1　記述統計学(1)：データの要約

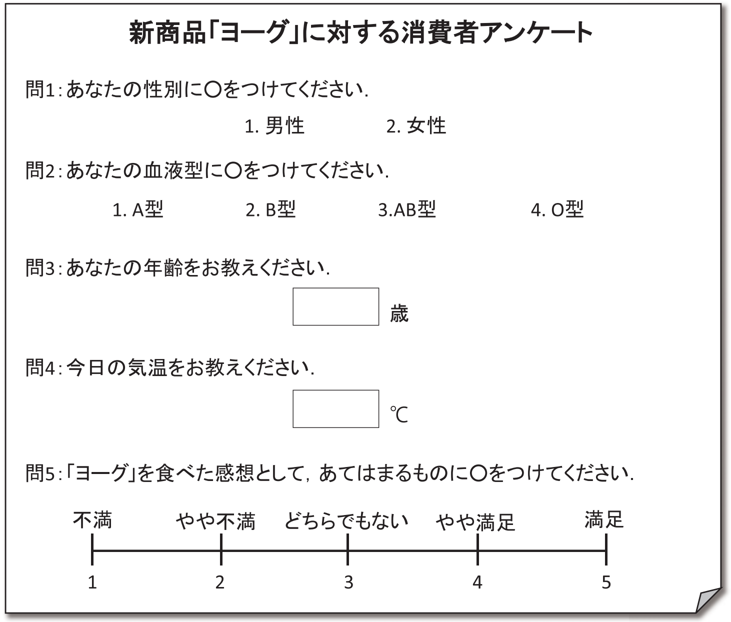
◉本章の目標◉

1. 変数の形式について理解する．
2. 位置およびばらつきを表す指標について理解する．
3. 標準化の方法とその意味を理解する．

本章では，統計学の出発点として，平均値，中央値，分散あるいは標準偏差といったデータを要約するための指標について述べる．

1.1　統計学におけるデータのとり扱い

ある食品企業が新商品のヨーグルト「ヨーグ」に対して，購入者に次のようなアンケート調査を行った．



新商品「ヨーグ」に対する消費者アンケート

問1：あなたの性別に〇をつけてください．

1. 男性　　2. 女性

問2：あなたの血液型に〇をつけてください．

1. A型　　2. B型　　3.AB型　　4. O型

問3：あなたの年齢をお教えください．

歳

問4：今日の気温をお教えください．

℃

問5：「ヨーグ」を食べた感想として，あてはまるものに〇をつけてください．

不満　やや不満　どちらでもない　やや満足　満足

1　2　3　4　5

このとき，調査項目のことを**変数**といい，各購入者のデータを**個体**という．つまり，今回は 100 個の個体が存在することになる．このとき，個体の総数を**標本サイズ** (**サンプルサイズ**，**個体数**) という．

変数は，**量的変数**と**質的変数**に大別される．量的変数とは，今回のアンケート調査で年齢 (問 3) と気温 (問 4) に該当するような，少なくとも個体間の差に意味があるような変数を指す．いいかえれば，量的変数とは変数の値の大きさ (量) が意味をもつ変数である．一方で，質的変数とは，今回のアンケート調査で性別 (問 1)，血液型 (問 2)，5 段階評価 (問 5) に該当するような変数間の差に意味がないようなデータを指す．いいかえれば，変数の値の大きさが意味をもたない変数である．

量的変数は，データの形式によって**間隔尺度**と**比例尺度**に分かれる．間隔尺度とは，値の比が意味をもたない変数を表し，気温 (問 4) が該当する．たとえば，購入者 A さんの調査時の気温が -3 ℃で，購入者 B さんの調査時の気温が 12 ℃だったとする．このとき，A さんの解答時の気温と B さんの解答時の気温について，その違いを $-3/12$ で表すことはできない．これに対して，比例尺度とは，値の比が意味をもつ変数を表し，年齢 (問 3) が該当する．たとえば，A さんの年齢が 40 歳で，B さんの年齢が 20 歳だったとする．このとき，A さんと B さんの年齢の違いを $40/20 = 2.0$ で表すことで，「A さんは B さんの 2 倍生きている」と解釈できる．

