確率統計と情報処理・演習（2007 年度後期）

1 変量データ分析

2007 年 10 月 05 日

日本女子大学理学部数物科学科　　今野 良彦

September 25, 2007

今日の講義の目的と概要

目的

（1）データが入力されたら，とりあえずデータにどんな値がどれくらいかるかを調べる．この様子をあらわしたものを「分布」という．表現法として，度数分布表や図（ヒストグラム）がある．

（2）分布を説明するために，分布の中央あたりの値を「分布の中心」として「分布の代表値」とする．

（3）つぎに，データのばらつきを表現する尺度を議論する．

今日の講義の目的と概要

目的

（1）データが入力されたら，とりあえずデータにどんな値がどれくらいかるかを調べる．この様子をあらわしたものを「分布」という．表現法として，度数分布表や図（ヒストグラム）がある．

（2）分布を説明するために，分布の中央あたりの値を「分布の中心」として「分布の代表値」とする．— 平均値と中央値．さらに，最小値と最大値，ボックスプロット（箱ひげ図）

（3）つぎに，データのばらつきを表現する尺度を議論する．. — 分散と標準偏差，平均偏差，範囲，四分位偏差

ヒストグラム

- 単峰型
  - 峰を中心に左右対称のもの
  - 左に偏ってもの（下段左）— 山の裾の部分に注目している！
  - 右に偏ったもの（下段右）
- 双峰型・多峰型 — 異種のデータが混在している場合が多い．

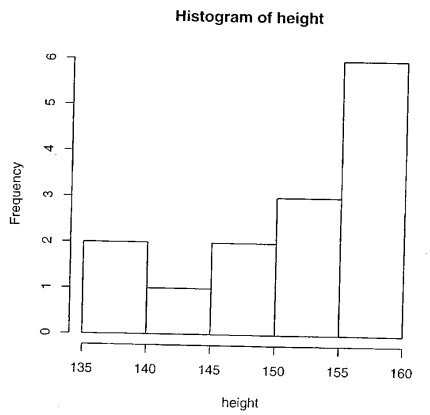

図 3.1　データ height のヒストグラム

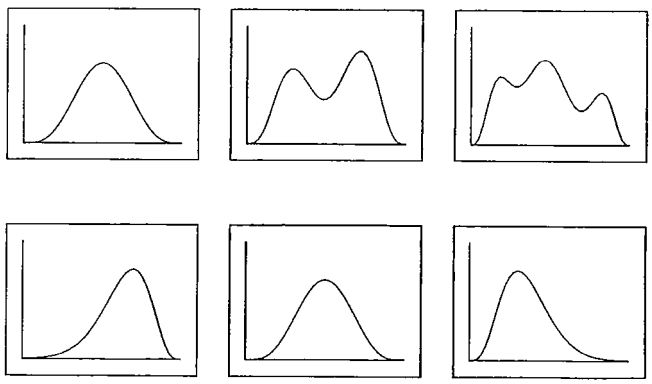

図 3.2　様々な分布形状

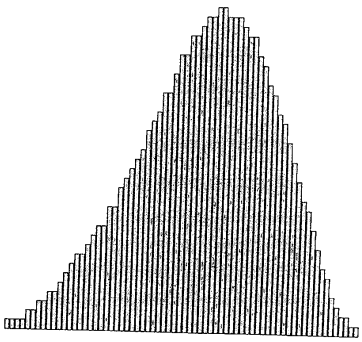

図 3.3　共通一次試験の総合得点の分布（昭和 55 年）

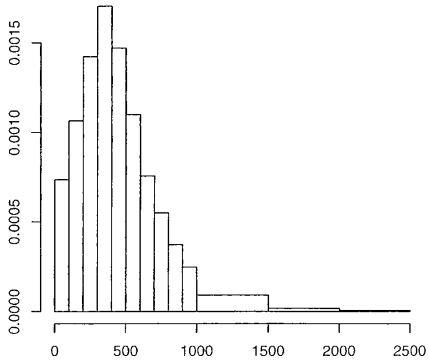

図 3.4　民間給与実態統計調査（平成 9 年）より

ヒストグラムの作成

ヒストグラムの作成

```
> a<-rnorm(100,10,10)
> hist(a)
> a<-round(rnorm(50,10,20))
> a
 [1]  -1  39  -4 -13  -2  40 -17   1  20  14
[11]   3  27   6  47  47 -23   6  -8  -5  16
[21] -17  -5  15   6   0   0  33 -16  23 -39
[31] 29  22   7 -10  26  16 -10  13  -1  -1
[41] 42 -23  -4   9  18 -34  14  13  -4   0
> hist(a)
```


