

箱ひげ図・分散の応用

樋口さぶろお

龍谷大学理工学部数理情報学科

確率統計☆演習 I L03(2014-10-10 Fri)

今日の目標

- データから箱ひげ図が手で描ける
- 箱ひげ図からヒストグラムが想像できる
- 線型変換のもとで平均値, 分散, 標準偏差を変換できる
- 偏差値が計算できる.



<http://hig3.net>

L02-S1

Quiz 解答:代表値

- ① $Q_2 = 17, Q_1 = 14.5, Q_3 = 18.$
- ② 最頻値は 18.
- ③ 平均値は $(14 + \dots + 25)/8 = 17.25.$

L02-S2

Quiz 解答:平均値中央値最頻値

- ① 22
- ② 10
- ③ 19.3

L02-S3

Quiz 解答:範囲

範囲は $Q_4 - Q_0 = 25 - 14 = 11$, 四分位範囲は $Q_3 - Q_1 = 18 - 14.5 = 3.5$, 標準偏差は $\left(\frac{((14 - 17.25)^2 + \dots + (25 - 17.25)^2)}{8}\right)^{1/2} = 3.34.$

ここまで来たよ

① 分散とその応用

- 分散の意味
- 平均値・分散・標準偏差の変換
- 変動係数
- 標準得点と偏差値

② 箱ひげ図とヒストグラム

- 箱ひげ図
- ヒストグラムと箱ひげ図の対応

分散の意味

分散が大きいとは?

分散が大きいとは「ばらつきが大きい」こと。

L03-Q1

Quiz(分散の意味)

あるクラスで行われたテストで、英語の平均点は 60 点、標準偏差 10 点。
数学の平均点は 60 点、標準偏差 20 点。

英語の 70 点と数学の 70 点、どちらのほうが価値ある?

- 1 たぶん英語のほうが価値ある
- 2 たぶん数学のほうが価値ある
- 3 どちらも同じ
- 4 これだけの情報ではまったくわからない
- 5 平均点が 60 点だと再テストがあるだろう

ここまで来たよ

① 分散とその応用

- 分散の意味
- 平均値・分散・標準偏差の変換
- 変動係数
- 標準得点と偏差値

② 箱ひげ図とヒストグラム

- 箱ひげ図
- ヒストグラムと箱ひげ図の対応

平均値・分散・標準偏差の変換

 x から y への変換

データ x_1, x_2, \dots, x_n , x の平均値 \bar{x} , 分散 s_x^2 , 標準偏差 s_x

データ y_1, y_2, \dots, y_n を $y_i = ax_i + b$ で作る. y の平均値 \bar{y} , 分散 s_y^2 , 標準偏差 s_y は?

a, b 定数.

例: 身長の変換 $y = 1.8(\text{m}) \leftarrow x = 80(\text{cm})$

$$y = ax + b,$$

平均値, 分散, 標準偏差の換算

 $y = ax + b$ のとき

- ① $\bar{y} = a\bar{x} + b$
- ② $s_y^2 = |a|^2 \times s_x^2$
- ③ $s_y = |a| \times s_x$

証明

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} \sum_{i=1}^n (ax_i + b) = \boxed{\phantom{a\bar{x} + b}} = a\bar{x} + b.$$

$$s_y^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{1}{n} \sum_{i=1}^n ((ax_i + b) - (a\bar{x} + b))^2 = \boxed{} = |a|^2 s_x^2$$

$$s_y = |a| s_x$$

ここまで来たよ

① 分散とその応用

- 分散の意味
- 平均値・分散・標準偏差の変換
- **変動係数**
- 標準得点と偏差値

② 箱ひげ図とヒストグラム

- 箱ひげ図
- ヒストグラムと箱ひげ図の対応

身長と靴のサイズじゃ標準偏差の意味が違う!