質的変数は，データの形式によって**名義尺度**と**順序尺度**に分かれる．名義尺度とは，カテゴリに順序関係が存在しない変数を表し，性別 (問 1) および血液型 (問 2) が該当する．また，性別は男性，女性のいずれかしかとらないため，2 値変数と呼ぶことがある．これに対して，血液型は A 型，B 型，AB 型，O 型の 4 個の選択肢が存在するため，多値変数と呼ぶことがある．これら名義尺度に対して，順序尺度とは，カテゴリに順序関係が存在する変数を表す．5 段階評価 (問 5) では，1 (不満) と回答した購入者よりも 4 (やや満足) と回答した購入者のほうが商品に満足している．したがって，5 段階評価 (問 5) は，順序尺度である．

1.2　量的変数の要約

新商品のヨーグルト「ヨーグ」のアンケート調査が，10,000 名の購入者を対象として実施された場合，すべての調査票に目を通して，その傾向を捉えることはほぼ不可能である．そのため，統計学では，要約指標を用いて，データの傾向を捉える．さまざまな要約指標のなかから，本書では，「平均的にどのような値をとるか」を表す指標 (位置を表す指標) と「データがどれぐらい散らばっているか」を表す指標 (散らばりを表す指標) について解説する．

表 1.1　Y 大学の 7 名の学生の身体測定における体重

ID	名前	体重 (kg)
1	A さん	54
2	B さん	67
3	C さん	66
4	D さん	62
5	E さん	58
6	F さん	59
7	G さん	61

1.2.1　位置を表す指標

位置を表す指標の代表的なものに，平均値，中央値，そして最頻値がある．ここでは，平均値および中央値について解説し，最頻値については第 2 章で触れる．

(a)　平均値

平均値 (**相加平均**，**算術平均**) は，日常生活においても一般的に用いられている指標の 1 つであり，統計学においても重要である．いま，Y 大学で身体測定が行われたとする．表 1.1 は，そのなかの 7 名の体重を表している．このとき，7 名の平均値 $\bar{x}$ は，

$$\bar{x} = \frac{54+67+66+62+58+59+61}{7} = \frac{427}{7} = 61$$

である．数式を用いて平均値を定義すると以下のようになる．

❖平均値の定義

いま，n 個の個体 $x_1, x_2, \ldots, x_n$ が与えられたとき，平均値 $\bar{x}$ は

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

で与えられる．ここで，$\sum$ は総和を表す演算記号である (1.4 節を参照)．

ちなみに，平均値と各個体の差をとって合計すると $(54-61)+(67-61)+\cdots+(59-61)=0$ になる．すなわち，平均値とはデータの重心を表している．

(b)　中央値

表 1.1 の身体測定における体重のデータをグラフで表したものが，図 1.1(a) である．データの重心を表す平均値は真ん中あたりを表しており，平均値が全体を代表する値であることがわかる．図 1.1(b) は，G さん (61kg) の代わりに巨漢の H さん (124kg) が入った場合である．このとき，平均値は 70(kg) となるが，H さん以外に，平均値よりも重い学生 (個体) は存在しない．このように，飛び抜けた値のことを**外れ値**という．外れ値が存在する場合に，平均値が全体を代表する値といえるかは疑問である．

外れ値が存在する場合に，位置を代表する指標として用いられるのが**中央値** (**メジアン**) である (中央値の利用は外れ値が存在する場合のみではないが，それについては第 2 章で触れる)．中央値とは，データを小さい順に並べ替えたときの，真ん中の値である．定義を以下に示す．

