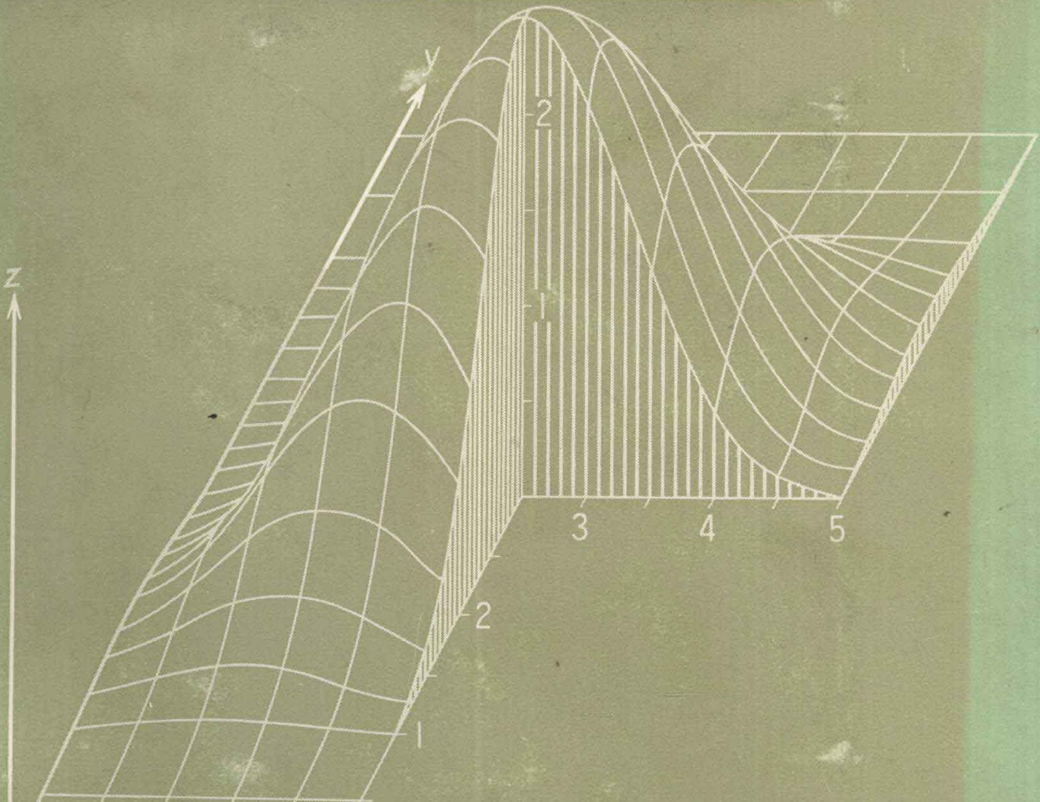
河田敬義
丸山文行 共著
鍋谷清治

大学演習 数理統計



裳華房

大学演習

数理統計

東京大学教授
理学博士
河田敬義

総理府技官
丸山文行

一橋大学助教授
鍋谷清治

共著

東京 裳華房 発行

著作者　河田敬義　丸山文行　鍋谷清治
発行者　吉野元章
印刷者　横山豊

大学演習 数理統計

定価 750円

発行所　東京都千代田区四番町8番地の1
振替口座東京107番
電話(331) 2811
6410
7854 番
合名会社 裳華房

印刷所　東京都墨田区亀沢町1～3　横山印刷株式会社

(福神製本)

社団法人 自然科学書協会会員

昭和37年9月10日 第1版印刷
昭和37年9月15日 第1版発行

ま え が き

本書は，大学における数理統計の基礎理論の教科書または参考書として，またはは社会人になられた方々が数理統計を勉強するときの独習書として書いたものである．

大学での統計学の講義は，数学について十分な予備知識をもたない時期に行なわれることがしばしばある．そこで本書の目的から，叙述はなるべく平易にして，しかも内容をできるだけ豊富にすることに意を用いた．予備知識としては VIII までは大学初年級でふつうに教えられている代数学，微分積分学より高度のものは仮定していない．複素函数論を用いて解いた例題や問題が二三あるが，それだけは例外である．IX では簡単な差分方程式の解法を用いている．

