統計的な推測【母平均の推定】　p.96～98

1 学習内容の説明 ⇒ 2 問題演習 ⇒ 3 振り返り（確認テスト・相互採点・リフレクションの記入）

【態度目標】しゃべる、質問する、説明する、動く、協力する、貢献する

【内容目標】標本平均から母平均がどのような範囲にあるか推定しよう

母集団が非常に大きいとき，母平均の真の値を求めるにはかなりの時間と労力を要する。

ここでは，標本平均を用いて，母平均がどのような範囲にあるか推定する方法を考えよう。

【参考】母集団の特性値を推測する場合、一般に「点推定」と「区間推定」の２種類のものが用いられる。

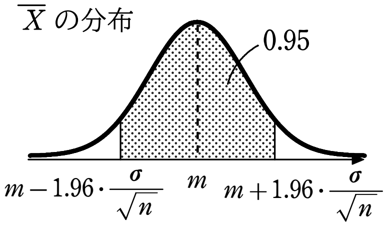
「点推定」
標本の値から一定の計算によって求められる値を、母集団の未知の特性値に対する推定量とする方法。不偏推定量、ベイズ推定量、最尤（さいゆう）推定量、一致推定量などがある。

「区間推定」
標本の値から計算によって求めた区間、すなわち、その中に母集団の特性値が含まれると期待できる区間を求める方法。利点としては、その区間がどの程度の正確さで推定しているかを明確にしていること。教科書で扱っているのはこの話。

□母平均の推定

92ページ

母平均 m，母標準偏差 σ をもつ母集団から抽出された大きさ n の無作為標本の標本平均 $\overline{X}$ 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。



すなわち，確率変数 $Z=\dfrac{\overline{X}-m}{\dfrac{\sigma}{\sqrt{n}}}$ は

近似的に標準正規分布 $N(0,\ 1)$ に従う。

巻末の正規分布表によると

$$P(|Z|\leqq 1.96)=0.95$$

例として、95%になるところを考える

であるから

$$P\left(|\overline{X}-m|\leqq 1.96\cdot\frac{\sigma}{\sqrt{n}}\right)=0.95$$

$\left|\dfrac{\overline{X}-m}{\dfrac{\sigma}{\sqrt{n}}}\right|\leqq 1.96$ の両辺に $\dfrac{\sigma}{\sqrt{n}}\ (>0)$ をかける（不等式を解いて $-1.96\leqq\dfrac{\overline{X}-m}{\dfrac{\sigma}{\sqrt{n}}}\leqq 1.96$ を変形するイメージでも）

よって，次の等式が成り立つ。

$$P\left(\overline{X}-1.96\cdot\frac{\sigma}{\sqrt{n}}\leqq m\leqq\overline{X}+1.96\cdot\frac{\sigma}{\sqrt{n}}\right)=0.95$$

すなわち，不等式

$$\overline{X}-1.96\cdot\frac{\sigma}{\sqrt{n}}\leqq m\leqq\overline{X}+1.96\cdot\frac{\sigma}{\sqrt{n}}\quad\cdots\cdots①$$

標本の平均を用いて、母集団の平均を推定する！

の成り立つ確率は 0.95 である。

① で示される範囲を，母平均 m に対する **信頼度** 95 %の **信頼区間** といい，次のように表す。

$$\left[\overline{X}-1.96\cdot\frac{\sigma}{\sqrt{n}},\ \overline{X}+1.96\cdot\frac{\sigma}{\sqrt{n}}\right]$$

閉区間[〇, △] ← $\{x \mid 〇\leqq x\leqq △\}$
開区間(〇, △) ← $\{x \mid 〇< x< △\}$

統計的な推測【母平均の推定】　p.96～98

まとめると，次のようになる。

母平均の推定

母標準偏差を σ とする。標本の大きさ n が大きいとき，
母平均 m に対する信頼度 95 %の信頼区間は

$$\left[\overline{X}-1.96\cdot\frac{\sigma}{\sqrt{n}},\ \overline{X}+1.96\cdot\frac{\sigma}{\sqrt{n}}\right]$$

