

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

使える統計! L04(2013-10-23 Wed)

今日の目標

- ① $X + b$, $X \times a$, の平均値・分散・標準偏差が, X の平均値・分散・標準偏差から求められる.
- ② データ全体の変動係数が求められる
- ③ 個々のデータの標準得点が求められる
- ④ 個々のデータの偏差値が求められる

<http://hig3.net>

L03-S6

Quiz 解答:平均値と分散

- $\frac{1}{5}(9 + 10 + 12 + 12 + 12) = 11$. よって, 平均値は 1011.
- 分散は $\frac{1}{5}((9 - 11)^2 + \dots + (12 - 11)^2) = \frac{1}{5}(4 + 1 + 1 + 1 + 1) = 1.6$.
標準偏差は, $\sqrt{1.6} = 1.26\dots$.

ここまで来たよ

- 1 復習:データのばらつきを表す値
 - 平均値・分散・標準偏差の変換

- 2 分散の応用
 - 変動係数
 - 標準得点
 - 偏差値

復習: 平均値, 分散, 標準偏差

復習

- 平均値 $= \frac{148+148.5+\dots+172.3}{77} = 158\text{cm}$
- 分散 $= \frac{(148-158)^2+(148.5-158)^2+\dots+(172.3-158)^2}{77} = 26.0 \text{ cm}^2$
- 標準偏差 $= \sqrt{\text{分散}} = 5.1 \text{ cm.}$

Q1

Quiz(平均値と分散と標準偏差)

データ 87, 93, 89, 91, 90 の, 平均値, 分散, 標準偏差を求めよう.

意味

平均値・分散・標準偏差の変換

$X \times a, X + b$ の平均値・分散・標準偏差は, X の平均値・分散・標準偏差からわかっちゃう!

データ	平均値		分散	標準偏差	
X	X の平均値		X の分散		X の標準偏差
$X + b$	X の平均値	$+b$	X の分散		X の標準偏差
$X \times a$	X の平均値	$\times a$	X の分散	$\times a^2$	X の標準偏差 $\times a$
$X \times a + b$	X の平均値	$\times a + b$	X の分散	$\times a^2$	X の標準偏差 $\times a$

定数を加える

100cm を除いた部分だけで考えると? $b = +100$. $X = 48, 48.5, \dots$

$$(X + b) \text{ の平均値} = X \text{ の平均値} + b$$

$$\begin{aligned} & \frac{[48 + 100] + [48.5 + 100] + \dots + [72.3 + 100]}{77} \\ &= \frac{48 + 48.5 + \dots + 72.3}{77} + 100 = (58 + 100)\text{cm} \end{aligned}$$

$$(X + b) \text{ の分散} = X \text{ の分散}$$

$$\begin{aligned} & \frac{[(48 + 100) - (58 + 100)]^2 + \dots + [(72.3 + 100) - (58 + 100)]^2}{77} \\ &= \frac{[48 - 58]^2 + \dots + [72.3 - 58]^2}{77} = 26.0\text{cm}^2 \end{aligned}$$

$$(X + b) \text{ の標準偏差} = X \text{ の標準偏差}$$

定数をかける

cm じゃなく m で考えると? $a = 0.01$. $X = 148, 148.5, \dots$

$$(X \times a) \text{ の平均値} = X \text{ の平均値} \times a$$

$$\begin{aligned} & \frac{(148 \times 0.01) + (148.5 \times 0.01) + \dots + (172.3 \times 0.01)}{77} \\ &= \frac{148 + 148.5 + \dots + 172.3}{77} \times 0.01 = 158 \times 0.01 \text{ m} \end{aligned}$$

$$(X \times a) \text{ の分散} = X \text{ の分散} \times a^2$$

$$\begin{aligned} & \frac{[(148 \times 0.01) - (158 \times 0.01)]^2 + \dots + [(172.3 \times 0.01) - (158 \times 0.01)]^2}{77} \\ &= \frac{[148 - 158]^2 + \dots + [172.3 - 158]^2}{77} \times 0.01^2 = 26.0 \times 0.01^2 \text{ m}^2 \end{aligned}$$

$$(X \times a) \text{ の標準偏差} = X \text{ の標準偏差} \times a = \sqrt{26.0 \cdot 0.01^2} = \sqrt{26.0} \cdot 0.01 \text{ m.}$$

ここまで来たよ

- 1 復習:データのばらつきを表す値
 - 平均値・分散・標準偏差の変換

- 2 分散の応用
 - 変動係数
 - 標準得点
 - 偏差値

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)

$$(\text{データ全体の}) \text{ 変動係数} = \frac{\text{標準偏差}}{\text{平均値}}$$

- 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.

