平成27年度 数学科リレー講座「統計入門」

5日目「推定」

担当: 網谷 泰治

1　推定とは何か？

統計調査には, 対象とする集団についてもれなく調べる**全数調査**と, 集団から一部分を取りだして調べる**標本調査**とがあります.

標本調査では, 本来調べたいものの集まりを**母集団**といいます. 母集団から抜き出された集まりを**標本**といいます.

推定

母集団における統計量 (平均・比率など) の値を, 標本の統計量 (平均・比率など) の値から求めること.

推定には, 大まかに 2 つの種類があります.

① 母集団の統計量の値をズバリ推測すること

② 母集団の統計量の値を いくらかの幅を持たせて推測すること

①を「**点推定**」といい, ②を「**区間推定**」といいます.

2　点推定

ある野生動物の個体数を推定するために, 古くから「捕獲–再捕獲法」という方法が用いられています.

捕獲–再捕獲法

ある決まった場所 (たとえば, 湖) に生息する野生動物の 1 種 A の個体数 N を推定する.

① まず, 動物 A を n 匹捕獲して, タグを付けたあと, 放す.

② 次に, 十分な時間が経ってから, 動物 A を捕獲する. 捕獲された動物 A の個体数を M, そのうちタグのついている (再捕獲された) ものの数を m とする.

この n, m, M の値から, N の値を推定する.

推定される個体数 $\widehat{N}$ は, どのように表せるでしょうか？

▶ **例 1**　ある湖にいる「オオクチバス」の個体数を推定してみよう.

	1 回目捕獲数 n	2 回目捕獲数 M	再捕獲された数 m
オオクチバス	213	104	13
コクチバス	232	329	16

▷ **練習 1**　例 1 で「コクチバス」の個体数を推定してみてください.

答　推定される個体数は $\widehat{N} =$_____

《メモ》

3　区間推定

問題

A 県で, 14 歳の男子 400 人を無作為に抽出して, 1 年間の身長の伸びを測りました. その 400 人の平均値は 4cm でした. A 県の 14 歳の男子の身長の伸びがつくる母集団の平均値とこの標本の平均値とぴったり一致することは稀 (まれ) なことです.
それでは, この値 4cm は母平均に どの程度近いと考えて良いでしょうか？

標本平均, さらに標本平均のばらつき (分布) を用いて, 母平均を推定します.

母平均 m, 母標準偏差 σ をもつ母集団から, 大きさ n の標本を抽出します. n が大きいとき, 標本平均 $\overline{X}$ の分布は　　分布と考えられます.

$$N\left(m, \frac{\sigma^2}{n}\right)$$

$$\left(\overline{X}\text{と } m \text{ の差が } c \times \frac{\sigma}{\sqrt{n}} \text{以下となる確率}\right) = 2p(c)$$

c の値を決めると $p(c)$ の値を正規分布表から求めることができます. 実際には, c ではなく, 確率の値を先に決めます. 確率の値を決めると, それに応じて c の値が決まります. 確率の値は大抵 95% や 99% が設定されることが多いです. 分布表から, 95% の場合は, $c = 1.96$ です. 区間

$$\left[\overline{X} - 1.96 \times \frac{\sigma}{\sqrt{n}},\ \overline{X} + 1.96 \times \frac{\sigma}{\sqrt{n}}\right]$$

を, 母平均 m の **信頼度 95% の信頼区間** と呼びます.

《メモ》

▶ **例 2**　先ほどの問題で母平均 m が不明で, 母標準偏差 $\sigma = 2.8\text{cm}$ であることは分かっているとします.

$$1.96 \times \frac{2.8}{\sqrt{400}} = \underline{\qquad\qquad\qquad}$$

となるので, 信頼度 95%の信頼区間は $\underline{\qquad\qquad\qquad\qquad}$ となります.