❖中央値の定義

いま，n 個の個体 $x_1, x_2, \ldots, x_n$ が与えられたとき，小さい順に並べ替えられたものを $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ で表す．このとき，中央値の順位は

$$m = \frac{n}{2} + 0.5$$

である．m の整数部分を m^- とするとき (n が奇数の場合には小数点以下が存在しないため $m = m^-$)，中央値 $\tilde{x}$ は

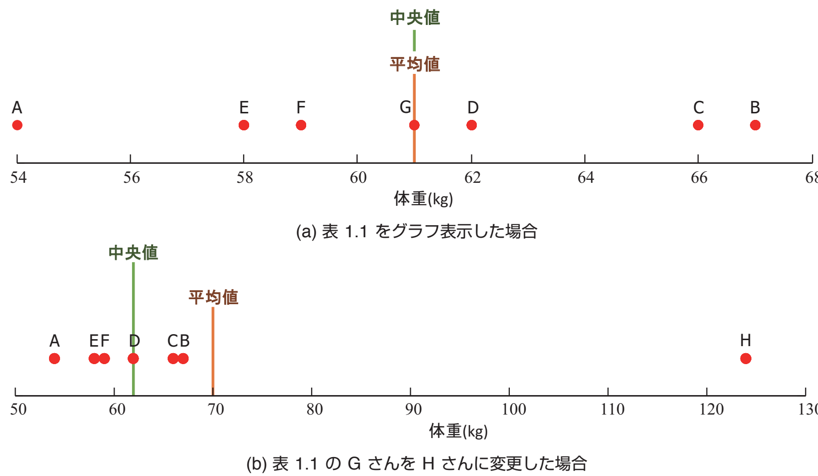
$$\tilde{x} = \begin{cases} x_{(m)}, & n \text{ が奇数の場合}, \\ \dfrac{1}{2}\left(x_{(m^-)} + x_{(m^-+1)}\right), & n \text{ が偶数の場合} \end{cases}$$

で与えられる．

表 1.1 の身体測定における体重のデータを小さい順に並べ替えると，

$$54 \quad 58 \quad 59 \quad 61 \quad 62 \quad 66 \quad 67$$

となる．標本サイズ n は $n = 7$ なので，中央値の番号 m は



(a) 表 1.1 をグラフ表示した場合

(b) 表 1.1 の G さんを H さんに変更した場合

図 1.1　7 名の学生の身体測定における体重のデータに対するグラフ表現

$$\frac{7}{2} + 0.5 = 4$$

である．よって，中央値 $\tilde{x}$ は，$\tilde{x} = 61$ である．

また，G さんを H さんに変更した場合に，データを小さい順に並べ替えると，

$$54 \quad 58 \quad 59 \quad 62 \quad 66 \quad 67 \quad 124$$

となる．中央値の番号 m は先ほどと同じ $m = 4$ なので，G さんを H さんに変更した場合の中央値は $\tilde{x} = 62$ である．

外れ値が存在しない場合のデータでは，平均値と中央値が同じであり，中心付近にあった (図 1.1(a))．したがって，平均値および中央値は位置を代表しているといえる．これに対して，外れ値が存在する場合 (G さんを H さんに変更した場合) には，中央値のほうが平均値よりも低いものの，データの真ん中にあることから，位置を代表しているといえる (図 1.1(b))．

例 1.1 ：先ほどのヨーグルト「ヨーグ」のアンケート調査において，6 名の被験者の年齢 (問 3) は，

$$42 \quad 31 \quad 63 \quad 29 \quad 56 \quad 38$$

だった．このときの平均値 $\bar{x}$ は

$$\bar{x} = \frac{42+31+63+29+56+38}{6} = \frac{259}{6} = 43.17$$

である．

次いで中央値を計算する．このデータを小さい順に並べ替えると

$$29 \quad 31 \quad 38 \quad 42 \quad 56 \quad 63$$

になる．中央値の番号 m は 3.5 であり，整数部分 m^- は 3 である．$x_{(3)} = 38$, $x_{(4)} = 42$ なので，中央値 $\tilde{x}$ は