クラス内で、身長の範囲 (range) は 50cm くらいだけど、靴のサイズの範囲は 5cm くらい。

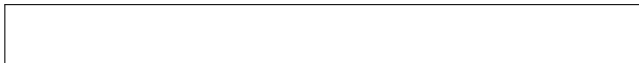
標準偏差が大きい = いろんな体格の人がいる

みたいに思いたいけど、身長と靴のサイズじゃ標準偏差の意味が違う。

変動係数 (coefficient of variation)

$$(\text{データ } x \text{ 全体の}) \text{ 変動係数} = \frac{s_x}{\bar{x}}$$

- これは**無次元の数**。すなわち単位がない量。



- $\frac{\text{分散}}{\text{平均値}}$ だと無次元の数にはならない。

L03-Q2

Quiz(次元のある数ない数)

次のうち次元のない数はどれ(とどれ)?

- ① 国内総生産 (GDP)
- ② 年間降水量
- ③ 物価上昇率
- ④ (道路の) 制限速度
- ⑤ 食糧自給率

L03-Q3

Quiz(変動係数)

次の2組のデータは、それぞれ100点満点、1000点満点のテストの点数。成績のばらつきが大きいのはどっち？ 変動係数を計算して答えよう。

- 805 780 805 795 795
- 87 93 89 91 90

ここまで来たよ

① 分散とその応用

- 分散の意味
- 平均値・分散・標準偏差の変換
- 変動係数
- 標準得点と偏差値

② 箱ひげ図とヒストグラム

- 箱ひげ図
- ヒストグラムと箱ひげ図の対応

標準得点

標準得点 (standard score)

$$\text{(値 } x_i \text{ の) 標準得点 } z_i = \frac{x_i - \bar{x}}{s_x}$$

- 平均値から、上下どちらに、標準偏差の何倍離れているかを表す値.
- z -得点 (z -score) などともいう.
- 標準得点は無次元の数.

標準得点の性質

標準得点 z の性質

- $\bar{z} = \square$
- $s_z^2 = \square$, $s_z = \square$
- z は無次元の数

なぜなら… いま

$$\bar{z} = a\bar{x} + b = \frac{1}{s_x} \cdot \bar{x} - \frac{\bar{x}}{s_x} = 0$$

$$s_z = |a|s_x = \left| \frac{1}{s_x} \right| s_x.$$

偏差値

0-100 の範囲の値をとるデータ (テストの点数や成績?) に使われる。
受験者 1 人 1 人の成績が, 平均値から上, または下に離れている程度を見られる。

偏差値

$$\begin{aligned}(\text{値 } x_i \text{ の) 偏差値} &= 10z_i + 50 \\ &= \frac{x_i - \bar{x}}{s_x} \times 10 + 50.\end{aligned}$$

$$a = \boxed{}, b = \boxed{}$$

- 異なるテスト, クラスでも比べられる。
- 偏差値の平均値は $\boxed{}$, 偏差値の標準偏差は $\boxed{}$
- 偏差値はまあ '無次元の数'(1000 点満点と 100 点満点を比較可能)

L03-Q4

Quiz(偏差値)

(学力) 偏差値について、次のうち正しいのはどれ(とどれ)?

- ① 偏差値の最低値は 0 である
- ② 偏差値の最高値は 75 である
- ③ 平均点 (をとった人) の偏差値は 50 である
- ④ 100 点のテストで満点を取った場合の偏差値は、他の人の成績しだいである
- ⑤ 偏差値 50 の人の順位は上から 1/2 程度である
- ⑥ 偏差値 60 の人の順位は上から 15% 程度である.

L03-Q5

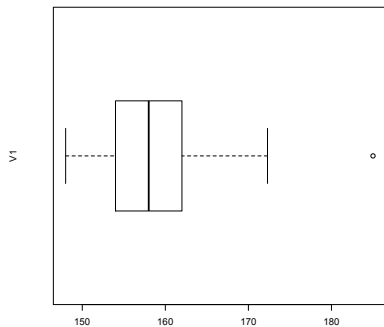
Quiz(標準得点と偏差値)

データ 87, 93, 89, 91, 90 で, 87 の標準得点と偏差値を求めよう.

