第５問　確率分布と統計的な推測

(1)　サイコロを１回振ったときに出る目は１，２，３，４，５，６のいずれかである．

確率変数 X は次のように定められている．

『サイコロを１回振って

１または６の目が出たときは $X=0$,

２または５の目が出たときは $X=1$,

３または４の目が出たときは $X=2$.』

不良品のサイコロＤについて，$P(X=0)=p$，$P(X=1)=q$，$P(X=2)=r$ とすると

$$p+q+r=\boxed{1} \qquad \cdots①$$

である．

X の期待値(平均)が $E(X)=\dfrac{7}{5}$ であるから

$$0\cdot p+1\cdot q+2\cdot r=\frac{7}{5}$$

つまり

$$q+\boxed{2}\,r=\frac{7}{5} \qquad \cdots②$$

が成り立つ．さらに，X の分散が $V(X)=\dfrac{16}{25}$ であるから

$$0^2\cdot p+1^2\cdot q+2^2\cdot r-\left(\frac{7}{5}\right)^2=\frac{16}{25}$$

$$q+4r=\frac{13}{5} \qquad \cdots③$$

が成り立つ．①，②，③より

$$p=\frac{\boxed{1}}{\boxed{5}},\quad q=\frac{\boxed{1}}{\boxed{5}},\quad r=\frac{\boxed{3}}{\boxed{5}}$$

である．よって，サイコロＤはＣグループに属することがわかる．

$P(X=0)=\dfrac{1}{5}$，$P(X=1)=\dfrac{1}{5}$，$P(X=2)=\dfrac{3}{5}$ であるサイコロを100回振って毎回 X の値を調べ，$X=2$ であった回数を Y とすると，確率変数 Y は二項分布 $B\left(100,\ \dfrac{3}{5}\right)$ に従うから，Y の期待値(平均)は

$$E(Y)=100\cdot\frac{3}{5}=\boxed{60}$$

であり，Y の分散は

$$V(Y)=100\cdot\frac{3}{5}\cdot\left(1-\frac{3}{5}\right)=\boxed{24}$$

である．

また，このサイコロを100回振って毎回 X の値を調べ，各回の X の値を X_k ($k=1,\ 2,\ 3,\ \cdots,\ 100$) とすると

全事象の起こる確率は１である．

期待値(平均)

確率変数 X のとり得る値を

$$x_1,\ x_2,\ \cdots,\ x_n$$

とし，X がこれらの値をとる確率をそれぞれ

$$p_1,\ p_2,\ \cdots,\ p_n$$

とすると，X の期待値(平均) $E(X)$ は

$$E(X)=\sum_{k=1}^{n}x_kp_k.$$

分散

確率変数 X のとり得る値を

$$x_1,\ x_2,\ \cdots,\ x_n$$

とし，X がこれらの値をとる確率をそれぞれ

$$p_1,\ p_2,\ \cdots,\ p_n$$

とすると，X の分散 $V(X)$ は，$E(X)=m$ として

$$V(X)=\sum_{k=1}^{n}(x_k-m)^2p_k \qquad \cdots(\text{i})$$

または

$$V(X)=E(X^2)-\{E(X)\}^2. \qquad \cdots(\text{ii})$$

ここでは(ii)を用いた．

(i)を用いると

$$\left(-\frac{7}{5}\right)^2p+\left(-\frac{2}{5}\right)^2q+\left(\frac{3}{5}\right)^2r=\frac{16}{25}$$

となる．

$$P(X=2)=\frac{3}{5}.$$

二項分布

確率変数 X が二項分布 $B(n,\ p)$ に従うとき，$q=1-p$ とすると，X の期待値(平均) $E(X)$，分散 $V(X)$，標準偏差 $\sigma(X)$ は

$$E(X)=np,$$
$$V(X)=npq,$$
$$\sigma(X)=\sqrt{npq}$$

である．

$$S = X_1 + X_2 + X_3 + \cdots + X_{100} = \sum_{k=1}^{100} X_k$$

である．

ここで

$$E(X_k) = \frac{7}{5}, \quad V(X_k) = \frac{16}{25} \quad (k = 1,\ 2,\ 3,\ \cdots,\ 100)$$

