

データの縮約値・代表値

樋口さぶろお

龍谷大学工学部数理情報学科

確率統計☆演習 I L02(2015-09-29 Thu)

最終更新: Time-stamp: "2016-09-28 Wed 19:29 JST hig"

今日の目標

- 高校 数学 I 塚田確率統計 1.4
- データから 手で平均値, 離散データの最頻値, ヒストグラムの最頻値が求められる
- データから 手で中央値, 四分位数が求められる
- データから 手で箱ひげ図が描ける



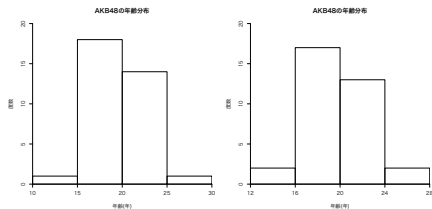
<http://hig3.net>

L01-Q1

Quiz 解答:度数分布表とヒストグラムを作ろう

階級 (歳)	度数	階級 (歳)	度数
10 より大きい 15 以下	1	12 より大きい 16 以下	2
15 より大きい 20 以下	18	16 より大きい 20 以下	17
20 より大きい 25 以下	14	20 より大きい 24 以下	13
25 より大きい 30 以下	1	24 より大きい 28 以下	2
計	34	計	34

このデータの場合はたまたま、以上未満でも同じ。



たまたま形が似たけど、階級の取り方でヒストグラムの形は変化する。

ここまで来たよ

1 略解:データの分布

2 データの縮約値・代表値

- 中央値と四分位値
- 最頻値と平均値
- 箱ひげ図

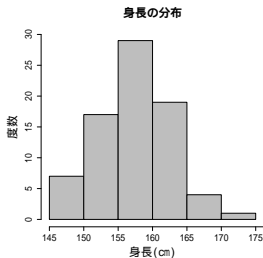
縮約値・代表値:データを1個の値で代表させたい!

塚田確率統計 1.4

縮約値=代表値某国民的アイドル集団の身長はだいたい 150cm? 170cm?

データ全体 148 152 ... 170

階級 (cm)	度数 f_j
145 より大きく 150 以下	7
150 より大きく 155 以下	17
155 より大きく 160 以下	29
160 より大きく 165 以下	19
165 より大きく 170 以下	4
170 より大きく 175 以下	1
合計	77



中央値 (median) と四分位数 (quartile)

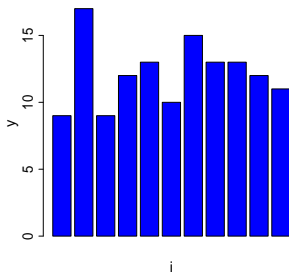
データを小さい順に並び替えたものを,

$y_1 \leq y_2 \leq \dots \leq y_n$ とする.

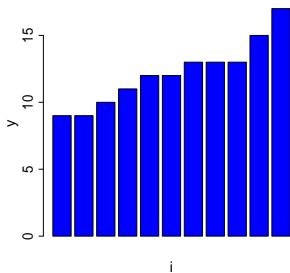
例

データ ($n = 11$): 9 17 9 12 13 10 15 13 13 12 11

i	1	2	3	4	5	6	7	8	9	10	11
y_i	9	9	10	11	12	12	13	13	13	15	17

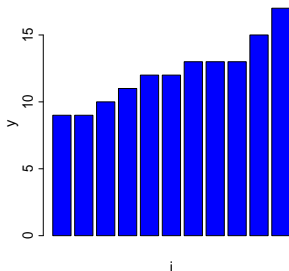


→ 順にならべる



四分位数の ABOUT な定義 塚田確率統計 1.5.1

- 最小値 $Q_0 = y_1 \approx y_{\frac{0}{4}n}$
- 第 1 四分位数 $Q_1 = y_{\frac{1}{4}n}$
- 第 2 四分位数 $Q_2 = y_{\frac{2}{4}n}$ = 中央値 塚田確率統計 1.4.1
- 第 3 四分位数 $Q_3 = y_{\frac{3}{4}n}$
- 最大値 $Q_4 = y_{\frac{4}{4}n}$

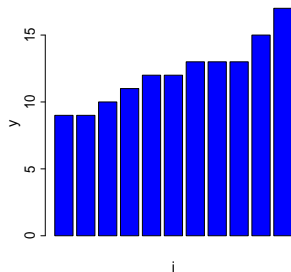
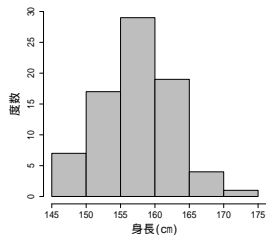
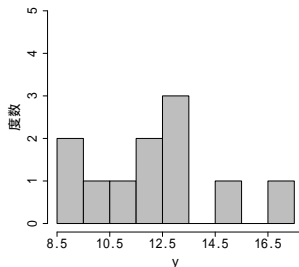


度数分布表からの中央値と四分位値の(だいたいの)求め方

階級値 = 階級の(上限値 + 下限値) / 2

階級 (cm)	階級値 $x_{(i)}$	度数 f_i
145 より大きく 150 以下	147.5	7
150 より大きく 155 以下		17
155 より大きく 160 以下		29
160 より大きく 165 以下		19
165 より大きく 170 以下		4
合計 n	—	77

中央値・四分位値のヒストグラムの意味

身長の分布yの分布

L02-Q1

Quiz(四分位値)

次のデータの四分位数 Q_1, Q_2, Q_3 を求めよう.

