

### 8.3 大標本による母比率の検定

母集団に属す各々の事物について、ある性質  $A$  を持つか持たないかのどちらかであるとする。このような母集団を2項母集団といった。標本調査の結果から、母集団全体のなかでこの性質  $A$  を持つ事物の占める割合、母比率について仮説検定をする。2項母集団に属す各々の事物について、性質  $A$  を持つときその属性は1であり、性質  $A$  を持たないとき属性は0であるとする。このとき、母比率は母平均と同じであり、標本比率は標本平均と同じである。

母比率が  $\rho$  である無限2項母集団から無作為抽出された標本の大きさ  $n$  が充分大きいとする。定理5.4.2により、標本比率を表す確率変数  $R$  は正規分布  $N\left(\rho, \frac{\rho(1-\rho)}{n}\right)$  に従う。標本調査では数学的根拠において確率が厳密な値か近似値かに拘ることにあまり意味がないので、 $R$  は正規分布  $N\left(\rho, \frac{\rho(1-\rho)}{n}\right)$  に従うとして確率を考える。 $R$  を標準化した確率変数

$$Z = \frac{R - \rho}{\sqrt{\frac{\rho(1-\rho)}{n}}} = \sqrt{\frac{n}{\rho(1-\rho)}}(R - \rho)$$

は標準正規分布  $N(0,1)$  に従う。標準正規分布を用いる仮説検定を  $Z$  検定という。

与えられた実数  $p$  に対して、仮説  $\rho \neq p$  及び仮説  $\rho > p$  及び仮説  $\rho < p$  を検討する。標本比率  $R$  を標準化した確率変数  $Z = \sqrt{\frac{n}{\rho(1-\rho)}}(R - \rho)$  は標準正規分布  $N(0,1)$  に従う。母比率  $\rho$  は分からないので、 $Z = \sqrt{\frac{n}{\rho(1-\rho)}}(R - \rho)$  は実現値を計算できないので検定統計量にできない。そこで、 $\rho$  を与えられた実数  $p$  で置き換えた確率変数

$$Q = \sqrt{\frac{n}{p(1-p)}}(R - p)$$

を検定統計量とする。0に近い正の実数  $\alpha$  を有意水準とする。

仮説  $\rho \neq p$  を検討する。帰無仮説  $\rho = p$  を立てる。

$$Z = \sqrt{\frac{n}{\rho(1-\rho)}}(R - \rho) = \sqrt{\frac{n}{\rho(1-\rho)}}(R - p) = Q.$$

検定統計量  $Q$  は標準正規分布  $N(0,1)$  に従うので、 $Q$  の値が0から離れるほど、つまり、 $Q$  の値の絶対値が大きいほど、確率的に不自然なので、帰無仮説  $\rho = p$  は疑わしい。

標準正規分布の確率密度関数のグラフ



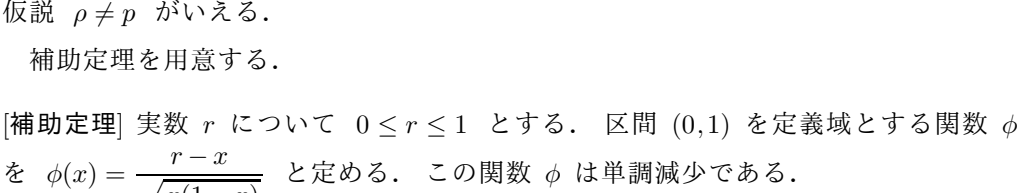
そこで  $Q$  の実現値の絶対値がある実数より大きいとき帰無仮説を棄却する。このように、数直線において帰無仮説が疑わしくなるような検定統計量の実現値の範囲が基準となる確率分布の中心（標準正規分布では0）の左右両側にあるとき、両側検定という。