$$\tilde{x} = \frac{1}{2}(38 + 42) = 40$$

である．

1.2.2 散らばり表す指標

統計学では，データの位置 (平均) とともに，散らばり (ばらつき) を評価することが重要である．表 1.2 は，表 1.1 に W 大学での結果を追加したものである．図 1.2 は，表 1.2 のデータをグラフで表したものである．平均値は，いずれの大学でも 61(kg) であるものの，W 大学のほうが Y 大学よりも散らばり具合が大きい．いいかえれば，W 大学での体重の個人差は，Y 大学よりも大きい．このように，平均値だけではデータの特徴を把握できない場合があるので，散らばりを吟味することは重要である．

(a) 範囲と四分位範囲

散らばりを表す指標として最も単純なものが**範囲 (レンジ)** である．範囲 R とはデータ全体が含まれる区間の長さであり，

$$R = (\text{データの最大値}) - (\text{データの最小値})$$

で計算される．表 1.2 における範囲は，

表 1.2　Y 大学および W 大学の身体測定における体重データ

Y 大学		W 大学	
名前	体重 (kg)	名前	体重 (kg)
A さん	54	I さん	71
B さん	67	J さん	65
C さん	66	K さん	53
D さん	62	L さん	72
E さん	58	M さん	46
F さん	59	N さん	52
G さん	61	O さん	68

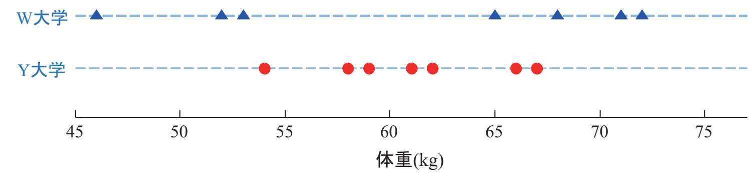


図 1.2　Y 大学および W 大学の体重データに対するグラフィカル表現

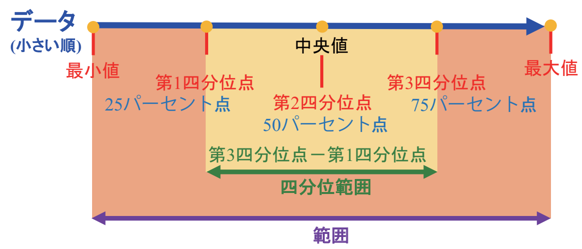


図 1.3　範囲および四分位範囲の概念図

$$\text{Y 大学の範囲 } R_{\mathrm{Y}} = 67 - 54 = 13, \quad \text{W 大学の範囲 } R_{\mathrm{W}} = 72 - 46 = 26$$

であり，W 大学のほうがデータの範囲は広かった．

範囲はデータの 100 パーセントが含まれる区間として定義されている．そのため，データに外れ値が存在する場合には，最大値および最小値を用いる範囲は，非常に大きな値をとる．この難点を回避するために，中央値を中心に 50 パーセントの値を含む区間として定義されているのが**四分位範囲**である．

図 1.3 は，範囲および四分位範囲の概念図である．中央値とはデータの真ん中であることから，中央値以下をとるデータの割合は 50 パーセントになる．そのため，中央値は **50 パーセント点**とも呼ばれる．さらに，最小値と中央値の真ん中の値，中央値と最大値の真ん中の値でデータを分けると，データを含む割合が 25 パーセントずつ 4 等分される．最初の区分点 (最小値と中央値の真ん中の値) は，**第 1 四分位点 (25 パーセント点)**Q_1 と呼ばれ，第 1 四分位点以下をとるデータの割合は 25 パーセントである．2 番目の区分点は中央値であるが，四分位数 (データを 4 等分した値) としては，**第 2 四分位点** Q_2 と呼ばれる．3 番目の区分点 (中央値と最大値の真ん中の値) は，**第 3 四分位点 (75 パーセント点)**Q_3 と呼ばれ，第 3 四分位点以下をとるデータの割合は 75 パーセントである．