ここまで来たよ

- 1 分散とその応用
 - 分散の意味
 - 平均値・分散・標準偏差の変換
 - 変動係数
 - 標準得点と偏差値
- 2 箱ひげ図とヒストグラム
 - 箱ひげ図
 - ヒストグラムと箱ひげ図の対応

箱ひげ図 (Box Plot)



横軸:身長 (cm), 縦軸:意味なし

外れ値 (outlier)

四分位点 Q_1, Q_2, Q_3 , 四分位範囲 $IQR=Q_3 - Q_1$

外れ値 (outlier)

- Q_1 の下に, IQR の 1.5 倍よりも離れたデータ
- Q_3 の上に, IQR の 1.5 倍よりも離れたデータ

(基本) 箱ひげ図の描き方

高校 数学 I

箱ひげ図を描く手順

- 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
- 平均値に $+$ を 1 個描く
- 外れ値を除いた最大値, 最小値までひげを描く
- 外れ値を \circ で描く

赤字部分を省略すると, 基本箱ひげ図. 高校の数学 I はそのレベル.

L03-Q6

Quiz(ヒストグラムと箱ひげ図を描こう)

次のデータから作ろう.

- ① 箱ひげ図
- ② 度数分布表
- ③ ヒストグラム

14 14 15 16 18 18 18 25

L03-Q7

Quiz(箱ひげ図)

下の 1 変量データについて, 3 つの四分位点を求め, 箱ひげ図を描こう.

X

2

8

10

11

12

12

12

14

15

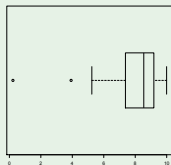
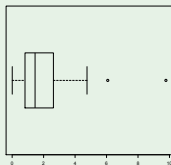
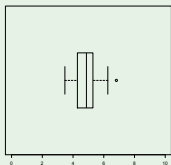
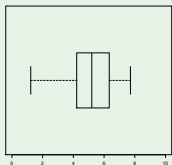
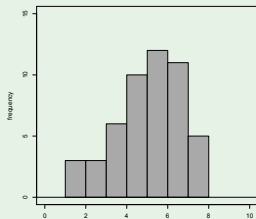
ここまで来たよ

- ① 分散とその応用
 - 分散の意味
 - 平均値・分散・標準偏差の変換
 - 変動係数
 - 標準得点と偏差値

- ② 箱ひげ図とヒストグラム
 - 箱ひげ図
 - ヒストグラムと箱ひげ図の対応

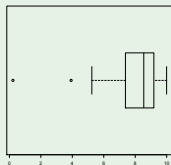
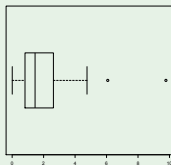
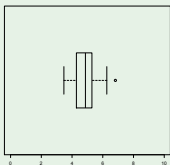
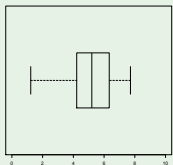
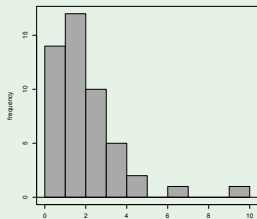
Quiz(ヒストグラムと箱ひげ図)

このヒストグラムに対応する箱ひげ図はどれ?



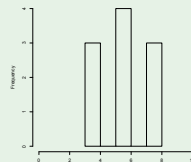
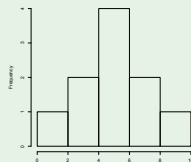
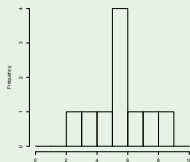
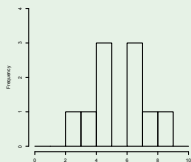
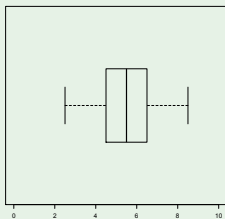
Quiz(ヒストグラムと箱ひげ図)

このヒストグラムに対応する箱ひげ図はどれ?



Quiz(ヒストグラムと箱ひげ図の対応)

この箱ひげ図に対応するヒストグラムはどれ?



連絡

- 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.
- Quiz の略解は授業終了後に <http://hig3.net> で配布しています.
- 次回からは, 加減乗除と平方根 (ルート) の使える電卓持って来てね. 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.
- 最初のころはいろいろ変更あるかも. メールに注意.
- 週のタイムラインで見たように, 予習問題を RaMMoodle に金 15:30 までに公開. 翌週水 09:20 までにやってね. それまで何回でも「受験」できます. 最後の受験が点数になります.

来週の非相談非参照 Quiz

- 箱ひげ図を描こう
- 偏差値を求めよう