

統計の分析と利用

1. データとその扱い

堀田 敬介

- 1-1. 一次元のデータ
 - 度数分布・ヒストグラム・幹葉プロット・箱ひげ図
 - 代表値と散らばり
 - データの尺度
- 2-2. 二次元のデータ
 - 散布図・○○・クロス集計
 - 二次元データの関係: 相関係数・相関比・連関係数

2010/9/24, Fri. ~

1-1. 一次元のデータ

- 度数分布
- ヒストグラム
- 幹葉プロット
- 箱ひげ図

$$\mathbf{x} = (\overset{n\text{個}}{x_1, x_2, \dots, x_n})$$

$$x_1, x_2, x_3, x_4, x_5, x_6$$

x	11	9	-3	14	5	23
---	----	---	----	----	---	----

$(n = 6)$

度数分布

週末はどのぐらいお客さんが来てくれたの？



○ データ [土日の来店客数の1年間のデータ]

292	373	282	251	322	392	366	300	226	314
325	300	356	319	213	229	244	347	283	372
253	317	306	390	287	268	257	247	318	232
306	274	231	370	275	186	327	297	260	300
285	365	272	335	167	289	352	321	341	313
319	351	299	327	405	259	376	360	259	252
339	301	337	229	244	279	243	272	211	303
316	311	287	248	199	274	286	367	317	311
434	346	329	338	319	244	329	329	274	262
288	306	189	248	344	262	385	302	366	249
250	297	292	261						

$x = (x_1, x_2, \dots, x_{104}) \quad (n = 104)$

データが多すぎて**全体の傾向**がよくわからない！



度数分布

○ 度数分布表[土日の来店客数の1年間のデータ]

階級
(class)
階級数:10
階級幅:30

階級値
各階級の上限・下限値の
中間値
[例] 344.5 ← 330-359
[例] 345 ← 330-360

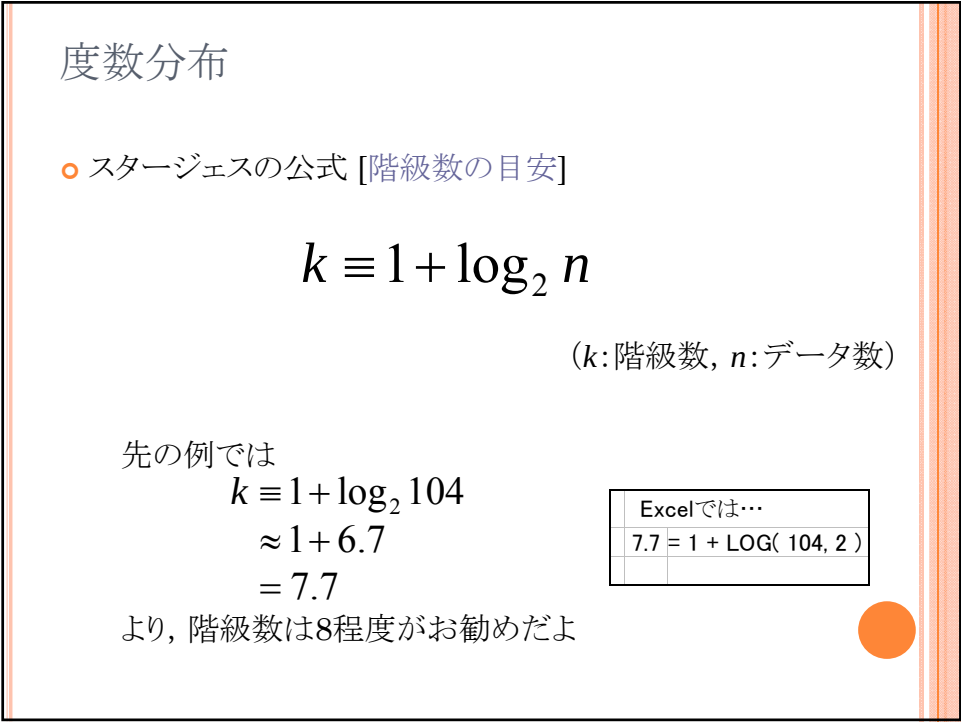
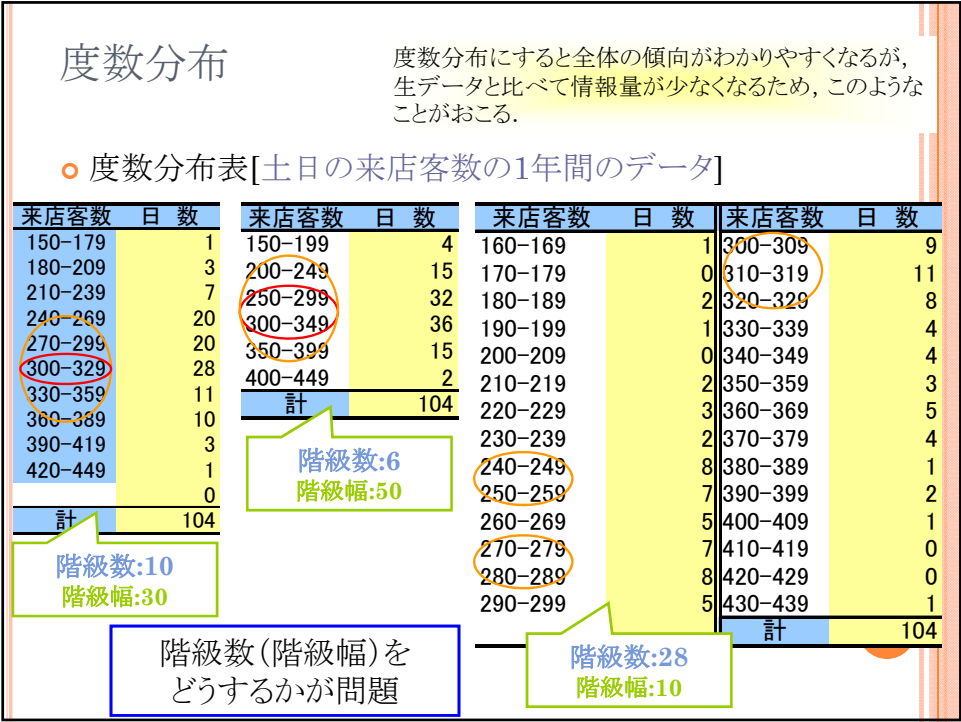
来店客数	日数
150-179	1
180-209	3
210-239	7
240-269	20
270-299	20
300-329	28
330-359	11
360-389	10
390-419	3
420-449	1
計	104