であるから，確率変数 S の期待値(平均)は

$$E(S) = E\left(\sum_{k=1}^{100} X_k\right) = \sum_{k=1}^{100} E(X_k)$$
$$= \sum_{k=1}^{100} \frac{7}{5}$$
$$= 100 \times \frac{7}{5}$$
$$= \boxed{140}$$

である．

また，確率変数 X_k $(k = 1,\ 2,\ 3,\ \cdots,\ 100)$ は互いに独立であるから，S の分散は

$$V(S) = V\left(\sum_{k=1}^{100} X_k\right) = \sum_{k=1}^{100} V(X_k)$$
$$= \sum_{k=1}^{100} \frac{16}{25}$$
$$= 100 \times \frac{16}{25}$$
$$= \boxed{64}$$

である．

期待値(平均)の性質

二つの確率変数 X，Y の和について

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

が成り立つ．

これを繰り返し用いることにより，n 個の確率変数 X_1，X_2，X_3，$\cdots$，X_n の和について

$$E\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} E(X_k)$$

が成り立つ．

分散の性質

二つの確率変数 X，Y が互いに独立ならば

$$V(X+Y) = V(X) + V(Y)$$

が成り立つ．

これを繰り返し用いることにより，n 個の確率変数 X_1，X_2，X_3，$\cdots$，X_n が互いに独立ならば

$$V\left(\sum_{k=1}^{n} X_k\right) = \sum_{k=1}^{n} V(X_k)$$

が成り立つ．

(2) 150 個の不良品に対する A グループに属するものの標本比率 R は $\dfrac{60}{150} = \dfrac{2}{5}$ であった．

150 個の標本のうち，60 個が A グループに属するものであった．

不良品全体に対する A グループに属するものの母比率を p_{A} とすると，R は近似的に正規分布 $N\left(p_{\mathrm{A}},\ \dfrac{p_{\mathrm{A}}(1-p_{\mathrm{A}})}{150}\right)$ に従う．したがって，母比率 p_{A} に対する信頼度 95% の信頼区間は

$$\left[R - 1.96\sqrt{\frac{R(1-R)}{150}},\ R + 1.96\sqrt{\frac{R(1-R)}{150}}\right]$$

である．$R = \dfrac{2}{5}$ のとき

$$1.96\sqrt{\frac{R(1-R)}{150}} = 1.96\sqrt{\frac{2}{5} \cdot \frac{3}{5} \cdot \frac{1}{150}}$$
$$= \frac{1.96}{25}$$
$$= 0.0784 \qquad \cdots④$$

であるから，求める信頼区間は

$$[0.4 - 0.0784,\ 0.4 + 0.0784]$$

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

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

正規分布表より，$\dfrac{0.95}{2}$ すなわち 0.475 となる z_0 の値は 1.96 である．

すなわち

$$[0.3216,\ 0.4784]$$

である．小数第3位を四捨五入すると

$$[\boxed{0}.\boxed{32},\ \boxed{0}.\boxed{48}]$$

である．

(3)　600個の不良品に対するAグループに属するものの標本比率 R は $\dfrac{240}{600}=\dfrac{2}{5}$ であった．

⬅ 600個の標本のうち，240個がAグループに属するものであった．

(2)と同様にして，母比率 p_A に対する信頼度95%の信頼区間は

$$\left[R-1.96\sqrt{\frac{R(1-R)}{600}},\ R+1.96\sqrt{\frac{R(1-R)}{600}}\right]$$

である．$R=\dfrac{2}{5}$ のとき

$$1.96\sqrt{\frac{R(1-R)}{600}}=1.96\sqrt{\frac{2}{5}\cdot\frac{3}{5}\cdot\frac{1}{600}}$$

$$=\frac{0.0784}{2}$$

$$=0.0392$$

⬅ ④を用いた．

より，母比率 p_A に対する信頼度95%の信頼区間は

$$[0.3608,\ 0.4392]$$

である．

よって，p_A が0.45を超える可能性は2.5%以下であるから，$\boxed{ネ}$ に当てはまるものは $\boxed{④}$ である．