗記問題 61	難易度 ★★	CHECK1	CHECK2	CHECK3

全国の **100** 万人の学生を対象に英語のテストを行った。この採点結果を母集団として，これから $n=256$ 人の成績を無作為に標本抽出した結果，標本平均 $\overline{X}$ が **74** 点，標本標準偏差 S が **12** 点であった。このとき，次の各問いに答えよ。(ただし，標本の大きさが変化しても，標本平均 $\overline{X}$ と標本標準偏差 S の値は変化しないものとする。)

(1) 母平均 m の (ⅰ) **95**％信頼区間と (ⅱ) **99**％信頼区間を求めよ。
(ただし，小数第 **3** 位を四捨五入せよ。)

(2) (1)の m の **95**％信頼区間の幅を半分にするための標本の大きさ n' の値を求めよ。

(3) (1)の m の **99**％信頼区間の幅を **2** 以下にするためには，少なくとも標本の大きさ n'' をいくつにすればよいか。

ヒント！ (1) 標本の大きさ $n=256$ は十分に大きいと考えられるので，標本標準偏差 S を用いて，m の (ⅰ) 95% 信頼区間と (ⅱ) 99% 信頼区間を求めればいいんだね。(2)では m の 95% 信頼区間の幅は $2\times1.96\dfrac{S}{\sqrt{n}}$ となること，(3)では 99% 信頼区間の幅が $2\times2.58\dfrac{S}{\sqrt{n}}$ となることを利用して解けばいいんだね。頑張ろう！

解答＆解説

(1) 標本の大きさ $n=256\,(=2^8)$，標本平均 $\overline{X}=74$，標本標準偏差 $S=12$ であり，n は十分に大きいので，

(ⅰ) 母平均 m の 95% 信頼区間は，

公式：
$\overline{X}-1.96\dfrac{S}{\sqrt{n}} \leqq m \leqq \overline{X}+1.96\dfrac{S}{\sqrt{n}}$

$$74-\underbrace{1.96\cdot\frac{12}{\sqrt{256}}}\leqq m \leqq 74+\underbrace{1.96\cdot\frac{12}{\sqrt{256}}}$$

$$\boxed{1.96\times\frac{12}{2^4}=1.96\times\frac{12}{16}=1.96\times\frac{3}{4}=1.47}$$

これから m の 95% 信頼区間の幅は，
$\cancel{\overline{X}}+1.96\dfrac{S}{\sqrt{n}}-\left(\cancel{\overline{X}}-1.96\dfrac{S}{\sqrt{n}}\right)$
$=2\times1.96\dfrac{S}{\sqrt{n}}$ となる。

$$74-1.47\leqq m\leqq 74+1.47$$

$\therefore\ 72.53\leqq m\leqq 75.47$ である。……………………………………(答)

(ⅱ) 母平均 m の 99% 信頼区間は，

$$74-\underbrace{2.58\cdot\frac{12}{\sqrt{256}}}_{\boxed{2.58\times\frac{3}{4}=1.935}}\leqq m\leqq 74+\underbrace{2.58\cdot\frac{12}{\sqrt{256}}}_{\boxed{1.935}}$$

公式：
$\overline{X}-2.58\frac{S}{\sqrt{n}}\leqq m\leqq\overline{X}+2.58\frac{S}{\sqrt{n}}$

これから m の 99% 信頼区間の幅は，
$\cancel{\overline{X}}+2.58\frac{S}{\sqrt{n}}-\left(\cancel{\overline{X}}-2.58\frac{S}{\sqrt{n}}\right)$
$=2\times2.58\frac{S}{\sqrt{n}}$ となる。

$$74-1.935\leqq m\leqq 74+1.935$$

$\therefore\ 72.07\leqq m\leqq 75.94$ である。………(答)

(2) (1)の(ⅰ)より，m の 95% 信頼区間の幅は，

$2\times1.96\frac{S}{\sqrt{n}}=2\times1.96\cdot\frac{12}{\sqrt{256}}$ ……① である。

この幅を半分にする標本の大きさを n' とおくと，①より，

$\cancel{2\times1.96}\times\frac{\cancel{12}}{\sqrt{n'}}=\frac{1}{2}\times\cancel{2\times1.96}\cdot\frac{\cancel{12}}{\sqrt{256}}$ となる。この逆数をとって，

$\sqrt{n'}=2\cdot\sqrt{256}$　これを 2 乗して，$n'=4\times256=2^2\times2^8=2^{10}=1024$

$\therefore\ n'=1024$ である。 ……………………………………………………………(答)

(3) (1)の(ⅱ)より，$n=256$ のときの m の 99% 信頼区間の幅は，

$2\times2.58\frac{S}{\sqrt{n}}=2\times2.58\cdot\frac{12}{\sqrt{256}}$ ……② である。

②の幅を 2 以下にするための標本の大きさを n'' とおくと，②より，

$\cancel{2}\times2.58\times\frac{12}{\sqrt{n''}}\leqq\cancel{2}$　　$\underbrace{12\times2.58}_{\boxed{30.96}}\leqq\sqrt{n''}$

この両辺を 2 乗すると，

$n''\geqq30.96^2=958.5216$ である。

$\therefore\ n''\geqq959$ より，m の 99% 信頼区間の幅を 2 以下にするためには，
少なくとも標本の大きさ n'' は 959 でなければならない。