そして，四分位範囲 IQR は，第 3 四分位点と第 1 四分位点の差 $IQR = Q_3 - Q_1$

で計算される．ただし，実際のデータではデータの番号の添え字に小数点以下の桁が存在することから，いくつかの計算方法が存在する．以下に，四分位範囲と四分位点の計算方法の 1 つを紹介する．

まず，第 1 四分位点 Q_1 は，最小値 $m_{(1)}$ と中央値 $\tilde{x} = x_{(m)}$ の真ん中に存在するため，Q_1 の番号 m_1 は

$$m_1 = \frac{m+1}{2} = \frac{n/2+0.5+1}{2} = \frac{n}{4} + 0.75$$

である．

次いで, 第 3 四分位点 Q_3 は, 中央値 $\tilde{x} = x_{(m)}$ と最大値 $x_{(n)}$ と中央値 $\tilde{x} = x_{(m)}$ の真ん中に存在するため，Q_3 の番号 m_3 は，

$$m_3 = \frac{m+n}{2} = \frac{n+n/2+0.5}{2} = \frac{3n}{4} + 0.25$$

である．

したがって，四分位範囲の定義および計算方法は，次のように与えられる．

❖四分位範囲の定義および計算方法

いま，n 個の個体 $x_1, x_2, \ldots, x_n$ が与えられたとき，小さい順に並べ替えられたものを $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ で表す．第 1 四分位点 Q_1 の番号 m_1 および，第 3 四分位点 Q_3 の番号 m_3 は，それぞれ，

$$m_1 = \frac{n}{4} + 0.75, \qquad m_3 = \frac{3n}{4} + 0.25$$

である．$m_j (j = 1, 3)$ の整数部分を m_j^-，小数部分を m_j' とすると，分位点 $Q_j (j = 1, 3)$ は

$$Q_j = (1 - m_j') \cdot x_{(m_j^-)} + m_j' \cdot x_{(m_j^- + 1)}$$

で与えられる．

また，中央値まわりでデータの 50 パーセントを含む範囲として定義される四分位範囲 IQR は

$$IQR = Q_3 - Q_1$$

である．

例 1.2 ：ここでは，表 1.2 の 2 大学の学生の体重のデータを用いる．それぞれのデータを小さい順 (昇順) に並べ替える．

番号	1	2	3	4	5	6	7
Y 大学	54	58	59	61	62	66	67
W 大学	46	52	53	65	68	71	72

第 1 四分位点の番号 m_1 および第 3 四分位点の番号 m_3 を計算する (Y 大学および W 大学はいずれも $n = 7$ なので，第 1 四分位点，第 3 四分位点の番号 m_1, m_3 は同じである).

$$m_1 = \frac{7}{4} + 0.75 = 2.5, \quad m_3 = \frac{3 \times 7}{4} + 0.25 = 5.5.$$

Y 大学の四分位範囲は,

$$\text{Y 大学の第 1 四分位点：} Q_{\mathrm{Y},1} = (1-0.5) \times 58 + 0.5 \times 59 = 58.5,$$

$$\text{Y 大学の第 3 四分位点：} Q_{\mathrm{Y},3} = (1-0.5) \times 62 + 0.5 \times 66 = 64.0$$

なので,

$$IQR_{\mathrm{Y}} = Q_{\mathrm{Y},3} - Q_{\mathrm{Y},1} = 64.0 - 58.5 = 5.5$$

である．また，W 大学の四分位範囲は,

$$\text{W 大学の第 1 四分位点：} Q_{\mathrm{W},1} = (1-0.5) \times 52 + 0.5 \times 53 = 52.5,$$

$$\text{W 大学の第 3 四分位点：} Q_{\mathrm{W},3} = (1-0.5) \times 68 + 0.5 \times 71 = 69.5,$$