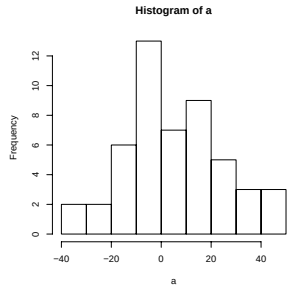

Figure 1: 作成されたヒストグラム

ヒストグラムの作成

```
>
> # コマンド hist のオプションを調べる
> ?hist
> # 割合で表示
> hist(a,freq=F)
>
```


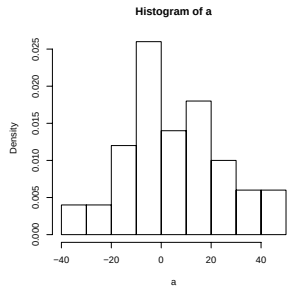

Figure 2: 作成されたヒストグラム

問題１

- 以下のようにデータを発生させ，そのヒストグラムを作成し，そこからわかることを述べよ．

ヒストグラムの作成

```
>
>
> a1<-round(rnorm(500,0,1))*5+あたたの誕生日
>
> a2<-round(rt(50,10)*5)+あたたの誕生日
>
```

オブジェクト a1 と a2 のヒストグラムを作成し，それを pdf file で保存する．ファイル名は 学籍番号下3桁-a1.pdf と 学籍番号下3桁-a2.pdf とせよ．さらに，ファイル ローマ字の名前-締め切り.txt

（例：mejiro-hanako-071012.txt)

代表値 — 平均値と中央値

n 個のデータの値を

$$x_1,\, x_2,\, \cdots,\, x_n$$

とする．

平均値　　$\bar{x}_n = \dfrac{1}{n}\sum_{i=1}^{n} x_i$

中央値　　$x_1,\, x_2,\, \cdots,\, x_n$ を昇順に並び替えたものを $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ としたとき，

$$Me = \begin{cases} x_{(\frac{n+1}{2})} & n \text{ が奇数} \\ \dfrac{x_{(n/2)}+x_{(n/2)+1}}{2} & n \text{ が偶数} \end{cases}$$

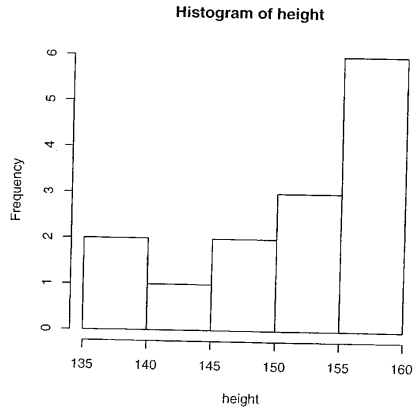

図 3.1　データ height のヒストグラム

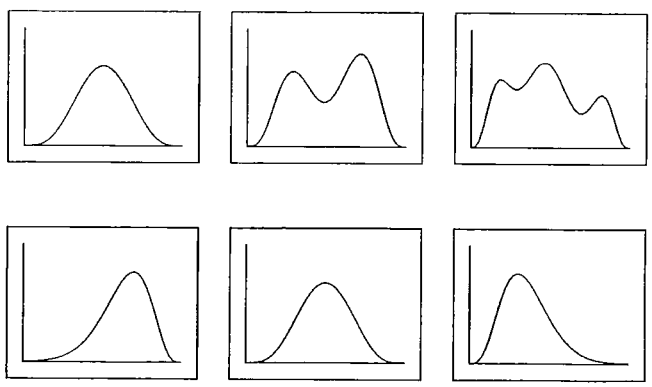

図 3.2　様々な分布形状

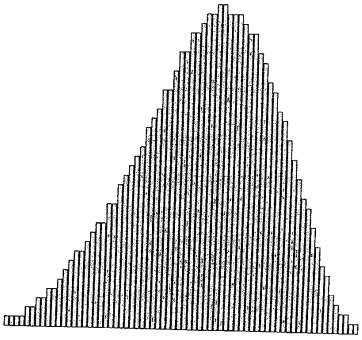

図 3.3 共通一次試験の総合得点の分布（昭和 55 年）

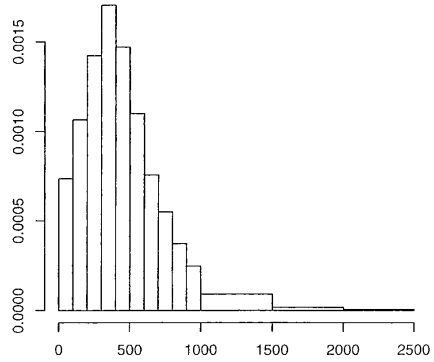

図 3.4 民間給与実態統計調査（平成 9 年）より