統計的な推測【母比率の推定】 p.102～103

青チャートB: 例題 91

【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】前回までの話を生かして標本比率から母比率を推定しよう

□母比率の推定

信頼区間の考え方は，母比率の推定にも用いることができる。標本比率から母比率を推定する方法を調べよう。

例えば，大量生産されたある製品の中から，大きさ n の無作為標本を抽出したとき，それらに対して，X_1，X_2，……，X_n の値を特質の有無で 1 か 0 かを定めると，$T=X_1+X_2+\cdots\cdots+X_n$ は大きさ n の標本の中で特性 A をもつものの個数を表す確率変数である。不良品が T 個あったとすると，標本における製品の不良率は $R=\dfrac{T}{n}$ である。この場合，R は「不良品」という特性の標本比率である。不良品の母比率が p である母集団から，大きさ n の無作為標本を抽出するとき，96 ページの標本平均の分布で学んだように，n が大きいと，R は近似的に正規分布 $N\left(p,\ \dfrac{p(1-p)}{n}\right)$ に従う。

したがって，99 ページの母平均の推定の場合と同様に

$$P\left(p-1.96\sqrt{\frac{p(1-p)}{n}} \leqq R \leqq p+1.96\sqrt{\frac{p(1-p)}{n}}\right) \fallingdotseq 0.95$$

よって $P\left(R-1.96\sqrt{\dfrac{p(1-p)}{n}} \leqq p \leqq R+1.96\sqrt{\dfrac{p(1-p)}{n}}\right) \fallingdotseq 0.95$

となる。ここで，n が十分大きいとき，大数の法則により，R は p に近いとみなしてよいから，根号の中の p を R でおき換えると（説明は後述）

$$P\left(R-1.96\sqrt{\frac{R(1-R)}{n}} \leqq p \leqq R+1.96\sqrt{\frac{R(1-R)}{n}}\right) \fallingdotseq 0.95$$

標本比率を用いて、
母集団の比率を推定する！

別解 T は二項分布 $B(n,\ p)$ に従うから，n が大きいとき，

近似的に正規分布 $N(np,\ np(1-p))$ に従う。 86ページ

母平均の推定の場合と同様に， 母平均の推定と同様にして…

$$P\left(-1.96 \leqq \frac{T-np}{\sqrt{np(1-p)}} \leqq 1.96\right)=0.95$$

$Z=\dfrac{X-(\text{平均})}{(\text{標準偏差})}$

ここで，n が十分に大きいならば大数の法則により 標本比率 R は母比率 p に近いとみなしてよい。よって，根号の中の母比率 p の代わりに標本平均 R を用いてよいことが知られているから，次のように変形できる。

$$P\left(-1.96 \leqq \frac{\dfrac{T}{n}-p}{\sqrt{\dfrac{R(1-R)}{n}}} \leqq 1.96\right)=0.95 \quad \cdots\cdots ①$$

$$P\left(\frac{T}{n}-1.96\cdot\sqrt{\frac{R(1-R)}{n}}\leqq p\leqq\frac{T}{n}+1.96\cdot\sqrt{\frac{R(1-R)}{n}}\right)=0.95$$

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

$R=\frac{T}{n}$

すなわち，標本調査で得られる 1 つの標本比率 R を用いて，母比率は次の区間内にあると推定され

$R-1.96\cdot\sqrt{\frac{R(1-R)}{n}}\leqq p\leqq R+1.96\cdot\sqrt{\frac{R(1-R)}{n}}$ の成り立つ確率は 0.95 である。

標本比率を用いて、
母集団の比率を推定する！

したがって，次のことが成り立つ。

母比率の推定

標本の大きさ n が大きいとき，標本比率を R とすると，
母比率 p に対する信頼度 95 %の信頼区間は

$$\left[R-1.96\sqrt{\frac{R(1-R)}{n}},\ R+1.96\sqrt{\frac{R(1-R)}{n}}\right]$$

※ 信頼度 99 %では

$$R-2.58\sqrt{\frac{R(1-R)}{n}}\leqq p\leqq R+2.58\sqrt{\frac{R(1-R)}{n}}$$

例題１２）ある県の高校 3 年生から無作為に 400 人を選び，むし歯がある生徒を数えたところ，200 人であった。この県の高校 3 年生のむし歯の保有率 p を，信頼度 95 % で推定せよ。

解答 標本比率 R は　$R=\frac{200}{400}=0.5$

$n=400$ であるから

$$1.96\sqrt{\frac{R(1-R)}{n}}=1.96\sqrt{\frac{0.5\times0.5}{400}}\fallingdotseq0.049$$

よって　$0.5-0.049=0.451$，　$0.5+0.049=0.549$

ゆえに，むし歯の保有率 p に対する信頼度 95 % の信頼区間は

$[0.451,\ 0.549]$

$n=400$ は十分大きな数である
→　近似的に正規分布
$N\left(p,\ \frac{p(1-p)}{n}\right)$ に従う

分数や指数の活用も
考えた方がよいかも

$$\frac{196}{100}\sqrt{\frac{1}{400}\cdot\frac{1}{2}\cdot\frac{1}{2}}$$
$$=\frac{196}{100}\cdot\frac{1}{20}\cdot\frac{1}{2}$$
$$=(200-4)\cdot\frac{1}{4}\cdot10^{-3}$$
$$=(50-1)\cdot10^{-3}$$
$$=0.049$$

信頼度 99 %とすると　$2.58\sqrt{\frac{0.5\times0.5}{400}}=2.58\times0.025\fallingdotseq0.0645$

