

東京理科大学教授 理博 国沢清典編

確率統計演習 2

統計

培風館

確率統計演習 2

統計

東京理科大学教授 理学博士

国沢清典 編

培風館

編者および執筆者紹介

国 沢 清 典
くに さわ きよ のり

1939 年 大阪大学理学部数学科卒業

1949 年 理学博士

現 職 東京工業大学名誉教授

主要著書

近代確率論 (岩波書店)

OR のための情報理論 (日本科学技術連盟)

OR による在庫管理 (日本事務能率協会)

ビジネス数学入門 (共著, 東洋経済新報社)

現代統計学 (上・下) (共著, 広川書店)

オペレーションズ・リサーチ演習
(共著, 培風館)

情報理論の進歩 (共編, 岩波書店)

オペレーションズ・リサーチ入門
(共著, 広川書店)

真 壁 肇 [第2巻 執筆]
ま かべ はじめ

1951 年 東京工業大学卒業

1962 年 工学博士

現 職 東京工業大学教授

主要著書

工業経営と装置計画 (共著, ダイアモンド社)

羽 鳥 裕 久 [第1巻 執筆]
は とり ひろ ひさ

1960 年 東京工業大学大学院卒業

1960 年 理学博士

現 職 東京理科大学教授

主要著書

初等確率論 (共著, 培風館)

確率論とその応用Ⅱ

(共訳, 紀伊国屋書店)

© 国 沢 清 典 1966

昭和41年6月10日 初版発行

昭和56年1月10日 初版第14刷発行

確率統計演習 2

統 計

編 者 国沢清典

発行者 山本健二

発 行 所 株式会社 培 風 館

東京都千代田区九段南 4-3-12・郵便番号 102
電話 (03) 262-5256 (代表)・振替東京 4-44725

定 価 ￥ 1600.

渥美活版組版・政弘印刷・三水舎製本

編者の承認をえて検印を省略しました

3033-0810-6955

ま え が き

統計学は確率論と同時代の17世紀の中葉、ドイツとイギリスにおこったと伝えられている。それはドイツ国勢学派とイギリス政治算術学派によって代表される J. Graunt や J. D. Süßmilch により、出生、婚姻、死亡などの集計整理のうちに、一定の統計的法則が存在することを実証し、「神の秩序」の賜物としたのである。J. Bernoulli は古典確率論の立場から、この法則を数式化していわゆる大数の法則を証明した。この大数の法則の発見により、それ以後の統計学の発展は全く方向づけられてしまったといって過言ではない。すなわち、大数の法則を守護神として、資料をたくさん集めさえすれば、ますます安定した結論が得られたわけである。つまり、資料のなかに含まれる偶然性を除去するために、大量の資料を集めさえすればよかったわけである。それによって安定した結論が得られることは大数の法則が保証してくれたわけである。

1894年 C. Darwin が「種の起原」を刊行し、このなかで進化論を発表した。進化論は完全に資料の整理分類の勝利であり、これによりそれ以前まではもっぱら統計学が政治、経済、社会に応用されていたのであるが、自然科学方面にもその重要性が認識される結果となった。これを機縁として、F. Galton, F. Edgeworth また K. Pearson などにより自然科学、とくに、生物学、遺伝学に応用されていった。

1920年頃から統計学は大きな転機を迎えた。それは統計学の守護神を大数の法則から偶然性におきかえたところに特長がある。大数の法則の庇護から離脱して偶然性のなかに統計学を投げこんだのである。これをなしとげた学者は反骨精神の旺盛な R. A. Fisher であった。自然科学のなかには、資料をたくさん集めれば集めるほど、不純な要因が混入して安定な結論を得ることが困難となることを彼は農事試験によって体験した。そこで彼は少数標本で結論を得ることのできる統計学を建設していったことは有名な話である。

統計学の発展の歴史をふりかえってながめてみるにつけ、統計学という学問の特長をひしひしと感ぜざるをえない。それは、統計学は現実界と関連なしには成立しえないことである。したがって統計学は時代の影響を受け、いつも時代の忠実な従僕たる宿命をもってきた。この点が数学とは異なっているわけで