より,

$$IQR_{\mathrm{W}} = Q_{\mathrm{W},3} - Q_{\mathrm{W},1} = 69.5 - 52.5 = 17.0$$

である．よって，Y 大学のほうが，W 大学よりも四分位範囲が狭い，すなわち個人差が小さいことがわかった．

(b) 分散と標準偏差

範囲および四分位範囲は中央値まわりの散らばりの大きさを表している．ここでは，平均値まわりの散らばりの大きさを考える．図 1.4 は，Y 大学の体重のデータ (表 1.1) をグラフで表している．平均値まわりの散らばりは，平均値からの差 $\epsilon_i = x_i - \bar{x}$ を評価することが考えられる．ちなみに $\epsilon_i (i = 1, \ldots, n)$ は**偏差**と呼ばれる．ただし，平均値の性質より，偏差の和は 0 であることから，偏差の平均値を散らばりの指標に利用できない．そのため，**分散**は，偏差を 2 乗したうえで，その平均値を用いる．

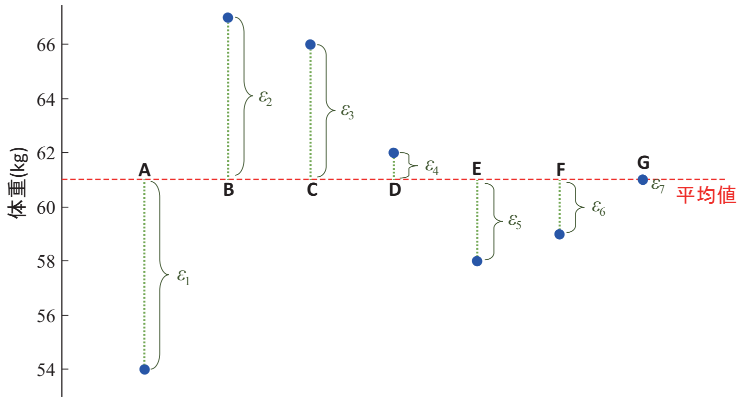


図 1.4　Y 大学の体重データに対する平均値まわりの散らばり

❖分散の定義

いま，n 個の個体 $x_1, x_2, \ldots, x_n$ が与えられたとき，分散 S^2 は

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

で与えられる．ここで，$\bar{x}$ は x_i の平均値である．

範囲および四分位範囲は，散らばりの指標と位置を表す指標の単位は同じである (たとえば，Y 大学における体重データの中央値まわりで 50 パーセントを含む範囲は 5.5kg である)．一方で，分散は，偏差を 2 乗するため，平均値の単位と同じ単位をもたない．平均値と単位を揃えることを意図して，分散に平方根をとった指標が**標準偏差**である．

❖標準偏差の定義

いま，n 個の個体 $x_1, x_2, \ldots, x_n$ が与えられたとき，標準偏差 S は

$$S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{S^2}$$

で与えられる．ここで，$\bar{x}$ は x_i の平均値である．

表 1.3　体重データにおける 2 乗和 (偏差平方和) の計算

(a) Y 大学			(b)W 大学		
x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
54	-7	49	71	10	100
67	6	36	65	4	16
66	5	25	53	-8	64
62	1	1	72	11	121
58	-3	9	46	-15	225
59	-2	4	52	-9	81
61	0	0	68	7	49

例 1.3：これまでと同様に，2 大学の学生の体重のデータを用いる (表 1.2)．分散および標準偏差を計算するためには，偏差 $\epsilon_i = x_i - \bar{x}$ の平方和 (**偏差平方和**) を計算しなければならない．表 1.3 は，偏差 $\epsilon_i = x_i - \bar{x}$ および偏差平方 $\epsilon_i^2 = (x_i - x)^2$ である．これらの値を用いることで，Y 大学の体重データの分散 S_{Y}^2 および標準偏差 S_{Y} は

$$\begin{aligned} S_{\mathrm{Y}}^2 &= \frac{1}{7}(49+36+25+1+9+4+0) = \frac{124}{7} = 17.714 \\ S_{\mathrm{Y}} &= \sqrt{17.714} = 4.209 \end{aligned}$$