帰無仮説  $\rho = p$  の下で、検定統計量  $Q$  の値の絶対値が大きいほど確率的に不自然なので、 $Q$  の値が実数  $q$  以上に不自然な実数になる確率は

$$P(|Q| \geq |q|) = P(|Z| \geq |q|) = P(Z \geq |q|) + P(Z \leq -|q|).$$

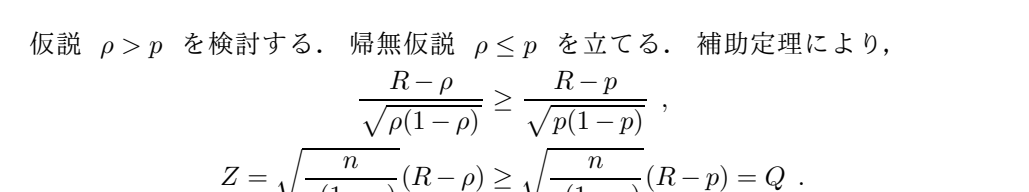
$Q$  の値  $q$  についてこの確率が小さいほど  $|q|$  が大きいので確率的に不自然である。

標準正規分布の確率密度関数のグラフ



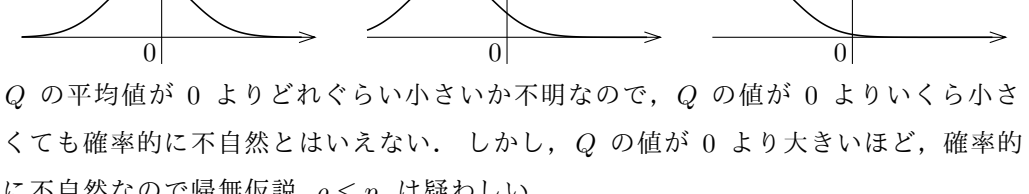
検定統計量  $Q$  の実現値  $q$  に対する確率  $P(|Z| \geq |q|)$  を  $p$  値として、この  $p$  値を有意水準  $\alpha$  と比べる。

標準正規分布の確率密度関数のグラフ



$p$  値  $P(|Z| \geq |q|)$  が有意水準  $\alpha$  より小さいとき、甚だ不自然と考えて、帰無仮説を棄却する。

標準正規分布の確率密度関数のグラフ



帰無仮説  $\rho = p$  の下で、検定統計量  $Q$  の実現値  $q$  に対して、 $p$  値は

$$P(|Z| \geq |q|) = P(Z \geq |q|) + P(Z \leq -|q|) = P(Z \geq |q|) + P(Z \geq |q|) = 2P(Z \geq |q|).$$

結局、有意水準  $\alpha$  の両側  $Z$  検定で、検定統計量  $Q$  の実現値  $q$  について  $P(|Z| \geq |q|) = 2P(Z \geq |q|) < \alpha$  のときに限り、帰無仮説  $\rho = p$  が棄却されて対立仮説  $\rho \neq p$  がいえる。

補助定理を用意する。

**[補助定理]** 実数  $r$  について  $0 \leq r \leq 1$  とする。区間  $(0,1)$  を定義域とする関数  $\phi$  を  $\phi(x) = \frac{r-x}{\sqrt{x(1-x)}}$  と定める。この関数  $\phi$  は単調減少である。

**[証明]** 実数  $x$  について  $0 < x < 1$  とする。

$$\phi'(x) = \frac{2rx - r - x}{2x(1-x)\sqrt{x(1-x)}}.$$

$0 \leq r \leq 1$  かつ  $0 < x < 1$  なので、 $-1 \leq \frac{2r-1}{2} \leq 1$  かつ  $-1 < 2x-1 < 1$ 、 $(2r-1)(2x-1) < 1$ 、よって  $2rx - r - x = \frac{(2r-1)(2x-1)-1}{2} < 0$ 。  $\phi'(x) < 0$  なので、 $\phi$  は単調減少である。 **[証明終了]**

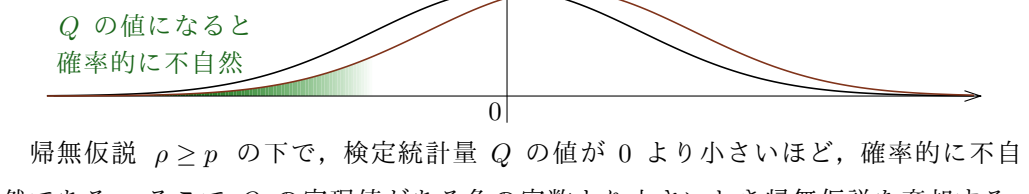
仮説  $\rho > p$  を検討する。帰無仮説  $\rho \leq p$  を立てる。補助定理により、

$$\frac{R-p}{\sqrt{\rho(1-\rho)}} \geq \frac{R-p}{\sqrt{p(1-p)}}, \\ Z = \sqrt{\frac{n}{\rho(1-\rho)}}(R-p) \geq \sqrt{\frac{n}{p(1-p)}}(R-p) = Q.$$