母平均に対する信頼区間

標本の大きさ n が大きいとき, 母平均 m についての信頼度 95%(99%) の信頼区間はそれぞれ

- $\left[\overline{\boldsymbol{X}} - \mathbf{1.96} \times \dfrac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}},\ \overline{\boldsymbol{X}} + \mathbf{1.96} \times \dfrac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}\right]$,
- $\left[\overline{\boldsymbol{X}} - \mathbf{2.58} \times \dfrac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}},\ \overline{\boldsymbol{X}} + \mathbf{2.58} \times \dfrac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}\right]$

区間の信頼度の意味

信頼度 95% とは具体的には何のことをいっているのでしょうか？この意味を理解するために, 皆さんと一緒に「公開実験」をしてみましょう！

次のページの乱数表 (100 個の整数が書かれている) を使います.

手順

(1) 100 個の整数のうち, 10 個を無作為に選んで○をつける.

(2) 同じように 100 個の整数のうち, 10 個を選んで△をつける.

(3) ○と△の平均をそれぞれ計算する. それぞれ, Ⓐ, Ⓑとします.

(4) ⒶとⒷの平均を計算する. これをⒸとします.

《計算用紙》

□ 計算結果

○の平均 Ⓐ _____　　△の平均 Ⓑ _____

ⒶとⒷの平均 Ⓒ _____

□ 黒板での計算結果

Ⓐ, Ⓑが 95% の信頼区間に入る確率 _____

Ⓒが 95% の信頼区間に入る確率 _____

信頼度 95% の信頼区間の意味

母集団から十分な大きさをもつ標本 (大きさ n) を抽出して, 信頼区間を作る操作を何回も繰り返す. これらの区間のうち母平均 m を含むものがほぼ 95% ある.

1 日目, 2 日目の講座にもありましたが, 確率は 1 通りではなく, 数学的確率 (先験的確率ともいいます) と統計的確率 (経験的確率) があるということでしたね. 生活していると, 身近にいろいろと確率を目にしますが, きちんと理解できているでしょうか.

• **クイズ**　受験生の A 君は ある模試を受けました. その結果, 志望する大学の合格率は 25% でした. さて, この合格率 25% の意味は次のうちどれでしょうか？

① 志望する大学を A 君が 4 回受ければ, そのうち 1 回は合格するということ

② 過去の模試で, A 君の成績と同程度にある階級の生徒たちのうちで志望大学に受かったものが統計上 25% いたということ

③ 今回の模試で, A 君の成績と同程度にある階級の生徒たちのうちで志望大学に受かるものが 25% いるということ

▶ **例 3**　例 2 では, 母標準偏差の値が分かっている場合を考えましたが, 実際にはその値が分からないことが多いのです. もし, 標本の大きさ n が大きいならば, σ の代わりに 標本標準偏差 s を用いて計算してもよいということが知られています. [1]

あるジュースの缶 40 缶について, A 成分の含有量を検査したところ, 平均値 32.5mg, 標準偏差 3.1mg を得ました. 1 缶あたりの A 成分の平均含有量を, 信頼度 95% で推定してみましょう. ここで, 簡単のため, $3.1/\sqrt{40} = 0.49$ としてみましょう.

答　95% の信頼区間は _____

[1]「分散」は標本をとったとき, $\sigma^2 = (n/(n-1))s^2$ となり若干目減りします. n が大きいと, $n/(n-1)$ はほぼ 1 と見なせます. 正規分布に近似し, 推定してよいことの根拠として, 中心極限定理があります.

母比率の推定に区間推定の考え方を用いてみましょう.

ある性質に注目します. その性質をもつ母比率 p となる母集団から, 大きさ n の標本を無作為に抽出します. n が大きいとき, その性質の標本比率 R の分布は正規分布と考えられます.

$$N\left(p,\ \frac{p(1-p)}{n}\right)$$

母平均のときと同様にして, 95% の信頼区間が

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

と求められます.

n が十分に大きいと, 大数の法則から, R を p に近いとみなせます. その結果, 次が分かります.

母比率に対する信頼区間

標本の大きさ n が大きいとき, 母比率 R についての信頼度 $95\%(99\%)$ の信頼区間はそれぞれ

- $\boldsymbol{\left[R-1.96\times\sqrt{\frac{R(1-R)}{n}},\ R+1.96\times\sqrt{\frac{R(1-R)}{n}}\right]}$,
- $\boldsymbol{\left[R-2.58\times\sqrt{\frac{R(1-R)}{n}},\ R+2.58\times\sqrt{\frac{R(1-R)}{n}}\right]}$

例題

ある地域で有権者 5000 人を無作為に抽出して, 政党 A の支持者を調べたところ, 1500 人でした. この地域の政党 A の支持率 p を, 信頼度 95% で推定してみましょう.