ある。数学は絶対的なものであり、その真実性はいつの時代でも変わるところがない。

さて、本演習書は第1巻の確率について大学初年級用、また一般社会人の統計学を勉強される方々のための参考書として編さんしたものである。統計学がその時代の忠実なサーバントであり、現実界と切り離しては意味のない点に鑑みて、例題、演習問題にしても、もっぱら、現実の具体的な問題から取材した。そのために個々の数値計算から、その処理方法にわたって、詳細に説明するように努めたわけである。とにかく統計学は現実上の問題をモデルとして考えると、きわめて理解しやすいものである。本書はこのような観点から統計学の特長を生かして、自然の姿で統計学が学べるように編さんしたのである。

学び方について一言するならば、初心者は第2章と第5章はやや理論的であるから、ここはざっととばし読みして、手法に自信がついてから第2章と第5章をさらに詳しく勉強すると便利である。なお、第1巻と同じく、問題にはすべて解答を付した。諸君諸賢のご批判を得て将来より完全なものにしてゆきたい念願である。

本書の編集、校正にあたっては多くの方々のご援助を受けた。まず付表に関しては、一般に公刊されている諸表のうち、本演習書を使ううえに必要と思われるものについて選択修正して利用させていただいた。またとくに、二項確率紙については、財団法人日本科学技術連盟のご好意をいただいた。ここに厚くお礼申し上げる。また、横浜国立大学 工学部助手 田口淑子、神沼靖子の両氏より校正および問題解答の計算をチェックする等について多大の援助を受けた。さらに、培風館森平勇三氏、藤野晁氏、松村良子氏をわずらわした点、多大である。ここに心より感謝の辞をささげたい。

1966年4月

編 者

目 次

まえがき

1. データのまとめ方 1~25

1-1 ヒストグラムの作り方と標本平均, 分散の求め方

- 1. データの表示法..... 1
- 2. ヒストグラムの作り方..... 2
- 3. 度数分布の量的表示法..... 3
- 4. 級分けされたときの平均値, 分散および積率..... 5
- 5. 要項4における計算の簡便法..... 5

1-2 度数分布と母集団分布

- 6. 度数分布と母集団分布.....17

1-3 相 関 関 係

- 7. 2次元正規母集団.....18
- 8. 級分けされないときの相関係数.....18
- 9. 級分けされたときの相関係数.....19
- 10. r の簡便計算法.....21

2. 統計的推定 26~64

2-1 推定法と標本平均および標本分散の性質

- 1. 推 定 法.....26
- 2. 標本変量, 統計量と推定量.....27
- 3. 母平均と母分散の推定.....28

2-2 点 推 定 法

- 4. 最 適 性.....33

2-3 推定量の求め方

- 5. 最 尤 法.....39
- 6. モーメント法.....40

2-4 推定量の性質

- 7. 最適推定量.....61
- 8. 最尤推定量.....61

3. 区間推定法

65~102

3-1 平均値 μ に関する区間推定

- 1. 区間推定とは.....65
- 2. 正規母集団 $N(\mu, \sigma^2)$ の母平均 μ の区間推定法66
- 3. 二つの正規母集団の母平均の差の区間推定法.....67

3-2 分散 σ^2 に関する区間推定

- 4. 正規母集団の母分散 σ^2 の区間推定法.....73

3-3 百分率 p に関する区間推定

- 5. 二項母集団の母百分率 p の区間推定76
- 6. 二つの二項母集団の母百分率の差の区間推定.....77

3-4 信頼区間の作り方

- 7. 信頼区間の定め方.....81

**3-5 指数分布および Poisson 分布の母平均に関する区間推定,
その他の区間推定**

- 8. 指数分布の母平均 μ の区間推定法91
- 9. Poisson 分布の母平均 λ の区間推定法.....91
- 10. 一様分布の幅の区間推定法.....92

3-6 母数によらない推定

- 11. 分布関数の区間推定..... 100

4. 統計的検定

103~157

4-1 統計的検定法

- 1. 仮 説 検 定 103
- 2. 二つの誤り方..... 104
- 3. 仮 説..... 104

I. 計量値に関する検定**4-2 平均値 μ に関する検定**

- 4. 正規母集団 $N(\mu, \sigma^2)$ の母平均 μ に関する検定 106

4-3 分散 σ^2 に関する検定

- 5. 正規母集団 $N(\mu, \sigma^2)$ の母分散 σ^2 に関する検定..... 112
- 6. 二つの正規母集団の母分散の等しいことをしらべる検定..... 113