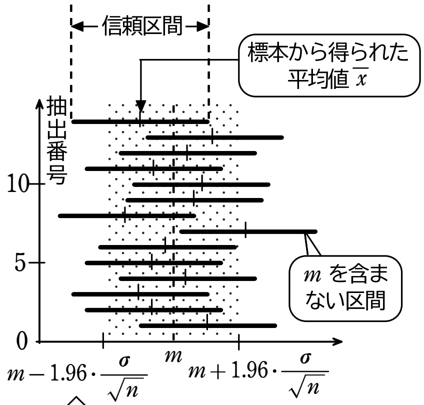
＜注意＞　「母平均 m に対して信頼度 95 %の**信頼区間を求めること**」を，
「母平均 m を信頼度 95 %で **推定 する**」ということがある。
また，必要に応じて，
信頼度 99 %や 90 %の信頼区間を求めることがある。

信頼度 99 %の信頼区間を求める場合は1.96 が 2.58 となる

母平均 m に対する信頼度 95 %の信頼区間の意味を考えよう。

$\overline{X}$ は確率変数であるから，標本から実際に得られる平均値 $\overline{x}$ は，抽出される標本によって異なる。

しかし，大きさ n の無作為標本を繰り返して抽出して，得られた平均値から上のような信頼区間を多数作ると，その中には m を含むものが 95 %あることが期待される。
これが，信頼度 95 %の信頼区間の意味である。



「無作為に標本抽出を 100 回繰り返し信頼区間を 100 個作ると、m を含む区間が 95 個ぐらいある」ということ

信頼区間

上で示した「母平均の推定」では，母標準偏差 σ を用いて，母平均を推定しているが，実際には σ の値もわからない場合が多い。標本の大きさ n が大きいときは，母標準偏差 σ の代わりに標本の標準偏差 S を用いても差し支えないことが知られている。（補足 あり）

標本標準偏差 S を用いる場合は「$t-$分布」(標本データが小さいときに使われる)を考える必要があるが、この場合も n が大きいときは σ を S で代用して得られる信頼区間はほぼ一致する

例題５）　大量に生産されたある製品の中から，400 個を無作為抽出して重さを量ったところ，平均値 98.8 g，標準偏差 2.0 g であった。この製品の平均重量 m g に対して，信頼度 95 %の信頼区間を求めよ。

① 母標準偏差が不明のときは、代わりに標本の標準偏差 S を用いる。
② $\overline{X}$ には確率変数 $\overline{X}$ の実現地である標本の平均値 $\overline{x}$ を用いる。

解答　標本の平均値は $\overline{x}=98.8$，標本の標準偏差は $S=2.0$，
標本の大きさは $n=400$ であるから

$$1.96\cdot\frac{S}{\sqrt{n}}=1.96\cdot\frac{2.0}{\sqrt{400}}\fallingdotseq 0.2$$

設問に合わせて小数第２位を四捨五入

よって，求める信頼区間は

$$[98.8-0.2,\ 98.8+0.2]$$

すなわち　$[98.6,\ 99.0]$　ただし，単位は g

＜注意＞　例題５ で求めた信頼区間は「98.6 g 以上 99.0 g 以下」であるともいう。

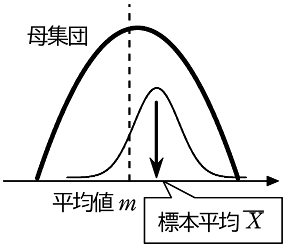
統計的な推測【母平均の推定】　p.96〜98

解説　「母平均の推定」の流れをイメージできるようにしよう！

母　集　団
平均値 m　：**未知**
分　散 σ^2 ：既知

平均値 m と分散 σ^2 をもつ様々な母集団（正規分布のものも含む）

大きさ n の標本を無作為に抽出し，標本平均 $\overline{X}$ とする。

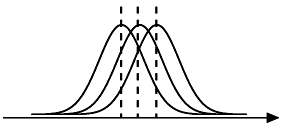


標本：大きさ n

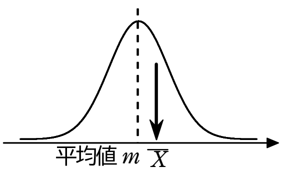
標本平均： $\overline{X}=\dfrac{X_1+X_2+\cdots\cdots+X_n}{n}$