順列，組合せ，二項定理，多項定理は，高等学校または大学ですでに教えられているが，統計学での重要性から，それらの簡単な解説も本書にいれることにした．

内容の排列にあたっては，基本的なことがらをまず基礎事項としてあげ，読者の便をはかって，それらをできる限り例題のなかで証明する方針をとった．

本書を教科書として用いるときには，例題を参照しながら基礎事項を説明していかれればよいであろう．問題のなかにも基礎事項に近い性格のものがあるが，それらは適宜に取捨選択されることを希望する．

本書は，河田が全体の企画を，鍋谷が本文を，丸山が計算と図を，それぞれ担当した．

最後に，本書のためにいろいろご協力下さった裳華房並びに横山印刷の方々に感謝の意を表明する．

昭 和 37 年 7 月

著　者

凡　例

1 全体を9つの章 I, II, $\cdots$ に分け，IX を除く各章を節 §1, §2, $\cdots$ に分け，各節は基礎事項，例題とその解，問題，解答からなっている．

2 各節内で式を (1), (2), $\cdots$, 表を表1, 表2, $\cdots$, 図を図1, 図2, $\cdots$ とし，これらには節ごとに通し番号をつけてある．

3 各節の基礎事項のなかは項 1., 2., $\cdots$ に分けてある．基礎事項のなかで定理またはこれに準ずるものを (a), (b), $\cdots$ とし，各節内で通しの符号をつけてある．まえがきで述べたように，これらの大部分は例題のなかで証明されている．

4 問題は〔A〕と〔B〕に分けてあるが，番号は節ごとの通しになっている．〔A〕は比較的易しい問題，〔B〕はやや難しい問題である．問題〔B〕は後になって問題〔B〕以外の個所で引用されることはない．

5 引用の記号はつぎの例による．この例では引用されるものは，I の §2 にあると仮定する．

引用されるもの	同じ節での引用	同じ章内の他の節での引用	他の章での引用
式 (3)	(3)	(2.3)	(I.2.3)
表 3	表 3	表 2.3	表 I.2.3
図 3	図 3	図 2.3	図 I.2.3
項 3	§ 2.3	§ 2.3	§ I.2.3
定理等 (a)	(a)	(2.a)	(I.2.a)
例題 3	例 3	例 2.3	例 I.2.3
問題 3	問 3	問 2.3	問 I.2.3

6 巻末には本書にとって必要な分布の表をのせてある．付表 3 並びに付表 5 は，E. S. Pearson 教授の許可を得て，それぞれ

C. M. Thompson : Table of percentage points of the χ^2 distribution, Biometrika, Vol. 32, 1941,

M. Merrington and C. M. Thompson : Tables of percentage points of the inverted beta (F) distribution, Biometrika, Vol. 33, 1943

から転載したものである．付表 4 は古田義雄氏の許可を得て，

R. A. Fisher and F. Yates : Statistical tables for biological, agricultural and medical research, Oliver and Boyd, 1952

から転載したものである．両氏の好意に深く感謝する．

目　次

I　データの整理

§1　1変量の場合

§2　2変量の場合

§3　多変量の場合

II　予備的公式

§1　順列と組合せ

§2　二項定理と多項定理

§3　Γ 函数，B 函数，Stirling の公式

III 算術的確率

§1 定義と基本的性質

§2 条件つき確率と独立性

IV 確率変数

§1 一般の確率変数

§2 1次元の確率分布

§3 多次元の確率分布

V 基本分布

§1 1次元の基本分布

§ 2 多次元の基本分布

VI 標 本 分 布

§ 1 一般の母集団の場合

§ 2 正規母集団の場合

VII 推 定

§ 1 点 推 定

§ 2 信 頼 区 間

VIII 検　定

§1 基本的な検定

§2 適合度の検定

§3 分散分析

IX 時系列

付　表

I データの整理

§1 1変量の場合

1. 度数分布

ある目的のために調査なり実験なりを行なって，ある種の量 X について多数の測定値が得られたとしよう．このとき，これら多数の測定値を1つ1つ書き並べることなく，それらの数値全体としての性質を適確に表現できることが望ましい．そのような目的で度数分布がつくられる．

ここで問題とする量 X はいろいろな値をとり得る量であって，このような量を一般に**変量**という．変量のなかには，とり得る値が離散的な値（多くの場合は整数値）に限られているものもあれば，連続した範囲内のどの値をもとり得るものもある．さいころを投げたときに出る目の数や世帯人員などは前者の例であり，学生の身長や電球の寿命などは後者の例である．

