

2010 年度 森林統計学 第 2 回課題 データの集計 2: 記述

今回は、前回の課題として集めたデータから度数分布の図表を作成することが課題である。

練習問題:

- 1) 教科書 p.29, 例 2. (a), (b)
- 2) 教科書 p.33, 問題 28. (a), (b)
- 3) 教科書 p.34, 問題 29. (a), ただし「問題 28 の(b)から(i)までの質問に答えよ」は(b)のみ

課題(基本点 7 点分):

- ① 前回の課題として集めたデータのそれぞれについて度数分布表を作成する。ただし、階級幅等は下記の注の方法(データ数に応じて決める方法)に従う。
- ② 度数分布表を図(ヒストグラム)化する。形式は教科書 2 章の図 2~4 (p.14~15) を参照。
※図は手描きでよい。目盛がそこそこ正確に書かれていれば、方眼紙等を使用せずに普通のレポート用紙に書いたものでもよい。ただし紙のサイズは A4とする(提出課題のコピー時に B5 等では支障があるため)。
- ③ 分布の中心、分布のひずみの有無などについて考察し、考察の結果を記述する。
- ④ 氏名・学生番号など必要な情報を付記して提出 (A4 又は A4 で印刷可能な体裁のファイル)
※前回課題の「データ収集」で指示した項目とあわせて、レポートには以下の項目が必要。
 - ・データの説明, データの一覧表, 採集方法・留意点の説明, 上記①~③の図表と考察

自主課題(追加点 3 点分):「練習問題」の 3 つのデータを同様の方法で集計(区間の数と幅は下記注の方法で決める)し、比較考察する。

注) 度数分布表の作り方(永田靖, 入門統計解析法, 日科技連, 1992, p.9-15 「1.3 度数表とヒストグラム」より)

o 度数分布表(略して「度数表」ともいう)の作成手順

1. データの測定単位 m (測定の最小きざみ) を明確にして、サンプリングを行う。データ数を n とする。
2. データより最大値 x_{max} 、最小値 x_{min} 、範囲 R を求める。(※ $R = x_{max} - x_{min}$)
3. 仮の区間の数 h を $h \cong \sqrt{n}$ をもとにして決める。

4. 区間の幅 c を $c \cong \frac{R}{h}$ をもとにして決める。この際、 c は測定単位 m の整数倍に丸める。

5. 各区間の境界値を次のように決める。

一番下の境界値を $x_{min} - \frac{m}{2}$ とする (※ 境界値とデータの値が一致しないようにするため)。これに区間の

幅 c を順次加えていって各区間の境界値とし、 x_{max} を含むまで繰り返す。

6. 区間の中心値をその区間の上側と下側の境界値の平均として求める。

7. 各区間に入るデータの個数(度数)を数え上げ、表にまとめる。

※ 区間数と幅が決まれば、度数は Excel の関数 FREQUENCY でも計算できる。

o ヒストグラムの作成手順

1. 度数表の区間の境界値を横軸にとり、度数を縦軸に取る。

2. 度数表に基づいて、それぞれの区間の度数に応じて柱を描く。

3. は次回に作成(次回に次回の課題として改めて説明する) → [3. 平均を示す線を書き込み、ヒストグラムの余白に n, \bar{x}, s (標準偏差; 不偏分散から求めた標準偏差 $\hat{\sigma}$ を用いるのが一般的) を記入する。]