

教養科目『統計学』（社会）任意提出レポート No.2 略解

担当：星野（経済学部講師）／2000年1月25日

1. 次のデータは、諸君に答えてもらった兄弟姉妹の数である。{ 4 2 2 3 2 2 2 1 3 3 2 3 3 2 2 2 2 2 2 3 2 4 2 2 3 3 2 3 3 3 2 2 3 1 3 2 2 3 4 3 2 3 3 2 3 3 6 2 3 4 2 4 2 3 2 3 2 4 3 2 1 3 2 } このデータについて下記の計算をせよ。

(a) 大きさ順に並び替えて見よ。(略)

(b) 度数分布表を作れ。

•

兄弟姉妹数	1	2	3	4	5	6
度数	3	30	24	6	0	1

(c) ヒストグラムを描け。(略)

(d) 五数要約をせよ。

- 最小値 $x_{(1)} = 1$ 、第一四分位点 $(x_{(16)} + x_{(17)})/2 = 2$ 、メディアン $(x_{(32)} + x_{(33)})/2 = 2$ 、第三四分位点 $(x_{(48)} + x_{(49)})/2 = 3$ 、最大値 $x_{(64)} = 6$ 。データ数 64 が 2,4 で割り切れる事に注意。

(e) モード、平均、分散を計算せよ。

- モードは 2 (30 では無い!)、平均は $(1 \times 3 + 2 \times 30 + \dots + 6 \times 1)/64 \doteq 2.58$ 。分散は概算で $((1-2.58)^2 \times 3 + (2-2.58)^2 \times 30 + \dots + (6-2.58)^2 \times 1)/64 \doteq 0.71$ のようである。64 で割るのを忘れている人が何人か目に付いたが、私が黒板で間違えた式を書いたのだろうか? 分散 $V(x)$ の定義: $V(x) = ((x_1 - \bar{x})^2 + \dots + (x_n - \bar{x})^2)/n$ 、但し \bar{x} は x の平均、を確認。なお $\sqrt{V(x)}$ は標準偏差である。

(f) 実は 6 というデータは、異常値である。(家族の数だそう) この 6 を除いた残りのデータセットについて、五数要約をしてみよ。また平均を計算せよ。

- 最小値 $x_{(1)} = 1$ 、第一四分位点 $x_{(16)} = 2$ 、メディアン $x_{(32)} = 2$ 、第三四分位点 $x_{(48)} = 3$ 、最大値 $x_{(63)} = 4$ である。今度のデータ数 63 は奇数。平均は約 2.52 らしい。

(g) このデータから何が読みとれるか。自由に述べよ。

- 約 20 年前の出生に関するデータな事に注意。昨今の少子化を示すものでは必ずしもない。

2. 表 1 (略) は、諸君の中で特定のプロ野球チーム (一人一球団にしぼった) をひいきにする者の度数、及び 2000 年度観客動員数 (単位: 万人) を球団毎にまとめた物である。このデータについて、下記の分析を試みよ。

(a) 度数と観客動員数について、相対頻度のヒストグラムを書いて見よ。

- 「相対」頻度の総和は 1 である事に注意。(例えば巨人の相対度数は $15/(15+8+2+\dots+1) \doteq 0.3$) このように単位が異なる変数も、同じスケールに調整すれば比較出来る。

(b) 球団のファンの数は観客動員数に比例すると思われる。しかし地域的な偏りが有り、本拠地周辺に比較的多い事が知られている。データから、本講義に参加している学生の出身地域特性を推論せよ。

- 母集団と標本の話の「枕」にするつもりだったが、時間切れで詳しく扱えない。ここで諸君が推論に用いた暗黙の前提を明らかにするのが、統計的標本調査論である。強調したいのは、日常的な事を厳密に議論するという統計学のあり方である。賢明に生きる為に、統計学を勉強しましょう。(宣伝)

3. 本講義の理解を試す為の問題を作って見よ。適当なものは改変した上、期末試験に出題する。(略)