4-4 平均値の差 δ に関する検定

- 7. 二つの正規母集団の平均値の差に関する検定..... 116

目次	v
8. Welch の検定	117
II. 検定方式の定め方	
4-5 検定方式の作り方 (その I)	
9. 検定方式の作り方 (I)	125
10. 検定方式の作り方 (II)	127
4-6 検定方式の作り方 (その II)	
11. 尤度比検定法	131
4-7 寿命分布の平均値 μ に関する検定	
12. 指数分布の母平均 μ に関する検定	134
III. 計数値に関する検定	
4-8 百分率 p に関する検定	
13. 百分率に関する検定 (I)	139
14. 百分率に関する検定 (II)	140
4-9 適合度の検定	
15. 適合度の検定 (χ^2 -検定)	145
4-10 分割表による独立性の検定	
16. 分割表による検定	148
4-11 Poisson 分布の平均値に関する検定および Thorndike 曲線を用いた検定	
17. Poisson 分布の母平均 λ に関する χ^2 -検定	154
18. Thorndike 曲線を用いた検定	154
5. 標本分布論と標本調査	158~194
I. 標本分布論	
5-1 正規母集団からの標本変量	
1. 正規標本論	158
2. $\bar{X}$ の分布	159
3. S^{*2} の分布	159
4. S^2 の分布	159
5. T の分布	160
6. F の分布	160
7. 正規, t , χ^2 および F -分布の間の関係	160

5-2 Cochran の定理とその応用

- 8. Cochran の定理 163
- 9. $\bar{X}$ と S^2 の独立性 163

5-3 いろいろな分布をした母集団からの標本変量の性質

- 10. 二項分布の正規近似 172
- 11. 二項分布と F -分布 172
- 12. Poisson 分布と χ^2 -分布および正規近似 173
- 13. 指数分布と χ^2 -分布 173
- 14. ガンマ分布 173

5-4 順序統計量の分布

- 15. 順序統計量 176
- 16. 最大 (最小) 標本値の分布 177
- 17. 中央値の分布 177
- 18. 範囲の分布 177
- 19. 許 容 限 界 177
- 20. 平均範囲の分布へのカイ近似 178
- 21. $\bar{X}$ と R の独立性 179

II. 標 本 調 査

5-5 標 本 調 査

- 22. 任意抽出法 185
- 23. 層別抽出法 186
- 24. 2 段抽出法 187

6. 分 散 分 析

195~241

6-1 分散分析表の作り方

- 1. 因子と水準 195
- 2. 構 造 模 型 196
- 3. 1 元配置法 196
- 4. 2 元配置法 198
- 5. 分散分析表の作り方の手順の一例 201
- 6. 3 元配置法 204
- 7. Yates の簡便法による 2^n 型実験の解析法 206

6-2 母数の推定法

- 8. 点 推 定 206
- 9. 区 間 推 定 207

6-3 変量模型のときの分散分析表のとり扱い方

10. 変量模型	228
----------	-----

6-4 ラテン方格とグレコラテン方格および直交配列表との関係

11. ラテン方格とグレコラテン方格	232
12. ラテン方格を用いた分散分析	233
13. グレコラテン方格を用いた分散分析	234
14. 直交配列表とラテン方格との関係	234

7. 相関関係と回帰分析 242~266

7-1 相関係数とその推定, 検定

1. 単相関係数	242
2. 相関係数の検定および推定	243
3. 回帰直線	243
4. 回帰係数の検定および推定	244
5. 重相関係数と偏相関係数	245
6. 重相関係数と偏相関係数の検定	246
7. 回帰平面式	246

7-2 回帰分析

8. 単回帰 I	247
9. 単回帰 II	248
10. 重回帰	249

8. 管理図, 簡易解析法 267~279

—品質管理におけるいろいろな手法—

8-1 管理図

1. 管理図	267
2. $\bar{x}$ - R 管理図	267
3. p 管理図	268

8-2 簡易解析法

4. 管理図による平均値の差の検定	269
5. 二項確率紙	270
6. 二項確率紙の原理	271
7. 二項確率紙を用いた統計的手法	271

付 表 280~287

1. 正規分布	280
---------	-----

2. t -分 布	280
3. χ^2 -分 布	281
4. F -分 布	282
5. z -変 換	284
6. 乱 数 表	285
7. Thorndike 曲線	286
8. 二項確率紙	287