度数
(frequency)

なるほど、週末の来店客数はだいたいこのぐらいのことが多いんだ



全体の傾向がよくわかる！



度数分布

○ 階級数8(階級幅38)で書くと...

来店客数	日数	相対度数
150-187	2	1.9
188-225	4	3.8
226-263	24	23.1
264-301	25	24.0
302-339	28	26.9
340-377	16	15.4
378-415	4	3.8
416-453	1	1.0
計	104	100.0

なるほど、週末の来店客数の全体傾向はだいたいわかったぞ



でも、度数の多い階級は全体からみてどのぐらいの割合なの？

相対度数
(relative frequency)

度数分布

○ 度数分布表[相対度数]

来店客数	日数	相対度数	来店客数	日数	相対度数
150-179	1	1.0	150-179	2	1.0
180-209	3	2.9	180-209	6	3.0
210-239	7	6.7	210-239	21	10.5
240-269	20	19.2	240-269	24	12.0
270-299	20	19.2	270-299	40	20.0
300-329	28	26.9	300-329	54	27.0
330-359	11	10.6	330-359	32	16.0
360-389	10	9.6	360-389	13	6.5
390-419	3	2.9	390-419	6	3.0
420-449	1	1.0	420-449	2	1.0
計	104	100	計	200	100.0

Bさんのお店と比べて、うちのお客さんの来店傾向はどうなのか比較したいな...



データ数が異なる2つのグループの比較ができる



度数分布

- 累積度数分布表[累積度数, 累積相対度数]

来店客数	日数	相対度数	累積度数	累積相対度数
150-179	1	1.0	1	1.0
180-209	3	2.9	4	3.8
210-239	7	6.7	11	10.6
240-269	20	19.2	31	29.8
270-299	20	19.2	51	49.0
300-329	28	26.9	79	76.0
330-359	11	10.6	90	86.5
360-389	10	9.6	100	96.2
390-419	3	2.9	103	99.0
420-449	1	1.0	104	100.0
計	104	100.0		

累積度数
(cumulative frequency)

累積相対度数
(cumulative relative frequency)

演習1-1:度数分布

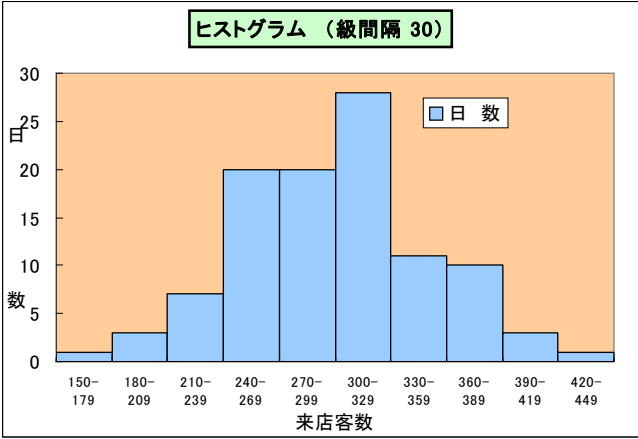
- 以下のデータの度数分布を作れ.