Q1

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

Q2

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 復習:データのばらつきを表す値
 - 平均値・分散・標準偏差の変換

- 2 分散の応用
 - 変動係数
 - 標準得点
 - 偏差値

標準得点

標準得点 (standard score)

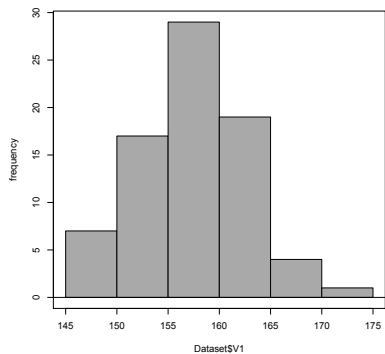
$$(\text{データ 1 個の}) \text{標準得点} = \frac{\text{データの値} - \text{平均値}}{\text{標準偏差}}$$

- 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.

Q3

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.



標準得点の性質

標準得点の性質

- 標準得点の平均値は
- 標準得点の分散は , 標準偏差は
- 標準得点は無次元の数

なぜなら…平均値と標準偏差の変換を思いだそう

元のデータを X , 標準偏差を s とすると, 標準得点 $Z = (X - m) \times \frac{1}{s}$.

$-m$ ずらす \rightarrow 平均値 , 標準偏差

$\frac{1}{s}$ 倍する \rightarrow 平均値 , 標準偏差

- $(X - m)$ の平均値 = X の平均値 $-m = 0$.
- $Z = (X - m)/s$ の平均値 = $(X - m)$ の平均値 $\times \frac{1}{s} = 0/0 = 0$.
- $(X - m)$ の標準偏差 = X の標準偏差 = s .
- $Z = (X - m)/m$ の標準偏差 = $(X - m)$ の標準偏差 $\times \frac{1}{s} = \frac{s}{s} = 1$.

ここまで来たよ

- 1 復習:データのばらつきを表す値
 - 平均値・分散・標準偏差の変換

- 2 分散の応用
 - 変動係数
 - 標準得点
 - 偏差値

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

$$\begin{aligned}
 (\text{データ 1 個の}) \text{偏差値} &= \text{標準得点} \times 10 + 50 \\
 &= \frac{\text{データの値} - \text{平均値}}{\text{標準偏差}} \times 10 + 50
 \end{aligned}$$

- 異なるテスト、クラスでも比べられる。
- 偏差値の平均は
- 偏差値の標準偏差は
- 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)

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。

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。

連絡

- 各追加 2 ピーナッツ=計 4 ピーナッツになる課題.

提出: 2013-10-16 水 or 2013-10-23 水 or 2013-10-30 水の授業

- ① 龍谷大学 e ラーニングシステム

<https://moodle.media.ryukoku.ac.jp/> → リメディアルコース統計学 → 第 1 章修了テスト.

- ② 龍谷大学 e ラーニングシステム

<https://moodle.media.ryukoku.ac.jp/> → リメディアルコース統計学 → 第 2 章修了テスト.

このサイトには, <http://hig3.net> → 龍大 Moodle, や Info Seta → e ラーニングサイト → 新 e ラーニングシステム でも到達できます. すべてを送信して終了する → レビューを終了する の後に出る, 「あなたの前回受験の要約」 ページ (下) を印刷して, 紙で提出. (スクリーンショットを課題にアップロードしてもいい)

- 今週は授業内で紙を 1 枚提出 (+修了テストも提出できます)
- 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.

あなたは樋口 三郎としてログインしています (ログ

フォーメーション

読み取れるように 修了テスト 第2章

評定方法: 最高評点

あなたの前回受験の要約

受験	状態	評点 / 100.00	レビュー
1	終了 送信日時 2013年 10月 8日(火曜日) 17:39	0.00	レビュー

最高評点: 0.00 / 100.00

もう一度受験する

クリッカー学籍番号送信の方法

- t012345 → 1012345
- c012345 → 4012345
- w012345 → 7012345

学期半ば授業アンケート

いま, スマートフォンで,

<http://hig3.net> or

→ 龍大 Moodle → 生活の中の

統計技術 → 2013-10-23 Wed → 学期半ば授業アンケート

または, あとで PC で, ポータル → 新 e ラーニングシステム (Ver2.4) →
生活の中の統計技術 → 2013-10-23 Wed → 学期半ば授業アンケート