…………(答)

母比率の区間推定

絶対暗記問題 62　難易度 ★★　CHECK1　CHECK2　CHECK3

ある地域の有権者から **1600** 人を無作為に抽出して，A 政党の支持者数を調べたところ **320** 人であった。この地域の A 政党の支持率 p を信頼度 **99**% で区間推定せよ。

ヒント！ これは，A 政党の支持率を母比率 p として，99% 信頼区間を求める問題なので，公式：$\overline{p}-2.58\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}} \leqq p \leqq \overline{p}+2.58\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}$ を用いればいいね。

解答＆解説

この地域での A 政党の支持率を母比率 p とおいて，この 99% 信頼区間を求める。

標本の大きさ $n=1600$，標本平均 $\overline{p}=\dfrac{320}{1600}=\dfrac{1}{5}=0.2$

公式：$\overline{p}-2.58\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}} \leqq p \leqq \overline{p}+2.58\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}$ を用いると，

$$\underbrace{0.2-2.58\sqrt{\frac{0.2\times0.8}{1600}}}_{} \leqq p \leqq \underbrace{0.2+2.58\sqrt{\frac{0.2\times0.8}{1600}}}_{}$$

$0.2-2.58\times\sqrt{\dfrac{0.16}{1600}}=0.2-0.0258=0.1742$

$\sqrt{\dfrac{1}{10000}}=\dfrac{1}{100}$

$0.2+0.0258=0.2258$

∴この支持率 p の 99% 信頼区間は，

$0.1742 \leqq p \leqq 0.2258$ である。 ……………………………………(答)

頻出問題にトライ・16　難易度 ★★★　CHECK1　CHECK2　CHECK3

ある地域の有権者から無作為に n 人を抽出して，A 政党の支持率を標本調査したところ支持率は **25**% であった。この地域の有権者全体の A 政党への支持率の **95**% 信頼区間の幅が **5**% 以下となるようにするためには，少なくとも何人を標本抽出すればよいか。

解答は P194

講義4 ● 確率分布と統計的推測　公式エッセンス

1. 期待値 $E(X)=m$，分散 $V(X)=\sigma^2$，標準偏差 $D(X)=\sigma$

(1) $E(X)=\sum_{k=1}^{n} x_k p_k$　(2) $V(X)=\sum_{k=1}^{n}(x_k-m)^2 p_k=E(X^2)-\{E(X)\}^2$

(3) $D(X)=\sqrt{V(X)}$

2. 新たな確率変数 $Y=aX+b$ の期待値，分散，標準偏差

(1) $E(Y)=aE(X)+b$　(2) $V(Y)=a^2V(X)$　(3) $D(Y)=|a|D(X)$

3. $E(X+Y)=E(X)+E(Y)$

4. 独立な確率変数 X と Y の積の期待値と和の分散

(1) $E(XY)=E(X)E(Y)$　(2) $V(X+Y)=V(X)+V(Y)$

5. 二項分布の期待値，分散，標準偏差

(1) $E(X)=np$　(2) $V(X)=npq$　(3) $D(X)=\sqrt{npq}$　$(q=1-p)$

6. 確率密度 $f(x)$ に従う連続型確率変数 X の期待値，分散

(1) $E(X)=\int_{-\infty}^{\infty} xf(x)\,dx$　(2) $V(X)=\int_{-\infty}^{\infty}(x-m)^2 f(x)\,dx$

$$=E(X^2)-\{E(X)\}^2$$

7. 正規分布 $N(m,\ \sigma^2)$ の確率密度 $f_N(x)$

$$f_N(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}\qquad (m=E(X),\ \sigma^2=V(X))$$

8. 標本平均 $\overline{X}$ の期待値 $m(\overline{X})$，分散 $\sigma^2(\overline{X})$，標準偏差 $\sigma(\overline{X})$

(1) $m(\overline{X})=m$　(2) $\sigma^2(\overline{X})=\dfrac{\sigma^2}{n}$　(3) $\sigma(\overline{X})=\dfrac{\sigma}{\sqrt{n}}$　$\left(\begin{array}{l} m：母平均 \\ \sigma^2：母分散 \end{array}\right)$

9. 母平均 m の（ⅰ）95% 信頼区間，（ⅱ）99% 信頼区間

（ⅰ）$\overline{X}-1.96\dfrac{S}{\sqrt{n}}\leqq m\leqq \overline{X}+1.96\dfrac{S}{\sqrt{n}}$

（ⅱ）$\overline{X}-2.58\dfrac{S}{\sqrt{n}}\leqq m\leqq \overline{X}+2.58\dfrac{S}{\sqrt{n}}$

10. 母比率 p の（ⅰ）95% 信頼区間，（ⅱ）99% 信頼区間

（ⅰ）$\overline{p}-1.96\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}\leqq p\leqq \overline{p}+1.96\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}$

（ⅱ）$\overline{p}-2.58\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}\leqq p\leqq \overline{p}+2.58\sqrt{\dfrac{\overline{p}(1-\overline{p})}{n}}$

$\left(\begin{array}{l} \overline{p}：標本比率 \\ n：標本数 \end{array}\right)$