35	35	35	40
30	40	35	15
15	40	40	15
50	50	25	30
35	30	15	40



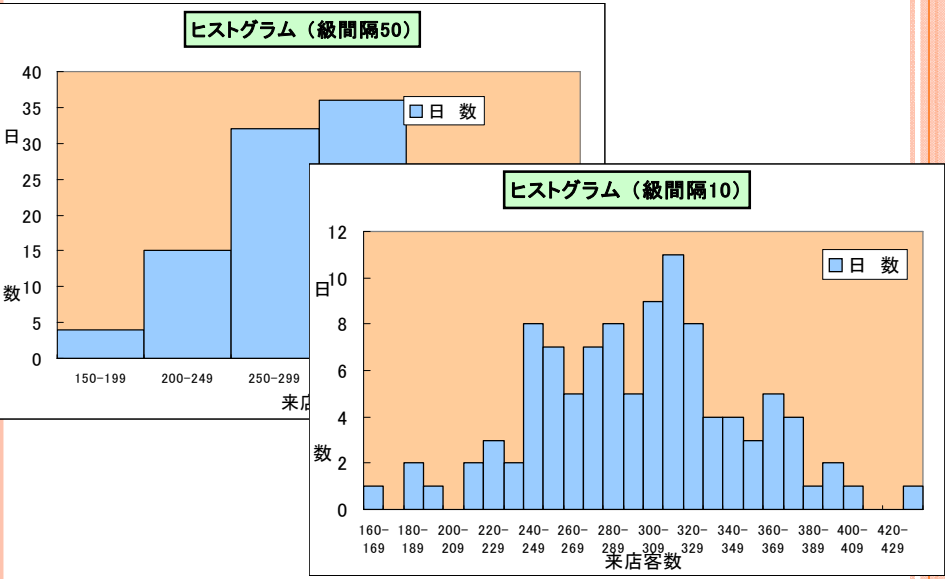
ヒストグラム

○ ヒストグラム(histogram)・柱状グラフ



ヒストグラム

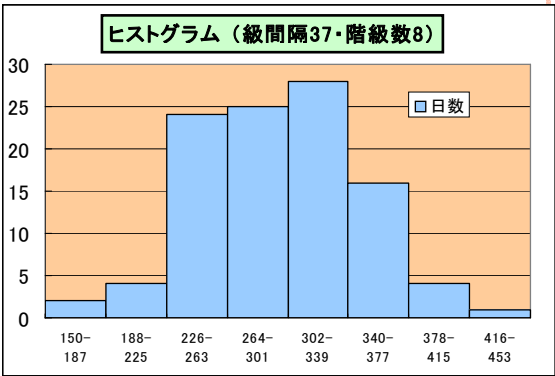
○ ヒストグラム(histogram)・柱状グラフ



度数分布

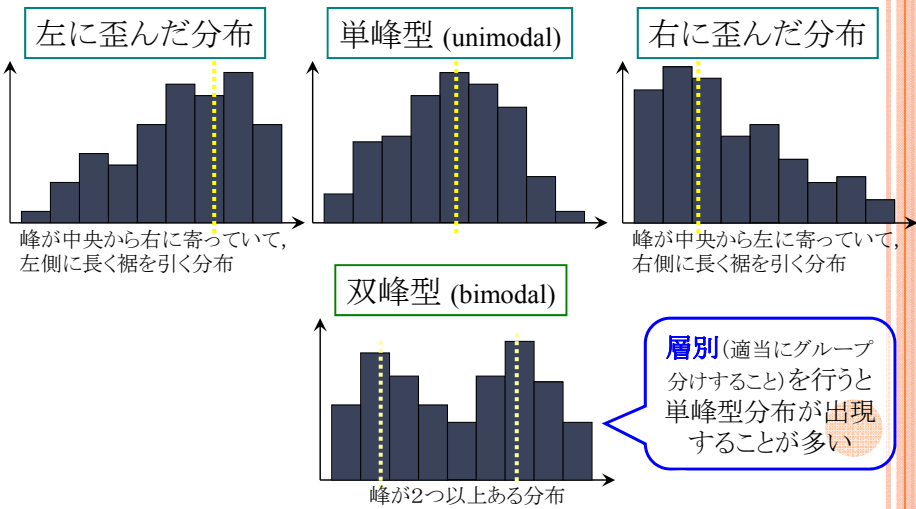
○ 階級数8で書くと...

来店客数	日数
150-187	2
188-225	4
226-263	24
264-301	25
302-339	28
340-377	16
378-415	4
416-453	1
計	104



ヒストグラム

○ ヒストグラムの形状



その他の手法1

幹葉プロットがヒストグラムより優れているのはどんなところ？

○ 幹葉プロット, ステムプロット (stem-and-leaf diagram[plot])

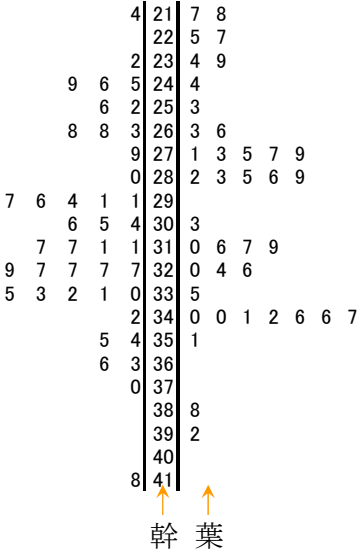
- 野球選手の打率一覧

○ Aチーム

0.275	0.347	0.266	0.263
0.271	0.225	0.283	0.324
0.286	0.351	0.346	0.342
0.388	0.319	0.303	0.279
0.217	0.273	0.244	0.234
0.277	0.392	0.326	0.32
0.282	0.289	0.218	0.285
0.316	0.335	0.34	0.31
0.346	0.239	0.127	0.263
0.317	0.341	0.34	0.253

■ Bチーム

0.317	0.327	0.37	0.355
0.291	0.28	0.297	0.311
0.317	0.306	0.245	0.366
0.232	0.342	0.335	0.263
0.304	0.311	0.294	0.214
0.327	0.327	0.252	0.331
0.268	0.291	0.279	0.296
0.363	0.33	0.329	0.246
0.354	0.249	0.332	0.333
0.256	0.418	0.268	0.305



その他の手法2

○ 箱ひげ図, 箱型図 (box plot)

- 野球選手の打率一覧

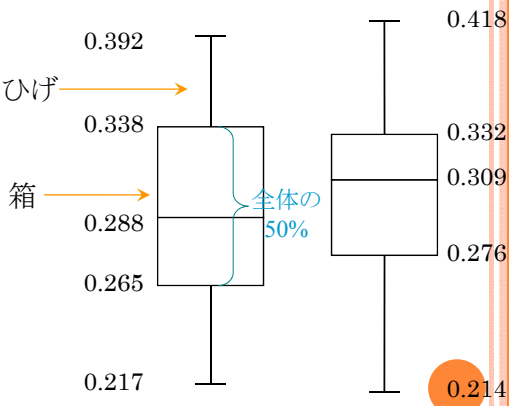
○ Aチーム

0.275	0.347	0.266	0.263
0.271	0.225	0.283	0.324
0.286	0.351	0.346	0.342
0.388	0.319	0.303	0.279
0.217	0.273	0.244	0.234
0.277	0.392	0.326	0.32
0.282	0.289	0.218	0.285
0.316	0.335	0.34	0.31
0.346	0.239	0.127	0.263
0.317	0.341	0.34	0.253

■ Bチーム

0.317	0.327	0.37	0.355
0.291	0.28	0.297	0.311
0.317	0.306	0.245	0.366
0.232	0.342	0.335	0.263
0.304	0.311	0.294	0.214
0.327	0.327	0.252	0.331
0.268	0.291	0.279	0.296
0.363	0.33	0.329	0.246
0.354	0.249	0.332	0.333
0.256	0.418	0.268	0.305

[Aチーム]	[Bチーム]
max.0.392	0.418 max.
Q ₃ 0.338	0.332 Q ₃
med.0.288	0.309 med.
Q ₁ 0.265	0.276 Q ₁
min. 0.217	0.214 min.