連続的変量 X について多数の測定値が与えられたとき，それらの測定値をすべて含むような区間を，一定の幅の小区間に分割する．この各小区間を**階級**といい，各階級の中点の値をその階級の**階級値**という．各階級に属する測定値の個数をその階級の**度数**といい，各階級あるいは階級値に，度数を対応させた表を**度数分布表**という．n 個の測定値 x_1, $\cdots$, x_n の度数分布表は，階級値を x_1', x_2', $\cdots$, x_k', 度数を f_1, f_2, $\cdots$, f_k で表わせば，表1の形になる．

表 1

Xの値	x_1'	x_2'	$\cdots$	x_k'	計
度数	f_1	f_2	$\cdots$	f_k	n

整数値だけをとり得る離散的変量については，さいころの目や普通世帯の世帯人員などのように，とり得る値の範囲が比較的狭ければ，x_1', x_2', $\cdots$, x_k' のかわりにとり得る個々の値を与えればよい．しかし事業所の従業人員などのように，とり得る値の範囲が広ければ，連続的変量のときと同様に，階級を考えなければならない．

〔注意〕 階級値を求めるには，階級の境界の正確な値が必要になる．連続的変量に関する測定値については，階級の境界は，測定の結果を読みとるときに生じ得る2つの隣り合った値の算術平均となる．例えば，身長の測定を cm 単位で小数第1位まで正確に行なっ

てあれば，境界となる点は，小数第2位が5のところである．整数値だけをとる離散的変量については，2つの隣り合った整数の算術平均を，階級の境界と考えることがある．

2. モーメント

X の n 個の測定値 $x_1, \cdots, x_n$ に対して，**平均値**あるいは**算術平均** $\bar{x}$ を

(1)　$$\bar{x}=\frac{1}{n}\sum_{\nu=1}^{n}x_\nu=\frac{1}{n}(x_1+\cdots+x_n)$$

で定義する．本章では ν について1から n までの和をこんご $\sum\limits_{\nu}$ または $\sum_\nu$ で表わすことにする．また，**分散** $s^2=s^2(x)$ を

(2)　$$s^2=\frac{1}{n}\sum_{\nu}(x_\nu-\bar{x})^2=\frac{1}{n}\{(x_1-\bar{x})^2+\cdots+(x_n-\bar{x})^2\}$$

で定義し，分散の非負の平方根 $s=s(x)=\sqrt{s^2}$ を**標準偏差**という．平均値 $\bar{x}$ は測定値全体の中心の位置を示すものとして，また分散 s^2 や標準偏差 s は，測定値全体のちらばりの程度を示すものとして用いられる．

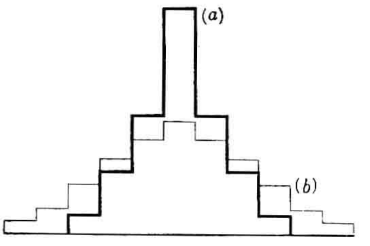


図 1　(a) と (b) は同じ平均値をもつ対称な柱状図で，(a) は分散が小，(b) は分散が大の場合を示す．

(a)　(例2参照)　測定値 $x_1, \cdots, x_n$ の分散はつぎの式から計算できる：

(3)　$$s^2=\frac{1}{n}\sum_{\nu}x_\nu{}^2-\bar{x}^2$$
$$=\frac{1}{n}(x_1{}^2+\cdots+x_n{}^2)-\bar{x}^2.$$

(b)　(例3参照)　2つの変量 X と U の間に $X=a+bU$ という関係があるとき，$x_\nu=a+bu_\nu$ ($\nu=1, \cdots, n$) とおけば，X と U の平均値の間，標準偏差の間につぎの関係がある：

(4)　$\bar{x}=a+b\bar{u}, \quad s(x)=|b|s(u).$

(以上)

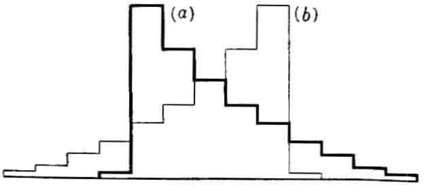


図 2　(a) と (b) は平均値も分散も同じ柱状図で，(a) は $c_3>0$，(b) は $c_3<0$ の場合を示す．

一般に原点のまわりの r 次の**モーメント** $a_r=a_r(x)$ を

(5)　$$a_r=\frac{1}{n}\sum_{\nu}x_\nu{}^r=\frac{1}{n}(x_1{}^r+\cdots+x_n{}^r),$$

平均値のまわりの r 次のモーメント $m_r=m_r(x)$ を

(6) $$m_r=\frac{1}{n}\sum_{\nu}(x_\nu-\bar{x})^r=\frac{1}{n}\{(x_1-\bar{x})^r+\cdots+(x_n-\bar{x})^r\}$$

