母集団比率の区間推定.

母集団において.ある条件Cをみたす比率を p とする.

また.大きさ n の標本において. Xを.Cをみたす標本の数とし.

$\overline{X}$をCをみたす比率とする. ここで $\overline{X}=\frac{1}{n}X$ である.

Xは二項分布 $B(n,p)$ に従うので、nが十分大きければ.($np \geqq 5$)

Xはほぼ $N(np, np(1-p))$ に従い. $\overline{X}$ は $N(p, \frac{p(1-p)}{n})$ に従う.

例. 信大生100人がW杯初戦をみたか調べたところ.37人見ていた.

この中継の視聴率 p の95%信頼区間を求めよ.

答. $n \cdot p = 100 \times 0.37 = 37 \geqq 5$ より.視聴率 $\overline{X}$ はほぼ $N(p, \frac{p(1-p)}{n})$ に従う.

∴ $P(\overline{X}-\delta \leqq p \leqq \overline{X}+\delta)=0.95$ となる δを求めればよい.変形すると.

$P(p-\delta \leqq \overline{X} \leqq p+\delta)=0.95$ となる.

$Z=\dfrac{\overline{X}-p}{\sqrt{\frac{p(1-p)}{n}}}$ とおけば Z は $N(0,1)$に従い.左辺は

$$P\left(\frac{-\delta}{\sqrt{\frac{p(1-p)}{n}}} \leqq Z \leqq \frac{\delta}{\sqrt{\frac{p(1-p)}{n}}}\right) = P\left(\frac{-\delta \cdot \sqrt{n}}{\sqrt{p(1-p)}} \leqq Z \leqq \frac{\delta\sqrt{n}}{\sqrt{p(1-p)}}\right)$$

$$=2 \cdot P\left(0 \leqq Z \leqq \frac{\delta\sqrt{n}}{\sqrt{p(1-p)}}\right)=0.95 \text{ となる.}$$

$P\left(0 \leqq Z \leqq \dfrac{\delta\sqrt{n}}{\sqrt{p(1-p)}}\right)=0.475$ より. $\dfrac{\delta\sqrt{n}}{\sqrt{p(1-p)}}=1.96$ になる.

ここで、$p=0.37$, $n=100$ として計算すると.

$\delta=\dfrac{\sqrt{0.37 \cdot (1-0.37)}}{\sqrt{100}} \cdot 1.96 \fallingdotseq 0.095$ となり

95%信頼区間は. $0.275 < p < 0.465$ である

<u>問題</u>. さきほどの例で、1万人中 3600人がみていたときの

pの99%信頼区間を求めよ.

<u>答</u>. $n \cdot p = 10000 \times 0.36 = 3600$ より 視聴率$\bar{X}$はほぼ$N(p, \frac{p(1-p)}{n})$に従う.

∴ $P(\bar{X}-\delta \leqq p \leqq \bar{X}+\delta) = 0.99$ となるδを求めればよい. 計算していくと.

$P(p-\delta \leqq \bar{X} \leqq p+\delta) = 0.99$. とできる

ここで $Z = \dfrac{\bar{X}-p}{\sqrt{\frac{p(1-p)}{n}}}$ とおけば、Zは$N(0,1)$に従い.

$$P\left(-\frac{\delta}{\sqrt{\frac{p(1-p)}{n}}} \leqq Z \leqq \frac{\delta}{\sqrt{\frac{p(1-p)}{n}}}\right) = 0.99$$

$P\left(0 \leqq Z \leqq \dfrac{\delta\sqrt{n}}{\sqrt{p(1-p)}}\right) = 0.495$ より $\dfrac{\delta\sqrt{n}}{\sqrt{p(1-p)}} = 2.5758$ となる.

ここで $p=0.36$, $n=10000$ を代入すれば、$\delta \fallingdotseq 0.012$ となり.

99%信頼区間は $0.348 < p < 0.372$ となる

<u>問題</u>. p104, 問1を解け.

<u>答</u>. 上の計算と同様にすれば. $\dfrac{\delta\sqrt{n}}{\sqrt{p(1-p)}} = 2.5758$ が求まる.

ここで $p=0.022$ $n=500$ を代入すれば. $\delta \fallingdotseq 0.017$. となり

99%信頼区間は $0.005 < p < 0.039$ となる.

統計的仮説検定

未知母数を具体的な数値と比較して.差異があるか判定したい.

例. 250 g 入りの缶詰 10本の内容量を調べたら、平均値は 248.2 g であった

メーカーの主張は平均 250 g , 標準偏差 3.2 g の正規分布である.

この表示は正当であるといえるか.

(1) [メーカーの主張] 平均値 $\mu = 250$

(2) [消費者の主張] 平均値 $\mu < 250$

(3) とりあえずメーカーの主張を認めて.内容量は $N(250, (3.2)^2)$ に従うとする.

すると.大きさ10の~~標本平均~~ $\overline{X}$ は $N(250, \frac{(3.2)^2}{10})$ に従う.

ここで $Z = \dfrac{\overline{X}-250}{\frac{3.2}{\sqrt{10}}}$ とおけば、Z は $N(0,1)$ に従う.

(4). 1より小さい値.例えば 0.05 をとり.

$P(\overline{X} < \xi) = 0.05$ となる ξ を求めると.

$P(Z < \frac{\sqrt{10}}{3.2}(\xi - 250)) = 0.05$

$P(\frac{\sqrt{10}}{3.2}(\xi - 250) < Z < 0) = 0.45$ より

$-\frac{\sqrt{10}}{3.2}(\xi - 250) = 1.6449$ となり $\xi \fallingdotseq 248.3$ となる.

(5) つまり. $\overline{X} = 248.2$ は確率的におこりにくい (5%) $\overline{X} < 248.3$ に入っている.

(6) 消費者の主張が正しく思える

問題 (4) で 0.05 を 0.01 とすればどうなるか.

答. $P(\frac{\sqrt{10}}{3.2}(\xi - 250) < Z < 0) = 0.49$ より. $-\frac{\sqrt{10}}{3.2}(\xi - 250) = 2.3263$ となり.

$\xi \fallingdotseq 247.6$ となる. $\overline{X} = 248.2 > \xi = 247.6$ なので、メーカーの主張が正しいと思える.