注: ひげの上端・下端は, 必ず max, min を使うわけではない.
r:=q3-q1 としたとき, 上端は区間(q3, q3+1.5r]内の最大値,
下端は区間[q1-1.5r, q1]内の最小値を用いる, など.

演習1-2: 幹葉プロット, 箱ひげ図

- 男女20人の身長データのデータがある。
 - 男女それぞれのデータについて, 10の位までを幹, 1の位を葉として幹葉プロットを描け.
 - 男女それぞれのデータについて, 箱ひげ図を描け.

男		女	
167	176	165	145
157	155	155	162
172	178	159	162
183	178	155	159
182	181	167	159
187	188	160	162
148	159	175	162
168	173	157	177
181	177	150	166
159	169	149	168



1-1. 一次元のデータ

- データの代表値
 - 算術平均
 - 中央値
 - 最頻値
- データの代表値(その他)
 - 四分位点
 - ミッド・レンジ
 - 幾何平均, 調和平均
 - 対数平均, identric平均

$$x = (x_1, x_2, \dots, x_n)$$

$$x_1, x_2, x_3, x_4, x_5, x_6$$

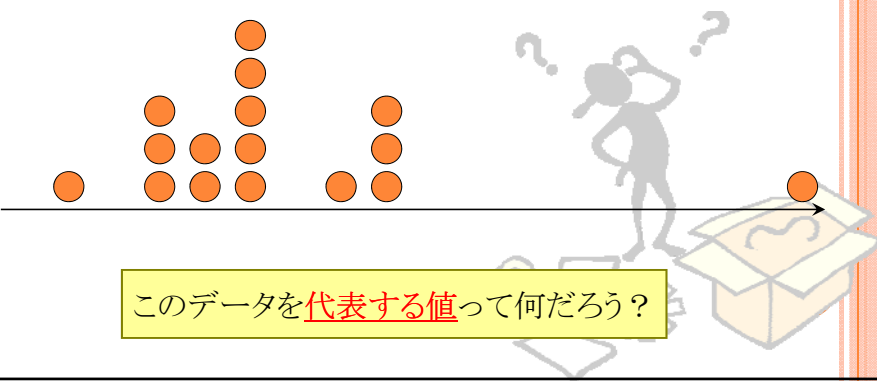
x	11	9	-3	14	5	23
---	----	---	----	----	---	----

$$(n = 6)$$

データの代表値を考える

■ 例:16個のデータ

X	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
データ	10	7	3	5	7	5	10	9	6	7	50	7	5	7	6	10



代表値 AVERAGES

○算術平均(相加平均) arithmetic mean

X	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
データ	10	7	3	5	7	5	10	9	6	7	50	7	5	7	6	10

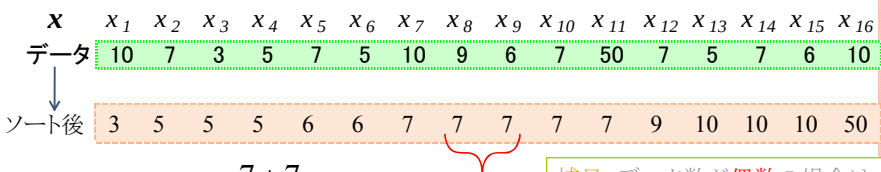
→ $\bar{x} = \frac{1}{16}(10 + 7 + \dots + 10) = 9.625$



代表値 AVERAGES

中央値 median

- データをソートして, ちょうど真ん中にある値



補足:ソート sort とは?
データを値の小さい(大きい)順に並べ替えること

→ $x_{\text{med}} = \frac{7+7}{2} = 7$

補足:データ数が偶数の場合は, 中央値は真ん中2つの算術平均

最頻値 mode

- データの中で最も頻繁に出てくる値

→ $x_{\text{mode}} = 7$

補足:最も頻繁に出てくる値がない場合は最頻値はなし

代表値 AVERAGES

中央値や最頻値は何故必要なのか?

- 例:年収(単位:万円)の代表値は?

700 500 1000 800 5000 700 300 800 700 800

□ 算術平均

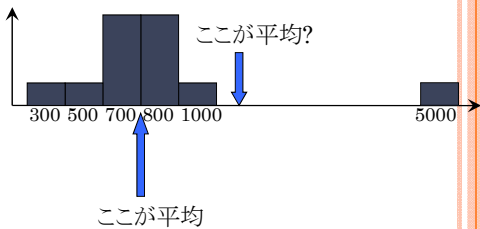
■ 1130万円

□ 中央値

■ $(700+800) / 2 = 750$ 万円

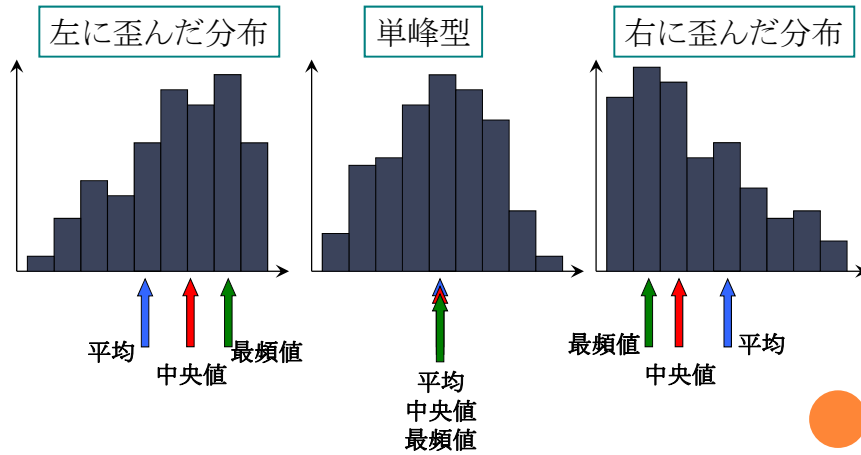
□ 最頻値

■ 700万円, 800万円



代表値 AVERAGES

○算術平均, 中央値, 最頻値の関係



代表値 AVERAGES

○幾何平均 geometric mean

補足: 対数を利用すると計算が楽になる

$$\log x_G = \log \sqrt[n]{x_1 \times \dots \times x_n} \\ = \frac{\log x_1 + \dots + \log x_n}{n}$$