確率変数  $Z$  は標準正規分布  $N(0,1)$  に従うので、

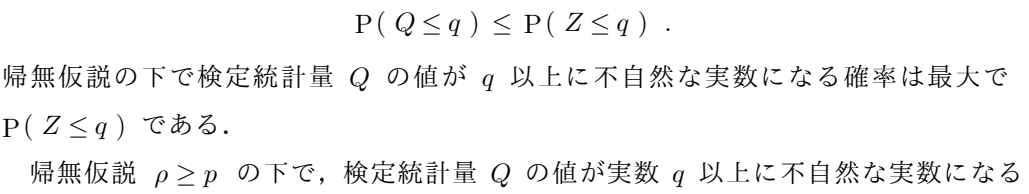
$$E[Q] \leq E[Z] = 0.$$

$Q$  の平均値が0以下なので、 $Q$  の確率密度関数のグラフは例えば次の図のようになる。



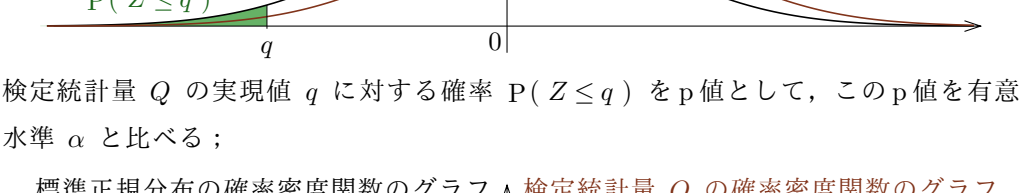
$Q$  の平均値が0よりどれくらい小さいか不明なので、 $Q$  の値が0よりいくらか小さくても確率的に不自然とはいえない。しかし、 $Q$  の値が0より大きいほど、確率的に不自然なので帰無仮説  $\rho \leq p$  は疑わしい。

検定統計量  $Q$  の確率密度関数のグラフ



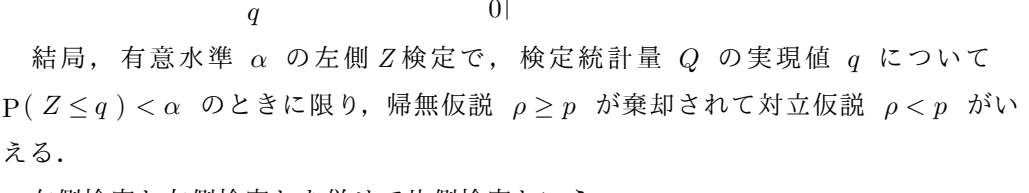
帰無仮説  $\rho \leq p$  の下で、検定統計量  $Q$  の値が実数  $q$  以上に不自然な実数になる確率は最大で  $P(Z \geq q)$  である。 $Q$  の値  $q$  についてこの確率が小さいほど  $q$  が大きいので確率的に不自然である。

検定統計量  $Q$  の確率密度関数のグラフ



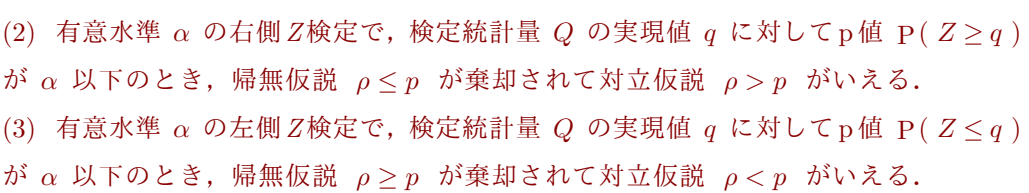
検定統計量  $Q$  の実現値  $q$  に対する確率  $P(Z \geq q)$  を  $p$  値として、この  $p$  値を有意水準  $\alpha$  と比べる。