で定義する．とくに $a_0=1$, $a_1=\bar{x}$, $m_0=1$, $m_1=0$, $m_2=s^2$ となる．さらに $s>0$ のとき，

(7) $$c_3=c_3(x)=m_3/s^3,$$

(8) $$c_4=c_4(x)=m_4/s^4$$

は，それぞれ度数分布の柱状図での**ひずみ**，**とがり**を表わすものとして利用される．ただし2つの柱状図のひずみやとがりを比較するのは，その2つの柱状図の面積を等しく，しかも標準偏差が横軸上の同じ長さの線分で表わされるようにした上である．

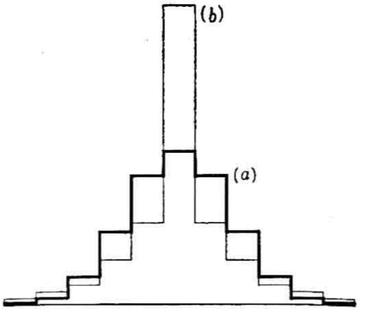


図 3 (a) と (b) は平均値も分散も同じ対称な柱状図で，(a) は c_4 が小，(b) は c_4 が大の場合を示す．

(c) (例4参照) 原点のまわりのモーメントと平均値のまわりのモーメントの間につぎの関係が成り立つ：

(9) $$m_2=a_2-\bar{x}^2,\quad m_3=a_3-3a_2\bar{x}+2\bar{x}^3,\quad m_4=a_4-4a_3\bar{x}+6a_2\bar{x}^2-3\bar{x}^4,$$

(10) $$m_r=\sum_{k=0}^{r}\binom{r}{k}a_{r-k}(-\bar{x})^k.$$

(d) (例5参照) 2つの変量 X と U の間に $X=a+bU$ $(b\neq0)$ という関係があるとき，$x_\nu=a+bu_\nu$ $(\nu=1, \cdots, n)$ とおけば，X と U のモーメントの間，c_3 の間，c_4 の間でつぎの関係が成り立つ：

(11) $$m_r(x)=b^rm_r(u),\quad c_3(x)=\mathrm{sgn}(b)c_3(u),\quad c_4(x)=c_4(u).$$ (以上)

［注意］ $\mathrm{sgn}(x)$ は，$x>0$ ならば 1，$x=0$ ならば 0，$x<0$ ならば -1 を示す記号である．

もとのデータのかわりに表1の度数分布表が与えられたならば，これからモーメントその他の特性値を計算するときには，つぎのように考える．すなわち，n 個のうち，x_1' が f_1 個，x_2' が f_2 個，$\cdots$，x_k' が f_k 個あるとする．したがってこのとき上にあげたいろいろな定義式はつぎのように書くことができる：

(12) $$\bar{x}=\frac{1}{n}\sum_{i=1}^{k}f_ix_i',\qquad s^2=\frac{1}{n}\sum_{i=1}^{k}f_i(x_i'-\bar{x})^2,$$

(13) $$a_r=\frac{1}{n}\sum_{i=1}^{k}f_ix_i'^r,\qquad m_r=\frac{1}{n}\sum_{i=1}^{k}f_i(x_i'-\bar{x})^r.$$

3. その他の特性値

与えられた測定値全体の中心の値を示すものとして，算術平均のほかに，メジアン，モード，幾何平均，調和平均などがある．**メジアン**とは，与えられた測定値のうちで大きさの順序からいって中央のものをいう．より正確には，$x_1, \cdots, x_n$ を大きさの順序に並べて $y_1 \leqq y_2 \leqq \cdots \leqq y_n$ としたとき，n が奇数ならば $y_{\frac{n+1}{2}}$ をメジアンといい，n が偶数ならば $(y_{\frac{n}{2}} + y_{\frac{n}{2}+1})/2$ をメジアンという．**モード**とは，与えられた測定値から等しい幅の階級を用いて度数分布をつくったとき，度数がもっとも多い階級の階級値のことをいう．正の測定値 $x_1, \cdots, x_n$ に対して，**幾何平均** $\bar{x}_G$，並びに**調和平均** $\bar{x}_H$ はそれぞれつぎの式で定義する：

