

## 2010 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知る。

基本課題 (7 点):

- ①  $N = 50$  以上のデータ ( $N \geq 50$ ) を集める。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(大小順の並べ替えは不要; 並べ替えない方がよい)。精度(最小計測単位)が明確にわかるように。
  - 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。
  - 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。
- ② 無作為抽出(教科書 p.122~123 参照; 下記追記も参照)により、母集団から以下の 4 パターンの標本抽出を行なう。
  - (1)  $n = 5$  を 5 回, (2)  $n = 10$  を 5 回, (3)  $n = 5$  を 10 回, (4)  $n = 10$  を 10 回。
  - それぞれについて、平均値  $\bar{x}$  (標本平均) と標準偏差  $s$  (標本標準偏差)、平均値の平均値  $\bar{\bar{x}}$  (標本平均の平均)、平均値の標準偏差  $s_{\bar{x}}$  (標本平均の標準偏差) を算出する(例: (1) の場合なら  $\bar{\bar{x}}$  と  $s$  はそれぞれ 5 個,  $\bar{x}$  と  $s_{\bar{x}}$  はそれぞれ 1 個求められる)。
  - 注: 標本なので、標準偏差は教科書 p.21, (9) 式を用いて求める ( $n-1$  で割る)。
- ③ 母集団の性質・分布形について考察する。
  - 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。
  - 注: 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)。
- ④ 結果について考察を加える。考察では以下の 2 点については必ず押えること。
  - (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 4 つの抽出パターンとの関係はどうか)
  - (2) もし標本の標準偏差を  $n-1$  でなく  $n$  で割って計算していたら、(1) の結果はどのようなになるか。

自主課題 (3 点):

- ② で抽出した標本のそれぞれについて度数分布表とヒストグラムを作成し、④ の (1) において「母集団の分布と標本の分布の相違」についても考察項目を追加。
- 注: 比較を容易にするために、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。

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。使いこなす自信のあるものは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)。