検定統計量  $Q$  の確率密度関数のグラフ



$p$  値  $P(Z \geq q)$  が有意水準  $\alpha$  より小さいとき、甚だ不自然と考えて、帰無仮説を棄却する。

検定統計量  $Q$  の確率密度関数のグラフ



結局、有意水準  $\alpha$  の右側  $Z$  検定で、検定統計量  $Q$  の実現値  $q$  について  $P(Z \geq q) < \alpha$  のときに限り、帰無仮説  $\rho \leq p$  が棄却されて対立仮説  $\rho > p$  がいえる。

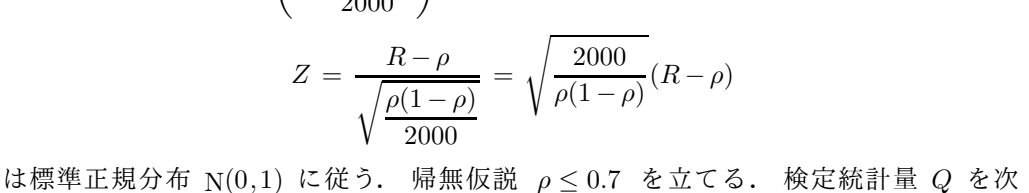
仮説  $\rho < p$  を検討する。帰無仮説  $\rho \geq p$  を立てる。先述の補助定理により、

$$\frac{R-p}{\sqrt{\rho(1-\rho)}} \leq \frac{R-p}{\sqrt{p(1-p)}}, \\ Z = \sqrt{\frac{n}{\rho(1-\rho)}}(R-p) \leq \sqrt{\frac{n}{p(1-p)}}(R-p) = Q.$$

確率変数  $Z$  は標準正規分布  $N(0,1)$  に従うので、

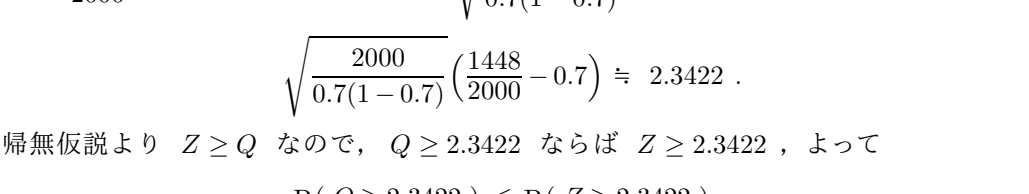
$$E[Q] \geq E[Z] = 0.$$

$Q$  の平均値が0以上なので、 $Q$  の確率密度関数のグラフは例えば次の図のようになる。



$Q$  の平均値が0よりどれくらい大きいかわからない。しかし、 $Q$  の値が0より小さいほど、確率的に不自然なので帰無仮説  $\rho \geq p$  は疑わしい。

標準正規分布の確率密度関数のグラフ



帰無仮説  $\rho \geq p$  の下で、検定統計量  $Q$  の値が0より小さいほど、確率的に不自然である。そこで  $Q$  の実現値がある負の実数より小さいとき帰無仮説を棄却する。このように、数直線において帰無仮説が疑わしくなるような検定統計量の実現値の範囲が基準となる確率分布の中心（標準正規分布では0）の左側だけにあるとき、左側検定という。

帰無仮説  $\rho \geq p$  の下で、検定統計量  $Q$  の値が小さいほど確率的に不自然なので、 $Q$  の値が実数  $q$  以上に不自然な実数になる確率は  $P(Q \leq q)$  である。帰無仮説より  $Z \leq Q$  なので、