標本の分散： S^2

大きさ n の標本の集合体を考える



標本平均 $\overline{X}$ の分布化



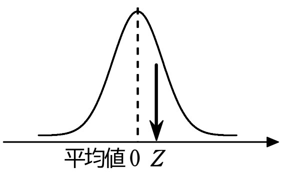
平均値： $E(\overline{X})=m$

分　散： $V(\overline{X})=\dfrac{\sigma^2}{n}$

標本平均 $\overline{X}$ の分布：正規分布 $N\left(m,\ \dfrac{\sigma^2}{n}\right)$ に従う

標準化 $Z=\dfrac{\overline{X}-m}{\dfrac{\sigma}{\sqrt{n}}}$

平均値： $E(Z)=0$
分　散： $V(Z)=1$



標準正規分布 $N(0,\ 1)$ に従う

標準正規分布の中で整理して推測する（母平均を推定）

母平均 m に対する信頼度（信頼係数）９５％の信頼区間

$$\overline{X}-1.96\cdot\frac{\sigma}{\sqrt{n}}\leqq m\leqq\overline{X}+1.96\cdot\frac{\sigma}{\sqrt{n}}$$

統計的な推測【母平均の推定】　p.96〜98

補足　$\overline{X}$ が正規分布 $N\left(m,\ \dfrac{\sigma^2}{n}\right)$ に従うとき，$Z=\dfrac{\overline{X}-m}{\dfrac{\sigma}{\sqrt{n}}}$ で標準化すると，

Z が標準正規分布 $N(0,\ 1)$ に従うか確認しよう。

標準化の式は　$Z=\dfrac{\sqrt{n}}{\sigma}\overline{X}-\dfrac{m\sqrt{n}}{\sigma}$ より

Z は $\overline{X}$ の１次式であるから，正規分布として対応する。

$$E(Z)=E\left(\frac{\overline{X}-m}{\dfrac{\sigma}{\sqrt{n}}}\right)=\frac{\sqrt{n}}{\sigma}E(\overline{X}-m)=\frac{\sqrt{n}}{\sigma}\{E(\overline{X})-m\}=0$$

$$V(Z)=V\left(\frac{\overline{X}-m}{\dfrac{\sigma}{\sqrt{n}}}\right)=\frac{n}{\sigma^2}V(\overline{X}-m)=\frac{n}{\sigma^2}V(\overline{X})=\frac{n}{\sigma^2}\cdot\frac{\sigma^2}{n}=1$$

Z は標準正規分布 $N(0,\ 1)$ に従うことが確認できる。

補足　一般に，「実際には σ の値もわからない場合が多い。標本の大きさ n が大きいときは，母標準偏差 σ の代わりに標本の標準偏差 S を用いても差し支えないことが知られている」とあるが，どのようなことか考えてみよう。

平均値 m，分散 σ^2 の母集団から，無作為に抽出した大きさ n の標本を $X_1,\ X_2,\ \cdots\cdots,\ X_n$ とし，標本の平均 $\overline{X}$，分散 S^2 とする。

標本分散 S^2 は　$S^2=\dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(X_k-\overline{X})^2$　である。

$$\begin{aligned}S^2&=\frac{1}{n}\sum_{k=1}^{n}(X_k-\overline{X})^2=\frac{1}{n}\sum_{k=1}^{n}\{(X_k-m)-(\overline{X}-m)\}^2\\&=\frac{1}{n}\sum_{k=1}^{n}(X_k-m)^2-\frac{2(\overline{X}-m)}{n}\sum_{k=1}^{n}(X_k-m)+\frac{1}{n}\sum_{k=1}^{n}(\overline{X}-m)^2\\&=\frac{1}{n}\sum_{k=1}^{n}(X_k-m)^2-\frac{2(\overline{X}-m)}{n}\left(\sum_{k=1}^{n}X_k-mn\right)+\frac{1}{n}\cdot n(\overline{X}-m)^2\\&=\frac{1}{n}\sum_{k=1}^{n}(X_k-m)^2-\frac{2(\overline{X}-m)}{n}(n\overline{X}-mn)+(\overline{X}-m)^2\\&=\frac{1}{n}\sum_{k=1}^{n}(X_k-m)^2-2(\overline{X}-m)^2+(\overline{X}-m)^2\\&=\frac{1}{n}\sum_{k=1}^{n}(X_k-m)^2-(\overline{X}-m)^2\end{aligned}$$