$$\bar{x}_G = \left(\prod_{\nu=1}^{n} x_\nu\right)^{\frac{1}{n}} = \sqrt[n]{x_1 \cdots x_n}, \quad \bar{x}_H = n\Big/\sum_\nu \frac{1}{x_\nu} = n\Big/\left(\frac{1}{x_1} + \cdots + \frac{1}{x_n}\right). \tag{14}$$

また，ちらばりの度合を表わすものとしては，分散，標準偏差のほかに，範囲，平均偏差などがある．**範囲**とは，与えられた測定値のうちの最大値から最小値を引いたものである．$x_1, \cdots, x_n$ の最大値を $\max(x_1, \cdots, x_n)$，最小値を $\min(x_1, \cdots, x_n)$ とすれば，範囲は $\max(x_1, \cdots, x_n) - \min(x_1, \cdots, x_n)$ と表わすことができる．また $x_1, \cdots, x_n$ の算術平均を $\bar{x}$ とするとき，**平均偏差** m_d はつぎの式で定義される：

$$m_d = \frac{1}{n}\sum_\nu |x_\nu - \bar{x}| = \frac{1}{n}(|x_1 - \bar{x}| + \cdots + |x_n - \bar{x}|). \tag{15}$$

例　　題

1. $x_1, \cdots, x_n$ の算術平均を $\bar{x}$ とすれば，x_ν の算術平均 $\bar{x}$ からの偏差 $x_\nu - \bar{x}$ の和は0となることを証明せよ．

[解] $(x_1 - \bar{x}) + \cdots + (x_n - \bar{x}) = x_1 + \cdots + x_n - n\bar{x}$ となり，$\bar{x}$ の定義式 (1) からこれは0に等しい．

2. (a) を証明せよ．

[解]
$$\begin{aligned}
s^2 &= \frac{1}{n}\{(x_1 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2\} \\
&= \frac{1}{n}\{(x_1^2 - 2\bar{x}x_1 + \bar{x}^2) + \cdots + (x_n^2 - 2\bar{x}x_n + \bar{x}^2)\} \\
&= \frac{1}{n}(x_1^2 + \cdots + x_n^2) - \frac{2}{n}\bar{x}(x_1 + \cdots + x_n) + \frac{1}{n}n\bar{x}^2 \\
&= \frac{1}{n}(x_1^2 + \cdots + x_n^2) - 2\bar{x}^2 + \bar{x}^2 \\
&= \frac{1}{n}(x_1^2 + \cdots + x_n^2) - \bar{x}^2.
\end{aligned}$$

3. (b) を証明せよ.

［解］ $\bar{x}=\dfrac{1}{n}(x_1+\cdots+x_n)=\dfrac{1}{n}\{(a+bu_1)+\cdots+(a+bu_n)\}$

$$=\frac{1}{n}\{na+b(u_1+\cdots+u_n)\}=a+b\bar{u}.$$

つぎにこの結果を利用して,

$$s^2(x)=\frac{1}{n}\{(x_1-\bar{x})^2+\cdots+(x_n-\bar{x})^2\}$$

$$=\frac{1}{n}\{(bu_1-b\bar{u})^2+\cdots+(bu_n-b\bar{u})^2\}$$

$$=\frac{1}{n}b^2\{(u_1-\bar{u})^2+\cdots+(u_n-\bar{u})^2\}=b^2s^2(u).$$

両辺の非負の平方根をとれば, $s(x)=|b|s(u)$ となる.

4. (c) を証明せよ.

［解］ 39 頁の二項定理 (II. 2. a) によって $(x_\nu-\bar{x})^r$ を展開して, 例2と同様に計算すると, (10) が得られる.

$$m_r=\frac{1}{n}\sum_\nu(x_\nu-\bar{x})^r=\frac{1}{n}\sum_\nu\sum_{k=0}^{r}\binom{r}{k}x_\nu^{r-k}(-\bar{x})^k$$

$$=\sum_{k=0}^{r}\binom{r}{k}(-\bar{x})^k\frac{1}{n}\sum_\nu x_\nu^{r-k}=\sum_{k=0}^{r}\binom{r}{k}(-\bar{x})^k a_{r-k}$$

$$=a_r-\binom{r}{1}a_{r-1}\bar{x}+\binom{r}{2}a_{r-2}\bar{x}^2-\cdots+(-1)^{r-1}\binom{r}{r-1}a_1\bar{x}^{r-1}+(-1)^r\bar{x}^r.$$

ここでとくに $r=2, 3, 4$ とおけば, (9) が得られる.