索 引	288
-----------	-----

重要記号一覧	292
--------------	-----



〔 1 卷 目 次 〕

1. 事象と確率
 2. 確率変数と分布
 3. 平均値, 分散
 4. 変数変換と和の分布
 5. 積率と積率母関数
 6. 大数の法則と中心極限定理
 7. 大数の強法則
- 付録 中心極限定理の別証明

1

データのまとめ方

本章では、母集団から抽出された**標本(データ)**を用いて母集団の分布形やその特性値を推測する手法を概説する。

ある会社で製造した肥料の窒素含有率を知るために、データとして $x_1, x_2, \dots, x_{100}$ なる測定値をえたとすれば、これらを適当な方法で整理してこの会社で現在作られている肥料の品質を知ることが統計学の仕事となる。母集団とは、この会社の、この工程においてつぎつぎと作り出されている無限に近い多くの肥料の窒素含有率の値全体をいう。したがって、しばしば**無限母集団**ともよばれる。 $x_1, x_2, \dots, x_{100}$ は**大きさ 100 の標本**(または**標本値**)とよばれるが、通俗的にデータとか資料ともいわれる。

この章では、統計のいちばん簡単な手法として、データをどのように整理したら母集団の実態がつかめるかを考え、さらに、期待値との関係 (1-2) や相関関係 (1-3) まで話を進める。

1-1 ヒストグラムの作り方と標本平均、分散の求め方

要項

1. データの表示法 通常は、二つに大別して述べられている。

a. 図的表示法 これは以下に述べるように、ヒストグラム(柱状図表ともいう)を作って、データを図的表現によって処理し、母集団分布の形を推定する方法である。

b. 量的表示法 データを計数的に処理して、母集団分布の特性値を推定する方法である。

2. ヒストグラムの作り方

これはつぎの手順によって実施される：

- (1) データを測定順に 10 個ぐらいずつ一つの組にしてまとめる。
- (2) データの最大値 $x_{\max}$ と最小値 $x_{\min}$ を求める（このためには、まず各組の最大値、最小値にそれぞれ *、√印をつけ、この中から最大、最小なものにそれぞれ ㊶、㊷をつける）。
- (3) $x_{\max} - x_{\min}$ ($=R$ ：これを範囲という) を求め、これを級の数 k でわる。
 k は通常 10 前後とされているが、データの数 N にも関係して
 $50 < N \leq 100$ のとき $8 \leq k \leq 12$, $100 < N$ のとき $10 \leq k \leq 20$
とされる。
- (4) $(x_{\max} - x_{\min}) \div k$ を計算し、この値の近くの適当な値 h を級の幅とする。この級の幅は測定単位の整数倍になるのがのぞましい。
- (5) 幅 h の級を小さい方から順次きめてゆく。級の境界値は
 - (i) この上に測定値がおちないように、測定値より単位が 1 桁下のものをとるようにし、
 - (ii) しかも、両端の級においては $x_{\max}$ と $x_{\min}$ がほぼその中程にくるようにすると、かたよりが少なくなる。
- (6) 各級に測定値をわりふる。これを表 1・1 にあるような度数表に作る。

表 1・1

級の番号 no.	級の境界値 $a_{i-1} \sim a_i$	わりふり チェック	度 数 f_i
1	$a_0 \sim a_1$	//	2
2	$a_1 \sim a_2$	////	4
3	$a_2 \sim a_3$	/// /// ///	13
4	$a_3 \sim a_4$	/// /// /// //	17
⋮	⋮	⋮	⋮
k	$a_{k-1} \sim a_k$	///	3
計			N (データの数)

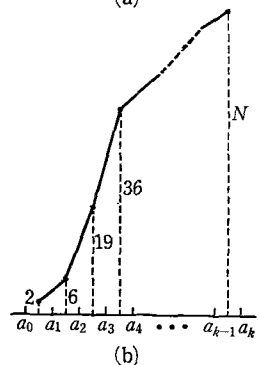
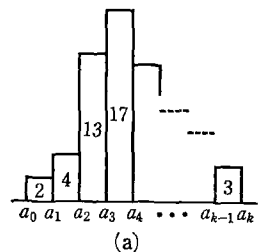