標本の大きさ n を大きくすると，$\dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(X_k-m)^2\longrightarrow\sigma^2$，$(\overline{X}-m)^2\longrightarrow 0$ だから

標本分散 S^2 は，母分散 σ^2 に近づく

したがって，標本の大きさが十分に大きい場合は，母標準偏差 σ の代わりに標本標準偏差 S を用いてもよい。

統計的な推測【母平均の推定】　p.96～98

参考　区間推定までの流れ

母集団（正規分布）
平均値 m （未知）
分　散 σ^2 （未知）

- 標本の大きさ n が大きい
  - 標本平均 $\overline{X}$
    標本分散 S^2
  - n が大きいので
    $\sigma^2 = S^2$ としてよい
  - $\overline{X}$ は，平均値 m，
    分散 $\dfrac{S^2}{n}$ の正規分布に
    従うので
    $Z = \dfrac{\overline{X}-m}{\dfrac{S}{\sqrt{n}}}$ として
    標準化すると，
    Z は標準正規分布
    $N(0,\ 1)$ に従う
  - 平均値 m の９５%の区間推定
    $\overline{X} - 1.96 \cdot \dfrac{S}{\sqrt{n}} \leqq m \leqq \overline{X} + 1.96 \cdot \dfrac{S}{\sqrt{n}}$
- 標本の大きさ n が小さい
  - 標本平均 $\overline{X}$
    標本分散 S^2
  - n が小さいので
    $\sigma^2 \fallingdotseq S^2$ より
    U^2 (不偏分散)を活用
  - $T = \dfrac{\overline{X}-m}{\dfrac{U}{\sqrt{n}}}$ と
    定義すると，
    T は自由度 $(n-1)$ の
    t 分布に従う
  - 平均値 m の９５%の区間推定
    自由度 $(n-1)$ の t 分布表から
    ９５%の数値 α を読み取る
    $\overline{X} - \alpha \cdot \dfrac{U}{\sqrt{n}} \leqq m \leqq \overline{X} + \alpha \cdot \dfrac{U}{\sqrt{n}}$

母集団（正規分布でない）
平均値 m （未知）
分　散 σ^2 （未知）

- 標本の大きさ n が大きい
  - 標本平均 $\overline{X}$
    標本分散 S^2
  - n が大きいので
    $\sigma^2 = S^2$ としてよい
  - $\overline{X}$ は，平均値 m，
    分散 $\dfrac{S^2}{n}$ の正規分布に
    従うので
    $Z = \dfrac{\overline{X}-m}{\dfrac{S}{\sqrt{n}}}$ として
    標準化すると，
    Z は標準正規分布
    $N(0,\ 1)$ に従う
  - 平均値 m の９５%の区間推定
    $\overline{X} - 1.96 \cdot \dfrac{S}{\sqrt{n}} \leqq m \leqq \overline{X} + 1.96 \cdot \dfrac{S}{\sqrt{n}}$
- 標本の大きさ n が小さい
  - 標本平均 $\overline{X}$
    標本分散 S^2
  - n が小さいので
    $\sigma^2 \fallingdotseq S^2$
  - n が小さいので
    中心極限定理も
    成り立たない
  - 平均値 m の区間推定は
    単純でなく，
    高校数学の範囲を超える

補足　$U^2 = \dfrac{n}{n-1} S^2$ と定義すると $E(U^2) = \sigma^2$ となる。

$U^2 = \dfrac{1}{n-1} \sum_{k=1}^{n} (X_k - m)^2$ で得られる分散を不偏分散という

$n \longrightarrow \infty$ とすると，$U^2 \longrightarrow \sigma^2$ より U^2 は一致性があり，
$E(U^2) = \sigma^2$ より U^2 は不偏性がある
U^2 は一致性と不偏性を有しており，
母分散の点推定や区間推定で用いられる。