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
データ	10	7	3	5	7	5	10	9	6	7	50	7	5	7	6	10

$$\rightarrow x_G = \sqrt[16]{10 \times 7 \times 3 \times 5 \times \dots \times 10} \approx 7.51$$

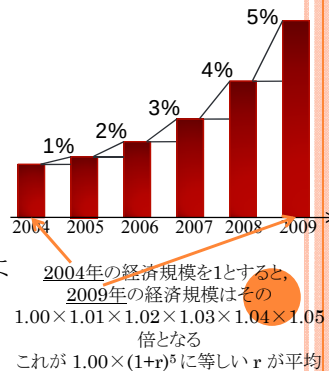
☆どんなときに幾何平均が役に立つ？

例題: 次の表から平均経済成長率を求めよ

年度	2005	2006	2007	2008	2009
経済成長率	1%	2%	3%	4%	5%

$$\text{答えは } \bar{x} = \frac{1+2+3+4+5}{5} = 3 \rightarrow 3\% \text{ じゃないよ}$$

$$\text{答えは } x_G = \sqrt[5]{1.01 \times 1.02 \times 1.03 \times 1.04 \times 1.05} \\ \approx 1.029 \rightarrow 2.9\% \text{ だよ}$$



代表値 AVERAGES

○調和平均 harmonic mean

x	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈	x ₉	x ₁₀	x ₁₁	x ₁₂	x ₁₃	x ₁₄	x ₁₅	x ₁₆
データ	10	7	3	5	7	5	10	9	6	7	50	7	5	7	6	10

→ $x_H = \frac{1}{\frac{1}{16}\left(\frac{1}{10} + \frac{1}{7} + \dots + \frac{1}{10}\right)} \approx 6.63$

☆どんなときに調和平均が役に立つ？

例題: 行き時速25km, 帰り時速15kmで走った車の平均速度を求めよ



答えは $\bar{x} = \frac{25+15}{2} = 20 \rightarrow 20\text{km/h}$ じゃないよ

答えは $x_H = \frac{1}{\frac{1}{2}\left(\frac{1}{15} + \frac{1}{25}\right)} = 18.75 \rightarrow 18.75\text{km/h}$ だよ

COFFEE BREAK

○和積の記号

- 和を表す記号: Σ (しぐま)

$$\sum_{i=1}^n x_i = x_1 + \dots + x_n$$

x_i を i を 1 から n まで動かして足す

- 積を表す記号: Π (ぱい)

$$\prod_{i=1}^n x_i = x_1 \times \dots \times x_n$$

x_i を i を 1 から n まで動かして掛ける

使用例)

$$\sum_{i=1}^4 x_i = x_1 + x_2 + x_3 + x_4$$

$$\sum_{k=1}^5 k = 1 + 2 + 3 + 4 + 5$$

$$\sum_{j=2}^4 5j = 5 \cdot 2 + 5 \cdot 3 + 5 \cdot 4$$

$$\frac{1}{n} \sum_{i=1}^n y_i = \frac{1}{n} (y_1 + y_2 + \dots + y_n)$$

$$\prod_{t=1}^6 t = 1 \times 2 \times 3 \times 4 \times 5 \times 6$$

COFFEE BREAK

○ 記号を用いた平均の定義

● 算術平均

$$\bar{x} \equiv \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + \cdots + x_n}{n}$$

● 幾何平均

$$x_G = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 \times \cdots \times x_n}$$

幾何平均
=
n個の積のn乗根

● 調和平均

$$x_H = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x_i}} = \frac{1}{\frac{1}{n} \left(\frac{1}{x_1} + \cdots + \frac{1}{x_n} \right)}$$

調和平均
=
逆数の算術平均
の
逆数

代表値 AVERAGES

○ 四分位点 quartile

- データをソートし、4等分したときの3つの分割点の値

- Q_1 : 第1四分位点, Q_3 : 第3四分位点

補足: Q_2 : 第2四分位点は
中央値 x_{med} である

- 注意: 四分位数の定義は複数ある

- $k_1 := 0.25 \times (n-1)$, $k_3 := 0.75 \times (n-1)$ とし,

$$\begin{cases} Q_1 = x_{[k_1]+1} + (k_1 - [k_1]) \times (x_{[k_1]+2} - x_{[k_1]+1}) \\ Q_3 = x_{[k_3]+1} + (k_3 - [k_3]) \times (x_{[k_3]+2} - x_{[k_3]+1}) \end{cases}$$

- $Q_1 = x_{[0.25 \times n]}$, $Q_3 = x_{n+1-[0.25 \times n]}$ など

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
データ	10	7	3	5	7	5	10	9	6	7	50	7	5	7	6	10
ソート後	3	5	5	5	6	6	7	7	7	7	7	9	10	10	10	50

※quartile: 四分位数
quantile: 分位数

MS Excel の関数QUARTILE() では, $Q_1=5.75$, $Q_3=9.25$
Mathematica の関数quantile[]では, $Q_1=5$, $Q_3=9$
Rの関数quantile()では, $Q_1=5.75$, $Q_3=9.25$

