

## 2012 年度 森林統計学 第 8 回課題 [課題 4] ランダムサンプリング

目的：標本抽出（無作為抽出；ランダムサンプリング）を実際に行なうことによって、母集団と標本集団の関係を考察する。[標本からいかにして母集団の詳細を推定することができるか?]

基本課題（7 点）：

- ①  $N \geq 50$  のデータを集める（多い方がよい）。これを母集団とみなして、以下で用いる。
- データは連番も付けて表にする。精度（最小計測単位）が明確にわかるように。ただし、この時点では大小順の並べ替えは不要（並べ替えない方がよい）。
  - 母集団として取り扱うことが適切な意味を持つデータとすることが好ましい。課題 1（ヒストグラム）や課題 2（基礎統計値）で用いたものと重複するデータでもよいが、全く同じにはしないこと。
- ※③で求める母平均・母標準偏差などが以前の課題と同じ数値にならないようにするため [結果を流用すると練習にならないので]；データを 1, 2 個追加あるいは削除する程度の調整でよい。
- ※データ数を  $n$  でなく  $N$  で表しているのは、母集団であることを明確にするため。
- ※引き続き②では、母集団は「未知」としてサンプリングを行ない、母平均についての推定を行なう。なので、データの詳細はこの時点では知らない方がよい（「わからない」という想定）。
- ② 無作為抽出（教科書 p.122～123 参照；下記追記も参照）により、母集団から標本抽出を行なう。標本抽出は、以下の 3 パターンとする [ $n$  と抽出回数的大小が与える影響を知るため]。
- (1)  $n = 5$  を 6 回, (2)  $n = 5$  を 12 回, (3)  $n = 10$  を 6 回。
  - (1)～(3)のそれぞれについて、抽出結果の一覧表を作成し、
    - 各標本抽出毎に平均値  $\bar{x}$ （標本平均）[と  $s$ （標本標準偏差）；応用課題用] を算出する。  
※標本なので、標準偏差  $s$  は教科書 p.21, (9)式を用いて求める ( $n-1$  で割る)。
    - 得られた抽出回数分の個数の平均値について、その平均値  $\bar{\bar{x}}$ （標本平均の平均）、と標準偏差  $s_{\bar{x}}$ （標本平均の標準偏差）を算出する（例：(1)の場合なら  $\bar{\bar{x}}$  は 6 個、(2)は 12 個、(3)では 6 個求められる；(1)～(3)のそれぞれについて、 $\bar{\bar{x}}$  と  $s_{\bar{x}}$  は各 1 個となる）。  
※標本平均も標本の一種なので、その標準偏差は教科書 p.21, (9)式を用いて求める ( $n-1$  で割る)。
    - 得られた結果を用いて母平均（母集団の平均値）を推定してみる（(1)～(3)のそれぞれについて；標本から考えて母平均の値はどのくらいと予想されるか）。また、(1)～(3)のそれぞれの推定の信頼性の高さについて、算出した  $\bar{\bar{x}}$ 、 $\bar{\bar{x}}$ 、 $s_{\bar{x}}$  を用いて考察する。
- ※合計 24 回の標本抽出を行なうことになるが、その一回一回が、実際の調査や実験におけるデータ採取に相当する。つまり、(1)あるいは(2)の場合ならば、 $n = 5$  の標本抽出（5 回の実験・5 個のプロットを設定しての調査・5 人を抽出してのアンケート調査、など）を実施して、その結果から母集団の平均値を推定することに相当する。
- ※実際の調査や実験では、データ採取は一回しか行なわない。今回は標本抽出という作業の性質を知るために、あえてデータ採取を同じ条件で 6 回((1), (3)の場合)、あるいは 12 回((2)の場合)、繰り返している。
- ③ 標本抽出が終わった時点で、母集団の性質・分布形について調べて概要を考察しまとめる。
- 母集団の集計は、②が終わったこの時点で行なう。考察には、度数分布表とヒストグラム、

基礎統計値（平均値[母平均  $\mu$ ]・標準偏差[母標準偏差  $\sigma$ ]・最大値・最小値・中央値・四分位数）を作成・算出して用いる。 $\mu \pm \sigma$ 、 $\mu \pm 2\sigma$ に入る確率についても確認しておく。

※標準偏差は、母集団であるため、 $N - 1$ でなく  $N$ で割って算出する（教科書 p.20, (5)式で分散を算出しその平方根をとる）。

※実際の調査や実験では、母集団の詳細は不明である（正解は不明）。今回は、普通にはないことだが、母集団の全数調査が可能なので正解を知ることができる。この③では、標本抽出実験で得られた結果が正しかったかどうかの答えあわせをするために、母集団の集計をする[これは課題 1~3 の復習も兼ねている]。

④ ②での母平均の推定結果と③の正しい母平均とを比較し、考察をする。考察の要点は:

- (1) 母平均と標本平均の相違、母集団の分布と標本平均の分布の相違（②の3つの抽出パターンとの関係はどうか； $n$ ,  $\bar{x}$ ,  $s_x$ , 抽出回数 の一覧表を作ると比較考察しやすい）

※ 標本平均の分布をより詳しく考察するためには、②で抽出した標本のそれぞれについてヒストグラムを作成するとよい。その場合、母集団との比較を容易にするためには、標本の度数分布表の集計区間は母集団のものと同じとする。

- (2) 教科書第 6 章の定理 1、定理 2 に照らして、得られた結果はどのように解釈できるか

応用課題（3点）：

- ②で抽出した標本のそれぞれの標準偏差  $s$ （標本標準偏差）は、母標準偏差  $\sigma$  の不偏推定値（教科書第 6 章 3 節参照）になっているかどうかについて考察する。

※ 手がかりとしては、 $s$  と  $\sigma$  の差異を比較する（ $s$  はその平均や分布を見るとよい）こと、もし  $n - 1$  でなく  $n$  で割っていたら（そのようにして算出した標本標準偏差を  $s'$  とする）、 $\sigma$  との差異はどのように変わるか、など

無作為抽出の方法に関する追記

1. 手作業による方法（最も確実）

データの一覧表を（できれば少し拡大して）印刷し、一行毎に短冊状に切る。できた  $N$  枚の短冊を袋などに入れてかきまぜ、 $n$  枚を取り出す。

2. 乱数表を用いる方法（教科書の方法；p.123 下から 7 行目～p.124 上から 13 行目を参照）

教科書 p.304～307 の表 XI を用いる。なんらかのルールを自分で決め、表の中から  $n$  個の数字を読み取る。それに相当する連番のデータが抽出された標本となる。

注：エクセルの乱数関数「=RAND()」を用いる方法もある。関数ヘルプ等で自習して、この方法を用いてもよい（[課題 3]の Web 資料（例 2）にも方法の解説あり）。