答　95% の信頼区間は _____

▷ **練習 2**　例題の支持率について, 信頼度 99% で推定してください.

答　99% の信頼区間は _____

4　区間推定の応用

応用はたいへん多岐にわたりますが, ここでは 2 つに絞って紹介します.

① 内閣支持率・総選挙の結果などの推定

日本国 (母集団) の内閣支持率を全数調査することは, 経費の面などで問題が出てきます. たとえば, 母比率が 40% と予想される場合, どの程度の人数に調査を行えばよいでしょうか.

このような場合に, 区間推定の考え方が有効です. 表 2 によると, 母比率が 40% と予想される場合, 無作為に 1000 人を選んで調査すれば, 誤差 ±3.0% 程度の結果が得られることが分かります. 3000 人の調査なら, 誤差を ±1.8 まで小さくできます.

	10%	20%	30%	40%	50%
500 人	2.6	3.5	4.0	4.3	4.4
1000 人	1.9	2.5	2.8	3.0	3.1
2000 人	1.3	1.8	2.0	2.1	2.2
3000 人	1.1	1.4	1.6	1.8	1.8

表 2　信頼度 95% を与える誤差 $(1.96 \times \sqrt{p(1-p)/n} \times 100)$ の値

衆議院や参議院の総選挙の夜にはテレビで結果の速報が報道されます. これは, 出口調査により得られた予想獲得議席数の区間推定の結果を報道しているわけです. 標本をとったことによる誤差は必然的に生じますが, 他にも, 誤差の要因はあります. 実際に投票したにも関わらず違った政党を答えてしまうなどの「測定誤差」や何らかの偏り「抽出バイアス」などの影響も考えられます.

② 薬効の推定

薬の効きがどの程度か, 推定の考え方を使う方法もあります. 表 3 を睡眠薬の効果実験の結果とします. たとえば, 不眠で困っている人がいたとき, 睡眠時間を増加させるために, A, B のうち, どちらの薬をすすめたらよいでしょうか.

患者	薬 A	薬 B
1	0.7	1.9
2	-1.6	0.8
3	-0.2	1.1
4	-1.2	0.1
5	-1.0	-0.1
6	3.4	4.4
7	3.7	5.5
8	0.8	1.6
9	0	4.6
10	2.0	3.4
$\overline{m}$	0.66	2.33
s	1.86	2.00

表 3　薬による睡眠の増加時間

今の場合ですと, データの数が少ないため, 正規分布ではなく「t 分布」を使って考えます. すると, 薬 A, B の 95% 信頼区間はそれぞれ次のようになります.

$$[-0.67,\ 1.99], \quad [0.90,\ 3.76]$$

その結果, 薬 A では平均的に睡眠時間が増加しない場合が区間に含まれてしまっているため, この薬を積極的に患者にすすめることはできません. 一方で, B では区間が 0.90 時間から始まっているため, B は効果があると考えられます.

ただ, 積極的に薬 B をすすめる場合, もっとデータをとったり, 薬理学的な領域 (薬効のメカニズムの分析) にまで踏み込むなど総合的に判断することが普通です.

以上 2 つの例を紹介しましたが, 区間推定の特色は, 分布を用いることで, 誤差を含めた上で, 推定の信頼性をきちんと数値化できることにあります. 区間推定の応用については, さまざまな応用がありますから, いろいろと関心を高めて, 探ってみましょうね.

参考文献

[1] 盛山 和夫, 『統計学入門』, ちくま学芸文庫, 2015.

[2] 田栗 正章・藤越 康祝・柳井 春夫・C.R.Rao, 『やさしい統計入門』, ブルーバックス B–1557, 講談社, 2007.

[3] 松井 敬, 『統計的推測』, 数学のかんどころ 11, 共立出版, 2012.

[4] 蓑谷 千凰彦, 『推定と検定のはなし』, 東京図書, 1988.

[5] 高橋 陸男, ほか 6 名,『四訂版 高等学校 確率・統計』, 数研出版, 1992.