であり，W 大学の体重データの分散 S_{W}^2 および標準偏差 S_{W} は

$$\begin{aligned} S_{\mathrm{W}}^2 &= \frac{1}{7}(100+16+64+121+225+81+49) = \frac{656}{7} = 93.714 \\ S_{\mathrm{W}} &= \sqrt{93.714} = 9.681 \end{aligned}$$

である．よって，範囲および四分位範囲と同様に，Y 大学のほうが，W 大学よりも分散および標準偏差が小さいことがわかった．

1.2.3　標準化

ある高校において，身体測定が行われ，男子の身長の平均値 $\bar{x}_{男}$ が $\bar{x}_{男} = 172$cm(標準偏差 $S_{男} = 12$)，女子の身長の平均値 $\bar{x}_{女}$ が $\bar{x}_{女} = 158$cm(標準偏差 $S_{女} = 8$) だった．かずお君の身長が 173cm，ゆみこさんの身長が 161cm であるとき，身長には性差がある (平均値および標準偏差が異なる) ため，2 人の身長を単純に比較することはできない．このような場合に用いられるのが**標準化** (**規準化**) である．

❖標準化の定義

いま，平均値 $\bar{x}$ および標準偏差 S が与えられたとき，x の標準化後の値 z は

$$z = \frac{x - \bar{x}}{S}$$

である．標準化後の値 z の平均値は 0，標準偏差は 1 になる．

かずお君とゆみこさんの身長を標準化すると

$$z_{\text{かずお}} = \frac{173 - 172}{12} = 0.083, \qquad z_{\text{ゆみこ}} = \frac{161 - 158}{8} = 0.375$$

となり，ゆみこさんのほうがかずお君よりも身長が相対的に高かった．

例 1.4： A 君と B 君は第 2 外国語にドイツ語を履修しており，C 君は中国語を履修している．ドイツ語の定期試験での平均値は 67 点であり，標準偏差は 15 だった．一方で，中国語の平均点は 52 点であり，標準偏差は 18 だった．3 人の定期試験での点数が

A 君の点数：$x_{\mathrm{A}} = 74$ 点，　B 君の点数：$x_{\mathrm{B}} = 61$ 点，　C 君の点数：$x_{\mathrm{C}} = 68$ 点

であるとき，点数のみを見ると，$x_{\mathrm{A}} > x_{\mathrm{C}} > x_{\mathrm{B}}$ である．ただし，科目ごとの成績を考慮するために標準化を行うと

$$z_{\mathrm{A}} = \frac{74 - 67}{15} = 0.467, \quad z_{\mathrm{B}} = \frac{61 - 67}{15} = -0.400, \quad z_{\mathrm{C}} = \frac{68 - 52}{18} = 0.889$$

となり，成績は $z_{\mathrm{C}} > z_{\mathrm{A}} > z_{\mathrm{B}}$ の順序であると考えることができる．

1.2.4　変動係数

分散あるいは標準偏差では，単位が異なる変数間の散らばりの大きさを相対的に評価できない．たとえば，象と鼠の体重の散らばり具合をそれぞれの分散および標準偏差のみでは比較できない．このような場合に用いられるのが，**変動係数**である．

❖受動係数の定義

いま，平均値 $\bar{x}$ および標準偏差 S が与えられたとき，変動係数 CV は

$$CV = \frac{S}{\bar{x}}$$

である．

例 1.5：ペット用品会社が開発したダイエット用のペットフードを 6 ヵ月間肥満の大型犬に与えたところ，平均 3,200g，標準偏差 1,256 の体重減少が認められ，肥満の小型犬では平均 600g，標準偏差 264 の体重減少が認められた．それぞれに対して変動係数を求めると

$$CV_{\text{大}} = \frac{1256}{3200} = 0.393, \qquad CV_{\text{小}} = \frac{264}{600} = 0.440$$