$$Q \leq q \text{ ならば } Z \leq q,$$

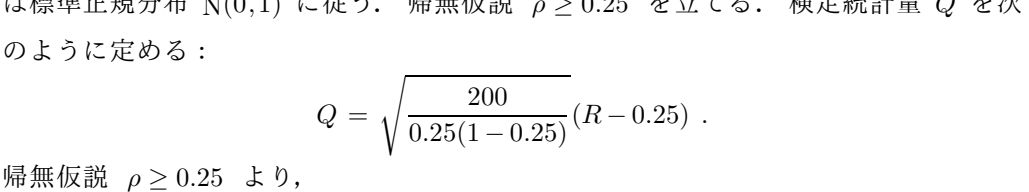
よって

$$P(Q \leq q) \leq P(Z \leq q).$$

帰無仮説の下で検定統計量  $Q$  の値が  $q$  以上に不自然な実数になる確率は最大で  $P(Z \leq q)$  である。

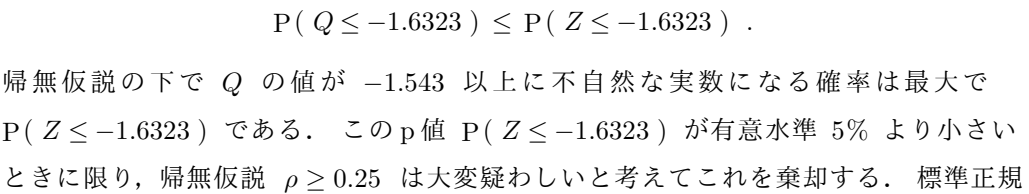
帰無仮説  $\rho \geq p$  の下で、検定統計量  $Q$  の値が実数  $q$  以上に不自然な実数になる確率は最大で  $P(Z \leq q)$  である。 $Q$  の値  $q$  についてこの確率が小さいほど  $q$  が小さいので確率的に不自然である。

標準正規分布の確率密度関数のグラフ



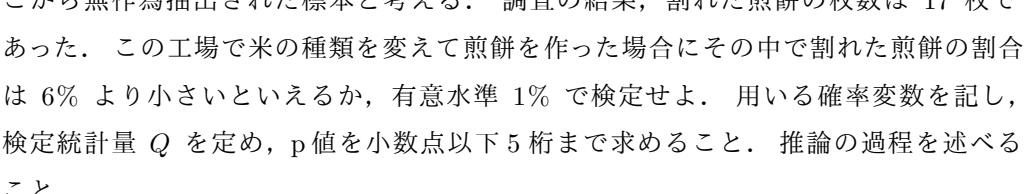
検定統計量  $Q$  の実現値  $q$  に対する確率  $P(Z \leq q)$  を  $p$  値として、この  $p$  値を有意水準  $\alpha$  と比べる；

標準正規分布の確率密度関数のグラフ



$p$  値  $P(Z \leq q)$  が有意水準  $\alpha$  より小さいとき、甚だ不自然と考えて、帰無仮説を棄却する。

標準正規分布の確率密度関数のグラフ



結局、有意水準  $\alpha$  の左側  $Z$  検定で、検定統計量  $Q$  の実現値  $q$  について  $P(Z \leq q) < \alpha$  のときに限り、帰無仮説  $\rho \geq p$  が棄却されて対立仮説  $\rho < p$  がいえる。

右側検定と左側検定とを併せて片側検定という。

**[公式8.3]** 無限2項母集団から無作為抽出された充分大きい大きさ  $n$  の標本の標本比率を表す確率変数を  $R$  とおく。標準正規分布  $N(0,1)$  に従う確率変数を  $Z$  とおく。実数  $\alpha$  は0に近い正の数とする。未知の母比率  $\rho$  及び与えられた実数  $p$  に対して、仮説  $\rho \neq p$  あるいは  $\rho > p$  あるいは  $\rho < p$  を検討するとき、検定統計量  $Q$  を  $Q = \sqrt{\frac{n}{p(1-p)}}(R - p)$  と定めると次のようになる。

(1) 有意水準  $\alpha$  の両側  $Z$  検定で、検定統計量  $Q$  の実現値  $q$  に対して  $p$  値  $P(|Z| \geq |q|) = 2P(Z \geq |q|)$  が  $\alpha$  以下のとき、帰無仮説  $\rho = p$  が棄却されて対立仮説  $\rho \neq p$  がいえる。

(2) 有意水準  $\alpha$  の右側  $Z$  検定で、検定統計量  $Q$  の実現値  $q$  に対して  $p$  値  $P(Z \geq q)$  が  $\alpha$  以下のとき、帰無仮説  $\rho \leq p$  が棄却されて対立仮説  $\rho > p$  がいえる。

(3) 有意水準  $\alpha$  の左側  $Z$  検定で、検定統計量  $Q$  の実現値  $q$  に対して  $p$  値  $P(Z \leq q)$  が  $\alpha$  以下のとき、帰無仮説  $\rho \geq p$  が棄却されて対立仮説  $\rho < p$  がいえる。