代表値 AVERAGES

○ミッド・レンジ **mid-range**

- データの最大値と最小値の算術平均

$$x_{MR} = \frac{\max\{x_1, \dots, x_n\} + \min\{x_1, \dots, x_n\}}{2}$$

x	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈	x ₉	x ₁₀	x ₁₁	x ₁₂	x ₁₃	x ₁₄	x ₁₅	x ₁₆
データ	10	7	3	5	7	5	10	9	6	7	50	7	5	7	6	10
ソート後	3	5	5	5	6	6	7	7	7	7	7	9	10	10	10	50

→ $x_{MR} = \frac{\max(10, 7, \dots, 10) + \min(10, 7, \dots, 10)}{2} = \frac{50 + 3}{2} = 26.5$

演習1-3:代表値

○統計データを使って代表値を計算する

- 総務省統計局 (<http://www.stat.go.jp>) から世帯収入, 世帯貯蓄などのデータを取得し, グラフ化せよ. グラフの形状はどのようなになるか?
- このデータの「算術平均」「中央値」「最頻値」を計算し, 分布の代表値として最も適切だと思われるのはどれか考察せよ.
- 「第1四分位数」「第3四分位数」「ミッドレンジ」を求めよ.

○簡単なデータを使って代表値を計算する

- 以下の10個のデータがある

1	20	20	22	23	24	25	26	26	53
---	----	----	----	----	----	----	----	----	----

- 「算術平均」「中央値」「最頻値」を求めよ.
- 「第1四分位数」「第3四分位数」「ミッドレンジ」を求めよ.

1-1. 一次元のデータ

■データの散らばり

- 範囲
- 四分位偏差
- 平均偏差
- 分散, 標準偏差

$x = (x_1, x_2, \dots, x_n)$

$x_1, x_2, x_3, x_4, x_5, x_6$

x	11	9	-3	14	5	23
-----	----	---	----	----	---	----

($n = 6$)

データの値らばりを考える

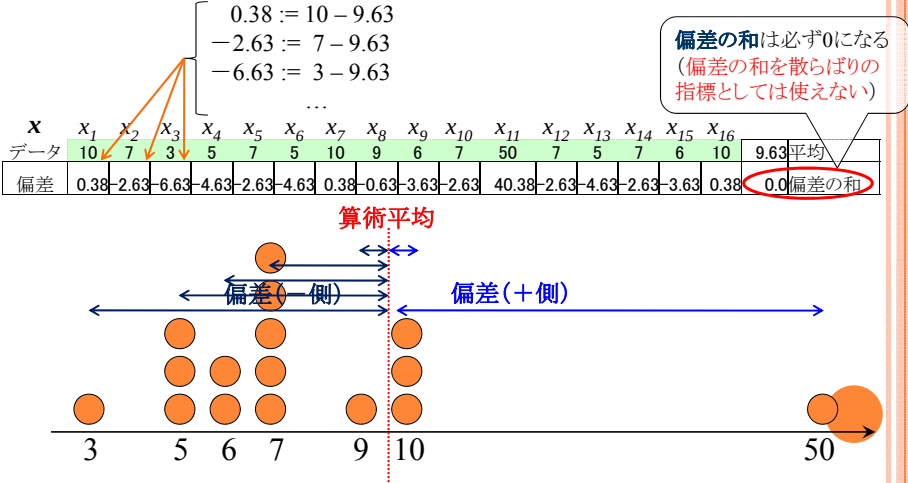
■ 例:16個のデータ

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}
データ	10	7	3	5	7	5	10	9	6	7	50	7	5	7	6	10

散らばり DISPERSION

○ 偏差 deviation

- データと平均の差

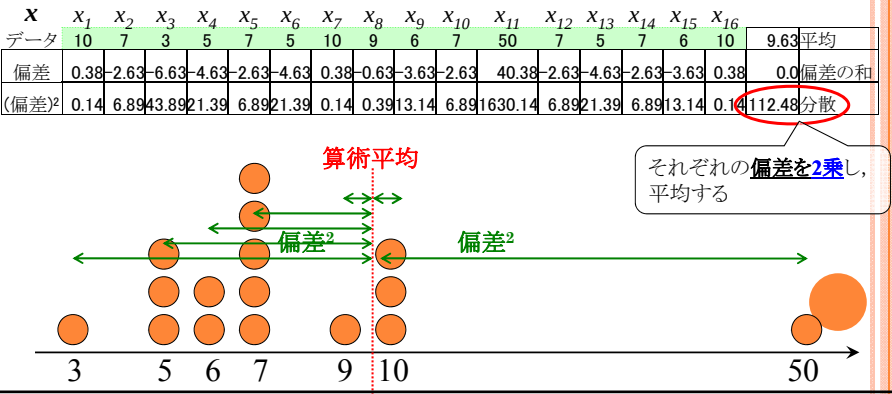


散らばり DISPERSION

○ 分散 variance

- 偏差の2乗和を平均化した値

$$S_x^2 = \frac{(10 - 9.63)^2 + (7 - 9.63)^2 + \dots + (10 - 9.63)^2}{16}$$



散らばり DISPERSION

標準偏差 standard deviation

- 分散の平方根

$$S_x = \sqrt{\frac{(10-9.63)^2 + (7-9.63)^2 + \dots + (10-9.63)^2}{16}}$$

X	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈	x ₉	x ₁₀	x ₁₁	x ₁₂	x ₁₃	x ₁₄	x ₁₅	x ₁₆		
データ	10	7	3	5	7	5	10	9	6	7	50	7	5	7	6	10	9.63	平均
偏差	0.38	-2.63	-6.63	-4.63	-2.63	-4.63	0.38	-0.63	-3.63	-2.63	40.38	-2.63	-4.63	-2.63	-3.63	0.38	0.0	偏差の和
(偏差) ²	0.14	6.89	43.89	21.39	6.89	21.39	0.14	0.39	13.14	6.89	1630.14	6.89	21.39	6.89	13.14	0.14	112.48	分散
																	10.61	標準偏差