図 1・1 (a) ヒストグラム
(b) 棒状累積分布図

(7) 度数表の度数を高さとした柱を各級の upper に立て、ヒストグラムを作る (図 1・1, (a))。

(8) ヒストグラムの各柱の頭をなるべくなめ

らかな曲線で結ぶと、度数分布曲線とよばれるものができる（要項6を見よ）。

いま、各級に対応する度数を高さとしてヒストグラムを作ったが、各級にそれ以下の級の度数の累計を柱に立てたり、線で結んだりして柱状、または棒状累積分布図を作ることもある（図1・1, (b)）。

3. 度数分布の量的表示法 要項2では、度数分布による母集団分布への図形的接近について述べたが、ここではこの母集団分布のある特性値をデータによって推測することを考える。分布の特性値として、まず、とり上げられるべきものに、**中心的傾向とバラツキ**がある。

a. 中心的傾向をあらわすもの

(イ) 平均値

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} \quad (1)$$

これを具体的に計算するときには、各測定値 x_i ($i=1, 2, \dots, n$) を

$$u_i = (x_i - x_0) \div h$$

として、変換しておいて

$$\bar{u} = \frac{u_1 + u_2 + \cdots + u_n}{n}$$

を求め、これを元の単位にもどして

$$\bar{x} = \bar{u} \cdot h + x_0 \quad (2)$$

とすればよい（例題2を見よ）。ここに x_0 を**仮の平均値**という。 x_0 は u_i が簡単な数になるように適当にきめる。

(ロ) 中央値（メジアンともいわれる）

標本を大きさの順にならべ $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n-1)} \leq x_{(n)}$ としたとき

$$\begin{aligned} \text{中央値 } \tilde{x} &= x_{(\frac{n+1}{2})}, & n: \text{奇数} \\ &= \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, & n: \text{偶数} \end{aligned} \quad (3)$$

となる。つまり、 n が奇数ならばちょうどまん中にくる数、偶数ならばまん中にある2数の平均となる（この他に**ミッド・レンジ**: $(x_{(1)} + x_{(n)})/2$ を用いることもある）。

b. バラツキ

(イ) 分散

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \cdots + (x_n - \bar{x})^2}{n}$$

または

$$= \frac{x_1^2 + x_2^2 + \cdots + x_n^2}{n} - \bar{x}^2 \quad (4)$$

具体的に計算するときには a, (イ) に述べた u_i を用いて

$$s_u^2 = \frac{u_1^2 + u_2^2 + \cdots + u_n^2}{n} - \bar{u}^2$$

を計算し、これを元の測定値の単位にもどして

$$s^2 = s_u^2 \cdot h^2 \quad (5)$$

とすればよい。

(ロ) 標準偏差

$$s = \sqrt{\frac{x_1^2 + \cdots + x_n^2}{n} - \bar{x}^2} \quad (6)$$

これは、 s^2 は次元が元の測定値の平方であることを、平方根に開いてとり扱い易くしたものである。しかし、理論的には分散のほうが扱い易いので、統計学では、むしろ s^2 のまま用いていることが多い。

(ハ) 範囲

$$R = x_{\max} - x_{\min} \quad (7)$$

つまり標本値の最大なものから、最小値を引いたもので、これはまた a, (ロ) の記号を用いて、 $R = x_{(n)} - x_{(1)}$ ともかける。

c. 積率 (モーメント)

a, b の他に、さらに

$$\text{原点のまわりの } \nu \text{ 次のモーメント: } m_\nu' = \frac{x_1^\nu + x_2^\nu + \cdots + x_n^\nu}{n}$$

や 平均値のまわりの ν 次のモーメント: (8)

$$m_\nu = \frac{(x_1 - \bar{x})^\nu + (x_2 - \bar{x})^\nu + \cdots + (x_n - \bar{x})^\nu}{n}$$

を計算することもある。ここで、とくに

$$\bar{x} = m_1', \quad s^2 = m_2$$

であって、一般に

$$m_\nu = m_\nu' - \binom{\nu}{1} \bar{x} m_{\nu-1}' + \binom{\nu}{2} \bar{x}^2 m_{\nu-2}' - \cdots + (-1)^{\nu} \binom{\nu}{\nu} \bar{x}^\nu$$