5. (d) を証明せよ.

［解］ (b) によれば, $\bar{x}=a+b\bar{u}$ となる. よって,

$$m_r(x)=\frac{1}{n}\sum_\nu(x_\nu-\bar{x})^r=\frac{1}{n}\sum_\nu\{(a+bu_\nu)-(a+b\bar{u})\}^r$$

$$=\frac{1}{n}\sum_\nu b^r(u_\nu-\bar{u})^r=b^rm_r(u).$$

これを利用して,

$$c_3(x)=\frac{m_3(x)}{s^3(x)}=\frac{b^3m_3(u)}{|b|^3s^3(u)}=\mathrm{sgn}(b)c_3(u),\quad c_4(x)=\frac{m_4(x)}{s^4(x)}=\frac{b^4m_4(u)}{|b|^4s^4(u)}=c_4(u).$$

6. (b), (c), (d) を利用して, 等しい幅の階級に分類した度数分布表から, $\bar{x}$, s, c_3, c_4 を計算するための簡便法を考えよ.

［解］ $\bar{x}$ に近いと考えられる1つの階級値を x_0' とし, これより小さい階級値を順次 x_{-1}', x_{-2}', $\cdots$ とし, x_0 より大きい階級値を x_1', x_2', $\cdots$ とするように番号をつけかえる. これに対応して度数も $\cdots$, f_{-2}, f_{-1}, f_0, f_1, f_2, $\cdots$ とする. 階級の幅を b (>0) とし, 任意の階級値を $x_i'=x_0+bu_i'$ とおけば, $u_i'=i$ は簡単な整数となる. そこで $X=x_0'+bU$ によって与えられる変量 U について, 表2にもとづいて原点のまわりのモーメント $\bar{u}$, $a_2(u)$, $a_3(u)$, $a_4(u)$ を求め, つぎに (c) を U に対して適用して, U の

平均値のまわりのモーメント $m_2(u)$, $m_3(u)$, $m_4(u)$ を求める．これより $s(u)$, $c_3(u)$, $c_4(u)$ が計算できるから，あとは (b), (d) を利用して，$\bar{x}=x_0'+b\bar{u}$, $s=s(x)=bs(u)$, $c_3=c_3(x)=c_3(u)$, $c_4=c_4(x)=c_4(u)$ が得られる．

表 2

x	f	u	uf	u^2f	u^3f	u^4f
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
x_{-2}'	f_{-2}	-2	$-2f_{-2}$	$(-2)^2f_{-2}$	$(-2)^3f_{-2}$	$(-2)^4f_{-2}$
x_{-1}'	f_{-1}	-1	$-1f_{-1}$	$(-1)^2f_{-1}$	$(-1)^3f_{-1}$	$(-1)^4f_{-1}$
x_0'	f_0	0	0	0	0	0
x_1'	f_1	1	$1f_1$	1^2f_1	1^3f_1	1^4f_1
x_2'	f_2	2	$2f_2$	2^2f_2	2^3f_2	2^4f_2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
計	n		$\sum uf$	$\sum u^2f$	$\sum u^3f$	$\sum u^4f$
計 $\div n$			$\bar{u}$	$a_2(u)$	$a_3(u)$	$a_4(u)$

$$m_2(u)=a_2(u)-\bar{u}^2,$$
$$m_3(u)=a_3(u)-3a_2(u)\bar{u}+2\bar{u}^3,$$
$$m_4(u)=a_4(u)-4a_3(u)\bar{u}+6a_2(u)\bar{u}^2-3\bar{u}^4;$$
$$s(u)=\sqrt{m_2(u)},$$
$$c_3(u)=m_3(u)/s^3(u),$$
$$c_4(u)=m_4(u)/s^4(u);$$
$$\bar{x}=x_0'+b\bar{u},$$
$$s=s(x)=bs(u),$$
$$c_3=c_3(x)=c_3(u),$$
$$c_4=c_4(x)=c_4(u).$$

