

Ⅱ-6. 出口調査での必要サンプル・サイズ

～ 新潟県知事選挙に関するトピックス ～

2025 年 1 月 19 日

2022 年 5 月 20 日作成

M.Fujii

[選挙に関する出口調査でのサンプル・サイズ決定方法]

1,870,419 人の有権者数（令和 4 年 5 月 11 日の選挙人名簿登録者数～新潟県選挙管理委員会資料）がいる新潟県で県知事選挙が行われます。

選挙日は 2022 年 5 月 29 日（日）です。

有権者を対象とする出口標本調査（または単純無作為抽出による標本調査、期日前投票と当日投票で標本調査）で、新潟県知事候補者の得票率を区間推定したいとすると、必要な標本数はいくらぐらいでしょうか。必要なサンプル・サイズを計算したいと思います。

信頼係数 99% の信頼区間が 誤差 1% 以下となるようにするには、少なくとも何人以上の有権者を調査すればよいでしょうか。ただし、調査された人は必ず H 氏または K 氏のいずれかに投票し回答したものとし、二項分布は近似的に正規分布に従うとしておきましょう。

H 氏、K 氏の得票率を p 及び $(1-p)$ としましょう。出口標本調査による H 氏の標本得票率を $\bar{x}$ とおき、その標本得票率の平均を $\bar{\bar{x}}$ としましょう。

おそらく、下記に示すとおりで、5 月 29 日夜、開票率 1% に満たないところで、当選確実、いわゆる「当確」が出ると考えます。

ケース I.

H 氏及び K 氏の支持率について事前の情報が全くない場合には、少なくとも何人以上の有権者を調査すればよいでしょうか。

母比率の 99% の信頼区間の推定は、標本得票率の平均を $\bar{\bar{x}}$ として、

$\bar{\bar{x}} \pm 2.576 * \{(\bar{\bar{x}} * (1-\bar{\bar{x}})/n)\}^{1/2}$ で計算できることが分かっています。

母集団の得票率 p の信頼区間 C.I. は、以下のとおりになります。

$$\bar{\bar{x}} - 2.576 * \{(\bar{\bar{x}} * (1-\bar{\bar{x}})/n)\}^{1/2} \leq p$$

$$(上記式続き) \quad p \leq \bar{\bar{x}} + 2.576 * \{(\bar{\bar{x}} * (1-\bar{\bar{x}})/n)\}^{1/2}$$

H 氏及び K 氏の支持率についての情報がハッキリしないため、母得票率が最も大きくなる $\bar{\bar{x}} = 0.5$ を用いて、下記式の $\bar{\bar{x}}$ に代入して計算します。

(注) 事前に他調査で同様な調査結果がある場合はその得票率を用いますが、事前に参加となる結果がない場合は必要な調査対象者数が最大となる 0.5 を入れます。

標本誤差には、調査結果で容認できる誤差を入れます。例えば、調査結果の誤差を 1% に抑えたいという場合であれば、0.01 を $(\bar{\bar{x}} - p)$ に代入します。

信頼区間 C.I. の式を変形して、必要サンプル・サイズ数 n を求めます。

$$n \geq (2.576)^2 * \{(\bar{\bar{x}} * (1-\bar{\bar{x}}))\} / 0.01^2 \quad \bar{\bar{x}} = 0.5 \text{ を代入}$$

答えは、16,590 人以上になります。

新潟県の有権者数に占める割合は、0.887% です。

0.887% の有権者を調査すれば、H 氏及び K 氏の得票率が誤差 1% として 99% 信頼区間が推定できます。

ケースⅡ.

調査の結果から、H 氏の支持率がおおよそ 80% であることがわかっている場合は、少なくとも何人以上の有権者を調査すればよいでしょうか？

今回は H 氏の支持率、つまり標本得票率が $\bar{x} = 0.8$ と分かっているので、それを使います。 計算自体は前項と全く同じです。

標本誤差には、調査結果で容認できる誤差を入れます。 例えば、調査結果の誤差を 1% に抑えたいという場合であれば、0.01 を $(\bar{x} - p)$ に代入します。

信頼区間 C.I. の式を変形して、必要サンプル・サイズ数 n を求めます。

$$n \geq (2.576)^2 * \{(\bar{x} * (1 - \bar{x}))\} / 0.01^2 \quad \bar{x} = 0.8 \text{ を代入}$$

答えは、10,618 人以上となります。

新潟県の有権者数に占める割合は、0.568% です。

0.568% の有権者を調査すれば、H 氏及び K 氏の得票率が誤差 1% として 99% 信頼区間が推定できます。

なお、ケースⅠ. 及びⅡ. とも、実際は調査対象者の全員から回答が得られるとは限らないため、想定される回収率を踏まえて、計算で得られた調査対象者数より多めに対象者数を見積もっておく必要はあります。

< 追加考察 >

上記考察の条件を緩めましょう。

信頼区間 95% の推定を考えることにして、調査結果の誤差は 3% を目指すことにしましょう。

ケース I plus.

信頼区間 C.I. の式を変形して、必要サンプル・サイズ数 n を求めます。

$$n \geq (1.960)^2 * \{(\bar{x} * (1 - \bar{x}))\} / 0.03^2 \quad \bar{x} = 0.5 \text{ を代入}$$

答えは、1,068 人以上になります。

新潟県の有権者数に占める割合は、0.057% です。

ケース II plus.

信頼区間 C.I. の式を変形して、必要サンプル・サイズ数 n を求めます。

$$n \geq (1.960)^2 * \{(\bar{x} * (1 - \bar{x}))\} / 0.03^2 \quad \bar{x} = 0.8 \text{ を代入}$$

答えは、683 人以上になります。

新潟県の有権者数に占める割合は、0.037% です。

たった 0.057% もしくは 0.037% の出口標本調査で誤差 3% として

H 氏及び K 氏の得票率の 95% 信頼区間が推定できます。

以上