無限2項母集団から無作為抽出された充分大きい大きさ  $n$  の標本の標本比率を表す確率変数を  $R$  とおく。標準正規分布  $N(0,1)$  に従う確率変数を  $Z$  とおく。未知の母比率  $\rho$  及び与えられた実数  $p$  について仮説  $\rho > p$  を検討するとき、検定統計量  $Q$  を  $Q = \sqrt{\frac{n}{p(1-p)}}(R - p)$  と定めた。帰無仮説  $\rho \leq p$  を仮定して、 $Q$  の実現値  $q$  に対して  $P(Q \geq q)$  を考えるが、この確率は計算できない； $P(Q \geq q)$  の最大の値が  $P(Z \geq q)$  なのでこれを  $p$  値とした； $P(Q \geq q) = P(Z \geq q)$  となるのは  $\rho = p$  のときなので、 $p$  値を求めるには  $\rho = p$  のときを考えればよいことになる。それ故、対立仮説  $\rho > p$  を検討するには帰無仮説として  $\rho = p$  を仮定して  $p$  値を求めることが多い。同様の理由で、対立仮説  $\rho < p$  を検討するには帰無仮説として  $\rho = p$  を仮定して  $p$  値を求めることが多い。

**[例]** 大きさの約50万の2項母集団の中である条件に該当するものの割合について推定する。母集団から無作為抽出された大きさ2000のある標本において該当する要素の個数は1448であった。母比率は70%より大きいといえるか、 $p$  値を求めて有意水準1%で検定する。

母集団の大きさに対する標本の大きさの割合はきわめて小さいので無限母集団と考える。母比率を  $\rho$  とおく。標本の大きさは充分大きいので、標本の比率を表す確率変数  $R$  は正規分布  $N\left(\rho, \frac{\rho(1-\rho)}{2000}\right)$  に従うとする。 $R$  を標準化した確率変数

$$Z = \frac{R - \rho}{\sqrt{\frac{\rho(1-\rho)}{2000}}} = \sqrt{\frac{2000}{\rho(1-\rho)}}(R - \rho)$$

は標準正規分布  $N(0,1)$  に従う。帰無仮説  $\rho \leq 0.7$  を立てる。検定統計量  $Q$  を次のように定める；

$$Q = \sqrt{\frac{2000}{0.7(1-0.7)}}(R - 0.7).$$

帰無仮説  $\rho \leq 0.7$  より、

$$Z = \sqrt{\frac{2000}{\rho(1-\rho)}}(R - \rho) \geq \sqrt{\frac{2000}{0.7(1-0.7)}}(R - 0.7) = Q.$$

$Z$  は標準正規分布に従うので、 $E[Q] \leq E[Z] = 0$ 。検定統計量  $Q$  の値が0より大きいほど帰無仮説  $\rho \leq 0.7$  は疑わしい。よって右側検定を行う。確率変数  $R$  の実現値が  $\frac{1448}{2000}$  なので、検定統計量  $Q = \sqrt{\frac{2000}{0.7(1-0.7)}}(R - 0.7)$  の実現値は

$$\sqrt{\frac{2000}{0.7(1-0.7)}}\left(\frac{1448}{2000} - 0.7\right) \approx 2.3422.$$

帰無仮説より  $Z \geq Q$  なので、 $Q \geq 2.3422$  ならば  $Z \geq 2.3422$ 、よって

$$P(Q \geq 2.3422) \leq P(Z \geq 2.3422).$$

帰無仮説の下で  $Q$  の値が2.3422以上に不自然な実数になる確率は最大で  $P(Z \geq 2.3422)$  である。この  $p$  値  $P(Z \geq 2.3422)$  が有意水準1%より小さいときに限り、帰無仮説  $\rho \leq 0.7$  は甚だ疑わしいと考えてこれを棄却する。標準正規分布の累積分布関数を  $\Phi$  とおく。 $p$  値は

$$P(Z \geq 2.3422) = 1 - P(Z \leq 2.3422) = 1 - \Phi(2.3422) \approx 1 - 0.9904 = 0.0096.$$

$P(Z \geq 2.3422) < 0.01$  なので、帰無仮説  $\rho \leq 0.7$  は棄却される。故に、有意水準1%の右側  $Z$  検定で、母比率が0.7より大きいといえる。 **[図]**