7.　n 個の測定値 $x_1, \cdots, x_n$ が与えられたとき，$\sum_\nu(x_\nu-a)^2$ を最小にする実数 a は，$\bar{x}$ となることを証明せよ．

〔解〕 $\sum_\nu(x_\nu-a)^2=\sum_\nu\{(x_\nu-\bar{x})+(\bar{x}-a)\}^2$

$$=\sum_\nu(x_\nu-\bar{x})^2+2(\bar{x}-a)\sum_\nu(x_\nu-\bar{x})+n(\bar{x}-a)^2.$$

ここで第3辺の第2項は例1によって0となるから，

(16)　$$\sum_\nu(x_\nu-a)^2=\sum_\nu(x_\nu-\bar{x})^2+n(\bar{x}-a)^2.$$

(16) の右辺の第1項は a に関係せず，第2項は $a=\bar{x}$ のときに0で，その他のときは正となる．よって (16) は $a=\bar{x}$ で最小となる．

〔別解〕 $\sum_\nu(x_\nu-a)^2$ は a に関する2次式で，a^2 の係数は正である．よって

$$0=\frac{d}{da}\sum_\nu(x_\nu-a)^2=-2\sum_\nu(x_\nu-a)=-2n(\bar{x}-a)$$

より a を求めれば，$a=\bar{x}$ となり，この a が $\sum_\nu(x_\nu-a)^2$ を最小にする．

〔注意〕 この例の結果から，任意の実数 a について，$\frac{1}{n}\sum_\nu(x_\nu-a)^2\geqq s^2$ となる．

問　　題〔A〕

1.　5つの測定値 2, 3, 3, 5, 7 の平均値 $\bar{x}$ と，平均値のまわりの 4 次までのモーメント m_2, m_3, m_4 を，(1) と (6) によって計算せよ．

2.　前問の測定値から，原点のまわりのモーメントを計算し，その値を利用して (9) によって m_2, m_3, m_4 を計算せよ．

3.　つぎの表3は500人の学生の試験の成績を，15点の幅の階級に分類した度数分布

表である．点数の平均値 $\bar{x}$，標準偏差 s，ひずみ c_3，とがり c_4 を計算せよ．

表 3

点数	15〜29	30〜44	45〜59	60〜74	75〜89	90〜104	105〜119	120〜134	135〜149	150〜164	165〜179	計
度数	3	11	29	53	100	119	79	42	33	20	11	500

4.　$\min(x_1, \cdots, x_n) \leqq \bar{x} \leqq \max(x_1, \cdots, x_n)$ を証明せよ．

5.　n 個の正の測定値 $x_1, \cdots, x_n$ の幾何平均の対数は，各測定値の対数の算術平均に等しいことを証明せよ．

6.　2つの数 x_1, x_2 の分散 s^2 は $s^2=(x_1-x_2)^2/4$ となることを証明せよ．

7.　n 個の数 $x_1, \cdots, x_n$ の分散 s^2 は $s^2=\dfrac{1}{n^2}\sum_{1\leqq\nu<\lambda\leqq n}(x_\nu-x_\lambda)^2$ となることを証明せよ．

8.　$x_1, \cdots, x_m$ の平均値を $\bar{x}$，分散を $s_x{}^2$ とし，$y_1, \cdots, y_n$ の平均値を $\bar{y}$，分散を $s_y{}^2$ とすると，$m+n$ 個の値 $x_1, \cdots, x_m, y_1, \cdots, y_n$ の平均値は $\dfrac{m\bar{x}+n\bar{y}}{m+n}$，分散は $\dfrac{ms_x{}^2+ns_y{}^2}{m+n}+\dfrac{mn}{(m+n)^2}(\bar{x}-\bar{y})^2$ となることを証明せよ．

9.　n 個の数 $1, 2, \cdots, n$ の平均値と平均値のまわりの4次までのモーメントを求めよ．

〔B〕

10.　n 個の値 $x_1, \cdots, x_n$ の平均偏差を m_d，標準偏差を s とすれば，$m_d \leqq s$ となることを証明せよ．さらに $m_d=s$ となるのはどのような場合かを吟味せよ．