よって，求める信頼区間は　$[0.5-0.0645,\ 0.5+0.0645]$

すなわち　$[0.4355,\ 0.5645]$

統計的な推測【母比率の推定】 p.102〜103

青チャートB: 例題 91

例） ある世論調査で，有権者から無作為抽出した 400 人について A 政党の支持者を調べたら 144 人いた。A 政党の支持者の母比率 p に対して，信頼度 95 %の信頼区間を求めよ。

解答　標本比率 R は　$R=\dfrac{144}{400}=0.36$

$n=400$ は十分大きな数である
→ 近似的に正規分布
$N\left(p,\ \dfrac{p(1-p)}{n}\right)$ に従う

$n=400$ であるから

$$1.96\sqrt{\frac{R(1-R)}{n}}=1.96\sqrt{\frac{0.36\times0.64}{400}}$$
$$=1.96\times0.024$$
$$\fallingdotseq 0.047$$

分数や指数の活用も考えた方がよいかも

$$\frac{196}{100}\sqrt{\frac{1}{400}\cdot\frac{36}{100}\cdot\frac{64}{100}}$$
$$=\frac{196}{100}\sqrt{\frac{1}{100}\cdot\frac{9}{100}\cdot\frac{64}{100}}$$
$$=\frac{196}{100}\cdot\frac{1}{10}\cdot\frac{3}{10}\cdot\frac{8}{10}$$
$$=(200-4)\cdot24\cdot10^{-5}$$
$$=(4800-96)\cdot10^{-5}$$
$$=4704\cdot10^{-5}$$
$$=0.04704$$

よって，求める信頼区間は

$$[0.36-0.047,\ 0.36+0.047]$$

すなわち　$[0.313,\ 0.407]$

信頼度 99 %とすると　$2.58\sqrt{\dfrac{0.36\times0.64}{400}}=2.58\times0.024\fallingdotseq 0.062$

よって，求める信頼区間は　$[0.36-0.062,\ 0.36+0.062]$　すなわち　$[0.298,\ 0.422]$

解説 標本における比率 $\dfrac{T}{n}$ を R とすれば n が十分に大きいとき $R \fallingdotseq p$ と見なしてよいことについて

母標準偏差 $\sqrt{p(1-p)}$ を確率変数 $\sqrt{R(1-R)}$ でおき換えたが,

$\sqrt{R(1-R)}$ は標本標準偏差 $S=\sqrt{\dfrac{1}{n}\sum_{k=1}^{n}(X_k-\overline{X})^2}$ と等しく，したがって，このおき換えは 101 ページで述べた「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 $S=\sqrt{\dfrac{1}{n}\sum_{k=1}^{n}(X_k-\overline{X})^2}$ の値を用いても差し支えないことが知られている。」の特別な場合になっている。このことは，次のようにして示される。

証明 X_k はとる値が 0 または 1 であるから，${X_k}^2=X_k$ となる。また，$R=\overline{X}$ であるから

$$\begin{aligned}
S^2&=\frac{1}{n}\sum_{k=1}^{n}(X_k-\overline{X})^2\\
&=\frac{1}{n}\sum_{k=1}^{n}({X_k}^2-2X_k\overline{X}+\overline{X}^2)\\
&=\frac{1}{n}\sum_{k=1}^{n}{X_k}^2-2\cdot\frac{1}{n}\sum_{k=1}^{n}X_k\overline{X}+\overline{X}^2\\
&=\frac{1}{n}\sum_{k=1}^{n}X_k-2\overline{X}\cdot\overline{X}+\overline{X}^2\\
&=\overline{X}-2\overline{X}^2+\overline{X}^2=\overline{X}-\overline{X}^2\\
&=R-R^2=R(1-R) \qquad \therefore \quad S=\sqrt{R(1-R)}
\end{aligned}$$

証明別解 $\left|\dfrac{T-np}{\sqrt{np(1-p)}}\right|\leqq 1.96$ であり，両辺がともに正であることから２乗すると

$\left(\dfrac{T-np}{\sqrt{np(1-p)}}\right)^2\leqq 1.96^2$ 　であるから　$(T-np)^2\leqq 1.96^2\cdot np(1-p)$

$T=n\cdot R$ より　$n\cdot(R-p)^2\leqq 1.96^2\cdot p(1-p)$

$\therefore\ (n+1.96^2)p^2-(2nR+1.96^2)p+nR^2\leqq 0$

$=0$ とみて解の公式を用いると

$$p=\frac{2nR+1.96^2\pm\sqrt{(2nR+1.96^2)^2-4(n+1.96^2)\cdot nR^2}}{2(n+1.96^2)}$$

$$p=\frac{2nR+1.96^2\pm\sqrt{4nR\cdot 1.96^2+1.96^4-4\cdot 1.96^2\cdot nR^2}}{2(n+1.96^2)}$$

$$p=\frac{2nR+1.96^2\pm 1.96\sqrt{4nR+1.96^2-4nR^2}}{2(n+1.96^2)}$$

$$p=\frac{2R+\dfrac{1.96^2}{n}\pm 1.96\sqrt{\dfrac{4R}{n}+\dfrac{1.96^2}{n^2}-\dfrac{4R}{n}}}{2\left(1+\dfrac{1.96^2}{n}\right)}$$

n を次第に大きくすると，n の次数を考慮すると $\dfrac{1.96^2}{n^2}$ が他より先に 0 になることから

$p\fallingdotseq\dfrac{2R\pm 1.96\sqrt{\dfrac{4R}{n}-\dfrac{4R^2}{n}}}{2}$ 　より　$p=R\pm 1.96\sqrt{\dfrac{R(1-R)}{n}}$

したがって　$R-1.96\sqrt{\dfrac{R(1-R)}{n}}\leqq p\leqq R+1.96\sqrt{\dfrac{R(1-R)}{n}}$