

2009 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。

課題:

- 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。精度(最小計測単位)が明確にわかるように。
 - 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、データ数は 50 以上となるようにする。
 - 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 母集団の性質を知るために、課題 1 と課題 2 に準じて度数分布表とヒストグラムを作成し、基礎統計値(必須: 平均値・分散・標準偏差; 他は任意)を算出する。
 - 度数分布表とヒストグラムについては、区間数は 5 とする(後で n が小さい標本と同じように比較するため)。
 - 基礎統計値については、平均値・分散・標準偏差については必須、他(4 分位数等)は任意とする。
 - 注: 分散と標準偏差は、母集団であるため、 $N - 1$ でなく N で割って算出する。(分散は教科書 p.20, (5)式を用いる; 標本の場合は p.20, (7)式を用いることになる。)
 - ③ 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、度数分布表とヒストグラムを作成し、平均値 \bar{X} と標準偏差 s 、平均値の平均値、平均値の標準偏差を算出する。
 - 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。
 - ④ 結果について考察を加える。
- 自主課題: ③で、(5) $n = 15$ を 10 回、(6) $n = 20$ を 10 回、の 2 パターンを追加。

無作為抽出の方法

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)

教科書 p.304~307 の表 XI を用いる。適当にルールを決め、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。具体的な方法は教科書 p.123 下から 7 行目~p.124 上から 13 行目に書かれている。

3. エクセルの乱数関数を用いる方法

「=RAND()」で 0~1 の乱数が作られる。データ一覧の隣に乱数の列を加え、連番とデータの列もあわせて乱数列をキーにして昇順(降順でもよい)で並べ替え、上から n 個を抽出する。