分散の平方根

散らばり DISPERSION

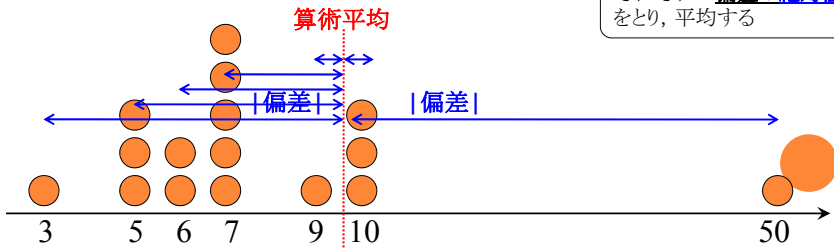
平均偏差 mean deviation

- 偏差の絶対値の合計を平均化した値

平均値からの
平均的な差

X	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈	x ₉	x ₁₀	x ₁₁	x ₁₂	x ₁₃	x ₁₄	x ₁₅	x ₁₆		
データ	10	7	3	5	7	5	10	9	6	7	50	7	5	7	6	10	9.63	平均
偏差	0.38	-2.63	-6.63	-4.63	-2.63	-4.63	0.38	-0.63	-3.63	-2.63	40.38	-2.63	-4.63	-2.63	-3.63	0.38	0.0	偏差の和
(偏差) ²	0.14	6.89	43.89	21.39	6.89	21.39	0.14	0.39	13.14	6.89	1630.14	6.89	21.39	6.89	13.14	0.14	112.48	分散
																	10.61	標準偏差
偏差	0.38	2.63	6.63	4.63	2.63	4.63	0.38	0.63	3.63	2.63	40.38	2.63	4.63	2.63	3.63	0.38	5.19	平均偏差

それぞれの偏差の絶対値
をとり、平均する

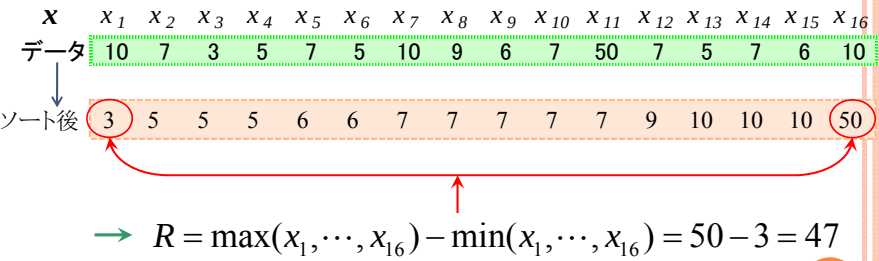


散らばり DISPERSION

○ 範囲 **range**

- 最大値と最小値の差

$$R = \max\{x_1, \dots, x_n\} - \min\{x_1, \dots, x_n\}$$

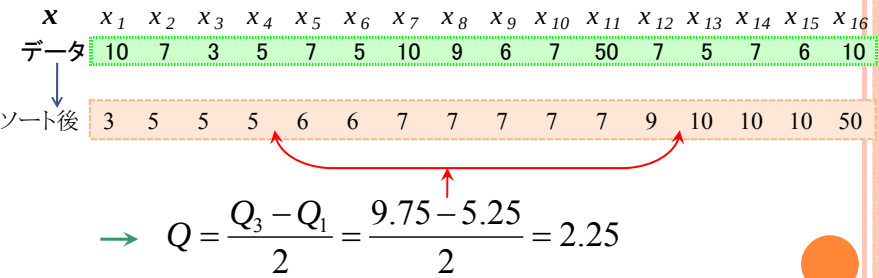


散らばり DISPERSION

○ 四分位偏差 **quartile deviation**

- 第3四分位点 Q_3 と第1四分位点 Q_1 の差の半分

$$Q = \frac{Q_3 - Q_1}{2}$$



演習1-4: 散らばり

- 以下のデータについて散らばりを計算したい

1 20 20 22 23 24 25 26 26 53

- このデータの「偏差」をだし、合計が0になることを確かめよ.
- このデータの「分散」を計算せよ.
- このデータの「標準偏差」を計算せよ.
- このデータの「平均偏差」を計算せよ.
- このデータの「範囲」を計算せよ.
 - 例) data[1, 5, 7, 9, 3] → 範囲: 9 - 1 = 8
- このデータの「四分位偏差」を計算せよ.

COFFEE BREAK

- 記号を用いた散らばりの定義

- 分散

$$S_x^2 = \frac{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}$$

- 標準偏差

$$S_x = \sqrt{\frac{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}}$$

- 平均偏差

$$d = \frac{|x_1 - \bar{x}| + \cdots + |x_n - \bar{x}|}{n}$$

1-1. 一次元のデータ

■ データの変換
標準化(正規化)
Cf. 偏差値

$x = (x_1, x_2, \dots, x_n)$

$x_1, x_2, x_3, x_4, x_5, x_6$

x	11	9	-3	14	5	23
-----	----	---	----	----	---	----

($n = 6$)

データの一次変換

○ 標準化 standardization

- 各データについて、平均を引き標準偏差で割る

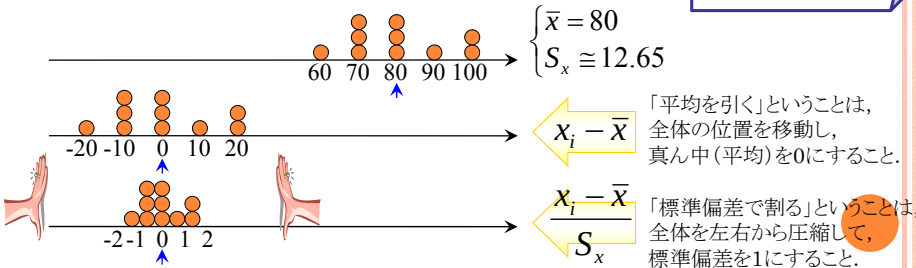
$$z_i = \frac{x_i - \bar{x}}{S_x} \quad (i = 1, \dots, n)$$