だった．よって，ダイエット用ペットフードの効果の個体差 (散らばり) は，小型犬のほうが大型犬よりも大きかった．

1.3　章末問題

問題 1.1： 次の項目 (変数) の尺度を答えなさい．

(1) 100m 走のタイム　(2) 学生の出身県

(3) 数学のテストの偏差値　(4) 疾病の重症度 (重度，中程度，軽度)

問題 1.2： 下記の変数に対して，質的変数である場合に ◯，量的変数である場合に × をつけなさい．

(1)　出身の都道府県

(2)　高校時代に入っていた部活

(3)　昨日の夕食での摂取カロリー

(4)　自動車の 1 リットルあたりの走行距離

(5)　マンションの築年数

問題 1.3： 次の表は平成 26 年春の全国交通安全運動 (4 月 6 日〜15 日) の都道府県別交通事故発生数である (出典：警察庁発表資料『平成 26 年春の全国交通安全運動期間中の交通事故発生状況』)．以下の問いに答えなさい

県	交通事故数	県	交通事故数
茨　城	348	神奈川	819
栃　木	152	新　潟	176
群　馬	418	山　梨	111
埼　玉	872	長　野	238
千　葉	512	静　岡	888

(1) 平均値と中央値を求めなさい．
(2) 分散と標準偏差を求めなさい．
(3) 四分位範囲を求めなさい．

問題 1.4 ： 以下の問題において，括弧内の正しい文章に ○ をつけなさい．

(1) 中央値を位置の指標として用いる場合に散らばりを表す指標として正しいものを選択しなさい：(分散あるいは標準偏差・四分位点範囲あるいは範囲)
(2) 外れ値の影響を受けない位置の指標はどちらか：(平均値・中央値)
(3) 個体の値を標準化したとき，その数字が負値になった．このときの正しい解釈を選びなさい：(平均より小さい値・中央値より小さい値)

問題 1.5 ： A さんは線形代数と統計学を受講している．定期試験の結果，A さんの線形代数の点数は 72 点であり，統計学の点数は 63 点であった．線形代数の平均値は 68 点，標準偏差は 10 点であり，統計学の平均点は 61 点，標準偏差は 15 点だったとするとき，以下の問いに答えなさい．

(1) A さんの線形代数と統計学の授業の点数を標準化しなさい．
(2) A さんは線形代数と統計学のどちらが優秀な成績を収めたと考えてよいか．

1.4　付録：総和 Σ の略説

いま，5 個の個体 $x_1, x_2, \ldots, x_5$ に次のような値

$$x_1 = 1,\quad x_2 = 2,\quad x_3 = 3,\quad x_4 = 4,\quad x_5 = 5$$

が代入されているとする．合計は，添え字 i を用いて

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

で表すことができる．すなわち，n 個の個体を合計する場合には

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

となる．また，2 乗和は

$$\sum_{i=1}^{n} x_i^2 = x_1^2 + x_2^2 + \cdots + x_n^2$$

である．総和 Σ のその他のいくつかの公式を以下に示す．

❖総和 Σ に対する簡単な公式

$$\sum_{i=1}^{n} a = na \quad (a \text{ は定数})$$

$$\sum_{i=1}^{n} ax_i = a(x_1 + x_2 + \cdots + x_n) = a\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} i = 1 + 2 + \cdots + n = \frac{n(n+1)}{2}$$

$$\begin{aligned}\sum_{i=1}^{n} (x_i + y_i) &= x_1 + y_1 + x_2 + y_2 + \cdots + x_n + y_n \\ &= \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i\end{aligned}$$