**[図8.3.1]** ある新聞社が行った内閣支持率の調査の結果では3000人から回答が得られた。この3000人の回答者を、有権者全体を無限母集団としてそこから無作為抽出された標本と考える。標本調査の結果、内閣を指示すると回答した人の人数は1842人であった。有権者全体において内閣支持率は60%より大きいといえるか、有意水準5%で検定せよ。用いる確率変数を  $Z$  とし、検定統計量  $Q$  を定め、 $p$  値を小数点以下5桁まで求めること。推論の過程を述べること。

**[例]** 大きさの約60万の2項母集団の中である条件に該当するものの割合について推定する。母集団から無作為抽出された大きさ200のある標本において該当する要素の個数は40であった。母集団の中で該当するものの比率は25%より小さいといえるか、 $p$  値を求めて有意水準5%で検定する。

母集団の大きさに対する標本の大きさの割合はきわめて小さいので無限母集団と考える。母比率を  $\rho$  とおく。標本の大きさは充分大きいので、標本の比率を表す確率変数  $R$  は正規分布  $N\left(\rho, \frac{\rho(1-\rho)}{200}\right)$  に従うとする。 $R$  を標準化した確率変数

$$Z = \frac{R - \rho}{\sqrt{\frac{\rho(1-\rho)}{200}}} = \sqrt{\frac{200}{\rho(1-\rho)}}(R - \rho)$$

は標準正規分布  $N(0,1)$  に従う。帰無仮説  $\rho \geq 0.25$  を立てる。検定統計量  $Q$  を次のように定める；

$$Q = \sqrt{\frac{200}{0.25(1-0.25)}}(R - 0.25).$$

帰無仮説  $\rho \geq 0.25$  より、

$$Z = \sqrt{\frac{200}{\rho(1-\rho)}}(R - \rho) \leq \sqrt{\frac{200}{0.25(1-0.25)}}(R - 0.25) = Q.$$

$Z$  は標準正規分布に従うので、 $E[Q] \geq E[Z] = 0$ 。検定統計量  $Q$  の値が0より小さいほど帰無仮説  $\rho \geq 0.25$  は疑わしい。よって左側検定を行う。検定統計量  $Q = \sqrt{\frac{200}{0.25(1-0.25)}}(R - 0.25)$  の実現値は

$$\sqrt{\frac{200}{0.25(1-0.25)}}\left(\frac{40}{200} - 0.25\right) \approx -1.6323.$$

帰無仮説より  $Z \leq Q$  なので、 $Q \leq -1.6323$  ならば  $Z \leq -1.6323$ 、よって

$$P(Q \leq -1.6323) \leq P(Z \leq -1.6323).$$

帰無仮説の下で  $Q$  の値が-1.543以上に不自然な実数になる確率は最大で  $P(Z \leq -1.6323)$  である。この  $p$  値  $P(Z \leq -1.6323)$  が有意水準5%より小さいときに限り、帰無仮説  $\rho \geq 0.25$  は大変疑わしいと考えてこれを棄却する。標準正規分布の累積分布関数を  $\Phi$  とおく。 $p$  値は

$$P(Z \leq -1.6323) = \Phi(-1.6323) = 1 - \Phi(1.6323) \approx 1 - 0.9488 = 0.0512.$$

$P(Z \leq -1.6323) < 0.05$  なので、帰無仮説  $\rho \geq 0.25$  は棄却されない。故に、有意水準5%の左側  $Z$  検定で、母比率が0.25より小さいといえない。 **[図]**

**[図8.3.2]** ある製菓工場で作られる煎餅は製造工程で一部が割れる。この製菓工場で作られる原料の米の種類を変えて試しに500枚の煎餅を作った。こうして作られた500枚の煎餅を、この工場で作る米の種類を変えて作った場合の煎餅の全体を無限母集団としてそこから無作為抽出された標本と考える。調査の結果、割れた煎餅の枚数は17枚であった。この工場で作る米の種類を変えて煎餅を作った場合にその中で割れた煎餅の割合は6%より小さいといえるか、有意水準1%で検定せよ。用いる確率変数を  $Z$  とし、検定統計量  $Q$  を定め、 $p$  値を小数点以下5桁まで求めること。推論の過程を述べること。