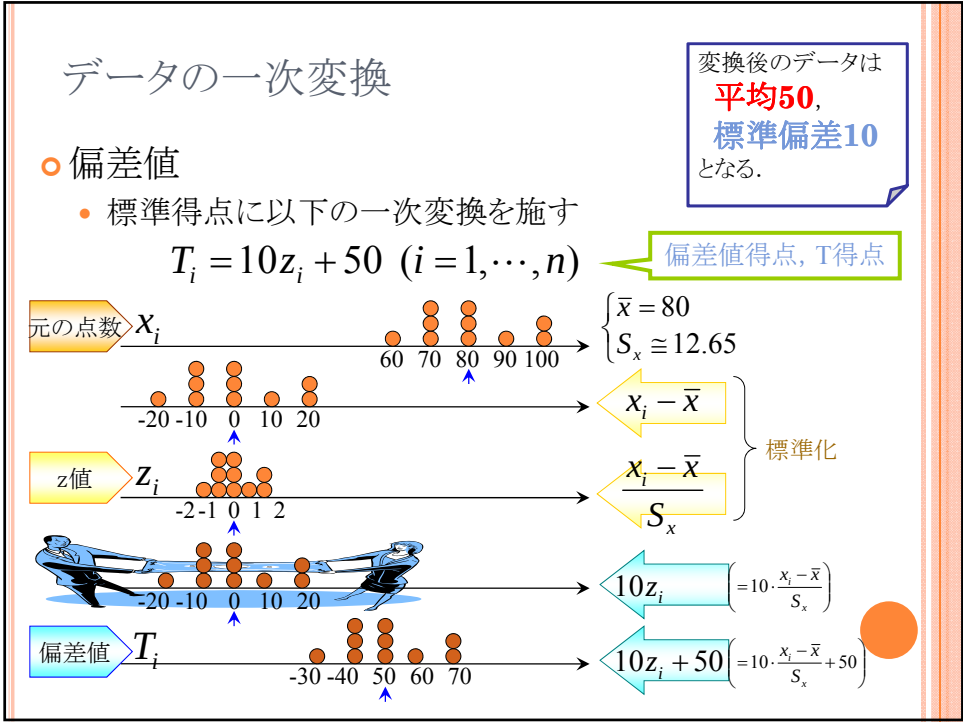
標準得点 standard score, Z得点

どんな1次元データも
標準化しさえすれば
同じ土俵で比較できるね！



変換後のデータは
平均0,
標準偏差1
となる。





データの一次変換

■ 例: 10人の中間・期末試験の得点, z得点と偏差値

平均88, 標準偏差9.8

中間試験

得点	100	90	80	80	90	100	80	90	100	70
z得点	1.2	0.2	-1	-1	0.2	1.2	-1	0.2	1.2	-2
偏差値	62	52	42	42	52	62	42	52	62	32

$1.2 = \frac{100 - 88}{9.8},$
 $62 = 1.2 \times 10 + 50$

平均33, 標準偏差16

期末試験

得点	40	20	60	20	40	10	50	45	25	15
z得点	0.5	-1	1.7	-1	0.5	-1	1.1	0.8	-0	-1
偏差値	55	42	67	42	55	36	61	58	45	39

演習1-5:データの標準化

- 偏差値を計算しよう
 - 以下のデータはある試験の5人の学生の結果である.
 - 英語の結果について, 各学生の得点を標準化し, z得点を出せ.
 - 英語のz得点をもとに, 各学生の偏差値を計算せよ.
 - 数学・国語についても同様に計算せよ.

	A	B	C	D	E
英語	22	28	36	74	50
国語	78	50	51	33	28
数学	27	74	38	26	95



1-1. 一次元のデータ

➤データの尺度

$$\mathbf{x} = (x_1, \quad x_2, \quad \cdots, \quad x_n)$$

$x_1, x_2, x_3, x_4, x_5, x_6$

x

11

9

-3

14

5

23

$(n = 6)$

データの測定尺度による分類

測定(measurement)と尺度(scale)

名義(名目)尺度 nominal scale

- 属性を表す基準(対象に区別がつけられる)
- 例:性別(男, 女, それ以外), パソコン保有(保有, 非保有)

質的(カテゴリ)データ

順序尺度 ordinal scale

- 対象間に順序がつけられる基準
- 例:成績(A>B>C>D), 居住性(住みやすい>まあまあ>すみにくい)

質的(カテゴリ)データ

間隔尺度 interval scale

- 間隔のみが意味を持つ基準
- 例:温度(摂氏℃, 華氏°F), 時刻(午後3時から1時間後)

量的(数値)データ

比率尺度 ratio scale

- 比が意味を持つ基準
- 例:身長(父は子の1.5倍の背), 体重(5kg重い), 絶対温度(°K, 絶対零度)

量的(数値)データ

測定が
厳密

データの測定尺度による集計例

質的データと量的データの集計例

質的データ

	性別		成績			
	男	女	A	B	C	D
データ例	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓
	✓	✓	✓	✓	✓	✓

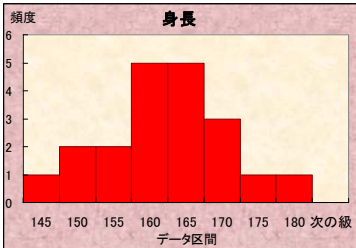
集計例

	A	B	C	D	計
男	3	2	1	0	6
女	1	0	2	2	5
計	4	2	3	2	11

量的データ

女性身長

165	155	159	155	167
160	175	157	150	149
145	162	162	159	159
162	162	177	166	168



演習1-6:データの尺度

- 身の回りにあるデータは, 4つの尺度のどれに相当するか考えてみよう.