17 18 16 18 25 18 14 14 15

ここまで来たよ

1 略解:データの分布

2 データの縮約値・代表値

- 中央値と四分位値
- 最頻値と平均値
- 箱ひげ図

最頻値=mode

塚田確率統計 1.4.1

最頻値の定義

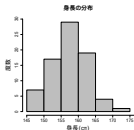
- 離散データの最頻値: '離散的な' データのとき いちばん多く繰り返し現れる値
- ヒストグラムの最頻値: '連続的または離散的な' データのとき 度数分布表/ヒストグラムで, 度数最大の階級の階級値

離散的な例 1(30 50 55 55 60 70 70 70 75 100) だと

ヒストグラムの最頻値の求め方

階級 (cm)	度数 f_i
145 より大きく 150 以下	7
150 より大きく 155 以下	17
155 より大きく 160 以下	29
160 より大きく 165 以下	19
165 より大きく 170 以下	4
170 より大きく 175 以下	1
合計	77

ヒストグラムの最頻値の意味



平均値=mean

平均値の定義 塚田確率統計 1.4.2

$$\text{平均値 } \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

\bar{x} のかわりに m , m_x などという記号もある.

例 1: 30 50 55 55 60 70 70 70 75 100 だと

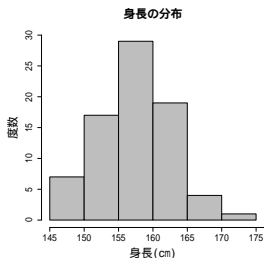
平均値より中央値のいい点

度数分布表からの平均値の(だいたいの)求め方

塚田確率統計 p.25

$$m \approx \frac{1}{n} \sum_{i=1}^k x_{(i)} f_i$$

階級内の値はぜんぶ階級値に等しい、と近似しちゃう。
平均値のヒストグラムの意味



$$x_G = \frac{\sum_i m_i x_i}{\sum_i m_i} \text{ で, } m_i = f_i.$$

力学

L02-Q2

Quiz(代表値)

次のデータを考える.

14cm, 14cm, 15cm, 16cm, 18cm, 18cm, 18cm, 25cm

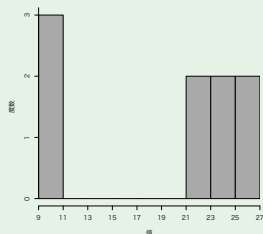
- ① 四分位数 Q_1, Q_2, Q_3 を求めよう.
- ② (離散データの) 最頻値を求めよう
- ③ 平均値を求めよう

L02-Q3

Quiz(平均値中央値最頻値)

次のヒストグラムから求めよう.

- ① 中央値
- ② (ヒストグラムの) 最頻値
- ③ 平均値



ここまで来たよ

1 略解:データの分布

2 データの縮約値・代表値

- 中央値と四分位値
- 最頻値と平均値
- 箱ひげ図

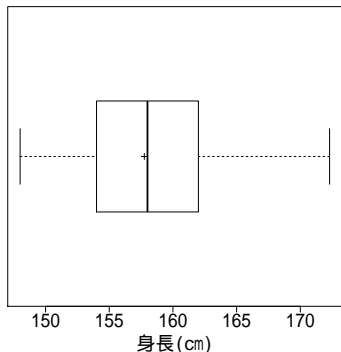
箱ひげ図 (Box Plot)

塚田確率統計 p.20

最小最大値 Q_0, Q_4 , 四分位点 Q_1, Q_2, Q_3

某アイドル集団の身長分布

某アイドル集団



箱ひげ図を描く手順高校 数学 I

- Q_0, Q_4 Q_1, Q_2, Q_3 と平均値 m を求める
- Q_2 に縦線をいれる
- Q_1, Q_3 を左右の端として箱を描く
- Q_0, Q_4 に短い縦線をいれ, 点線のひげで箱とつなぐ
- 平均値に + を 1 個描く

この他に「外れ値を○で描く」こともある。

いまの場合, 横軸:身長 (cm), 縦軸:意味なし,

L02-Q4

Quiz(四分位数と箱ひげ図)

下のチョコの重さのデータについて, 3つの四分位数と平均値を求め箱ひげ図を描こう.

7g 8g 10g 11g 12g 12g 12g 14g 15g

連絡

- 次回は 7-002 講義室
- 配布資料は 1-503 向かいの引出, <http://hig3.net> で再配布しています.
- 加減乗除と平方根 (ルート) の使える電卓持ってきてね. 関数電卓でなくてもいいです. 携帯電話の機能・アプリでもかまいません.
- Learn Math Moodle の予習復習問題で来週の trial に備えてね.
- 樋口オフィスアワー木 6 金昼 (1-502), Math ラウンジ月-木昼 (1-614)



[https://manaba.
ryukoku.ac.jp](https://manaba.ryukoku.ac.jp)