11.　$x_1, \cdots, x_n$ に新しいデータ x' をつけ加えると，平均値と分散はどう変化するか．

12.　n 個の値 $x_1, \cdots, x_n$ はいずれも区間 $[a, b]$ にあるとき，$x_1, \cdots, x_n$ の分散 s^2 の最大値は，n が偶数ならば $(b-a)^2/4$，n が奇数ならば $(n^2-1)(b-a)^2/4n^2$ となることを証明せよ．

13.　n 個の測定値 $x_1, \cdots, x_n$ が与えられたとき，$\sum_\nu|x_\nu-a|$ を最小にする a の値は，与えられた測定値を大きさの順序に並べかえて $y_1 \leqq y_2 \leqq \cdots \leqq y_n$ とするとき，n が奇数 $n=2m-1$ ならば y_m であり，n が偶数 $n=2m$ ならば，区間 $[y_m, y_{m+1}]$ のすべての値となることを証明せよ．

14.　n 個の値 $x_1, \cdots, x_n$ が与えられたとき，自然数 k に対して $\sum_\nu(x_\nu-c)^{2k}$ を最小にする c の値 c_k はただ1つ存在して，$\lim_{k\to\infty}c_k=\dfrac{1}{2}\{(\max(x_1, \cdots, x_n)+\min(x_1, \cdots, x_n)\}$ となることを証明せよ．

解答

1. 計算は表4にもとづいて行なわれる．第1列の合計は 20 であるから，$\bar{x}=20/5=4$ となる．第2列は，各測定値から $\bar{x}$ を引いたもの $x-\bar{x}$ であり，第3〜5列は，第2列の値の2乗〜4乗である．第3〜5列の合計を5で割ったものが，それぞれ m_2, m_3, m_4 となる．

表 4

	x	$x-\bar{x}$	$(x-\bar{x})^2$	$(x-\bar{x})^3$	$(x-\bar{x})^4$
	2	-2	4	-8	16
	3	-1	1	-1	1
	3	-1	1	-1	1
	5	1	1	1	1
	7	3	9	27	81
計	20	0	16	18	100
÷5	4		3.2	3.6	20
	‖		‖	‖	‖
	$\bar{x}$		m_2	m_3	m_4

2. 原点のまわりのモーメントの計算は表5にもとづいて行なわれる．これより，

表 5

	x	x^2	x^3	x^4
	2	4	8	16
	3	9	27	81
	3	9	27	81
	5	25	125	625
	7	49	343	2401
計	20	96	530	3204
÷5	4	19.2	106	640.8
	‖	‖	‖	‖
	$\bar{x}$	a_2	a_3	a_4

$m_2=19.2-4^2=3.2$,

$m_3=106-3\times 19.2\times 4+2\times 4^3$

$\quad =106-230.4+128=3.6$,

$m_4=640.8-4\times 106\times 4+6\times 19.2\times 4^2-3\times 4^4$

$\quad =640.8-1696+1843.2-768=20$.

[注意] この例では $\sum_\nu x_\nu=20$ を $n=5$ で割るときにきれいに割り切れたので，本問の方法による計算よりも，前問の方法による計算のほうが簡単であった．しかしこの割り算は割り切れないのがふつうであって，そのようなときには，本問の方法による方が，計算は簡単になる．なお $\bar{x}$ と m_2 だけ，または $\bar{x}$ と s だけを求めるときに，表5の第3〜4列の計算が不要になることはいうまでもない．

3. 階級の境界は 14.5, 29.5, 44.5, $\cdots$ となっているものとして，階級値 22, 37, 52, $\cdots$ を算出する．最初から6番目の階級値 97 を u の原点にとって，$x=97+15u$ とおくと，階級値 x に対しては，u は簡単な整数となる．そこで表6にしたがって，各階級について uf, u^2f, u^3f, u^4f を計算して，その合計をつくり，これを $n=500$ で割れば，それぞれ $\bar{u}$, $a_2(u)$, $a_3(u)$, $a_4(u)$ となる．

$m_2(u)=3.926-0.090^2=3.918$,

$m_3(u)=3.150-3\times 3.926\times 0.090+2\times 0.090^3=2.091$,

$m_4(u)=46.814-4\times 3.150\times 0.090+6\times 3.926\times 0.090^2-3\times 0.090^4=45.871$,

$\bar{x}=97+15\times 0.090=98.35$, $\qquad s=15\times\sqrt{3.918}=29.691$,

$c_3=2.091/3.918^{3/2}=0.270$, $\qquad c_4=45.871/3.918^2=2.988$.