なる関係が成立する。

4. 級分けされたときの平均値, 分散および積率 各級の境界値を $a_0, a_1, \dots, a_k$ とし, その代表値を

$$\xi_i = \frac{a_{i-1} + a_i}{2} \quad (i=1, 2, \dots, k)$$

としよう (図 1.2)。

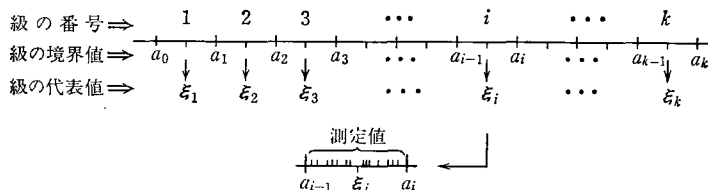


図 1.2

N 個の測定値のうち i 番目の級に入った測定値の個数つまり度数を f_i とすれば, この f_i 個の値はすべて ξ_i に等しいと近似的にみなせるから, つぎのように平均値その他を計算できる。

(イ) 平均値

$$\bar{x} = \frac{1}{N} (\xi_1 f_1 + \xi_2 f_2 + \dots + \xi_k f_k)$$

(ロ) 分散

$$\begin{aligned} s^2 &= \frac{1}{N} \{ (\xi_1 - \bar{x})^2 f_1 + (\xi_2 - \bar{x})^2 f_2 + \dots + (\xi_k - \bar{x})^2 f_k \} \\ &= \frac{1}{N} \sum_{i=1}^k \xi_i^2 f_i - \bar{x}^2 \end{aligned}$$

(ハ) ν 次の原点のまわりおよび平均値のまわりの積率

$$\begin{aligned} m'_\nu &= \frac{1}{N} \sum_{i=1}^N \xi_i^\nu f_i \\ m_\nu &= \frac{1}{N} \sum_{i=1}^N (\xi_i - \bar{x})^\nu f_i \end{aligned}$$

(ニ) ひづみ (歪度) および とがり (尖度)

ひづみ: $a = m_3/s^3$, とがり: $b = m_4/s^4 - 3$ * (ともに無名数である)

5. 要項 4 における計算の簡便法 各 ξ_i を変数変換して

* とがりとは $b' = m_4/s^4$ として定義されることも多い。 $b = m_4/s^4 - 3$ としているのは正規分布について求めたとがり $\beta' = \mu_4/\sigma^4$ が 3 だからである (表 1.14 参照)。

$$u_i = (\xi_i - \xi_0) \div h$$

とする。ただし、 ξ_0 はほぼ中央に位する級の代表値 ξ_i にとり、これを仮の平均値としたものである。 h は級の幅である。 ξ_i を u_i に変換したのは、 ξ_0 を 0 とし、その上、下の級に順々に整数 $-1, -2, \dots; 1, 2, \dots$ をわりふるためである（計算の具体的方法は例題 4 を見よ）。このようにすると、平均、分散などはつぎのように求められる。

$$(イ) \text{ 平均値: } \bar{u} = \frac{1}{N} \sum_{i=1}^k u_i f_i = (\sum u f) \div N$$

これを元の単位にもどせば

$$\bar{x} = \bar{u} \cdot h + \xi_0$$

$$(ロ) \text{ 分散: } s_u^2 = \frac{1}{N} \sum_{i=1}^k u_i^2 f_i - \bar{u}^2 = (\sum u^2 f) \div N - \bar{u}^2$$

これを元の単位にもどして

$$s^2 = s_u^2 \cdot h^2$$

(ハ) ν 次のモーメント

$$m_{\nu(u)}' = \frac{1}{N} \sum_{i=1}^k u_i^\nu f_i = (\sum u^\nu f) \div N$$

元の単位にもどせば

$$m_\nu' = m_{\nu(u)}' \cdot h^\nu$$

となり、さらにこれを用いて m_ν を計算できる。

$$m_\nu = h^\nu \left\{ m_{\nu(u)}' - \binom{\nu}{1} \bar{u} m_{\nu-1(u)}' + \binom{\nu}{2} \bar{u}^2 m_{\nu-2(u)}' - \dots + (-1)^\nu \binom{\nu}{\nu} \bar{u}^\nu \right\}$$

(ニ) その他の量

中心的傾向を知る他の量

中央値: $\sum_{i=1}^{m-1} f_i < \frac{N}{2}, \sum_{i=1}^m f_i > \frac{N}{2}$ なる m を求め、これを用いて

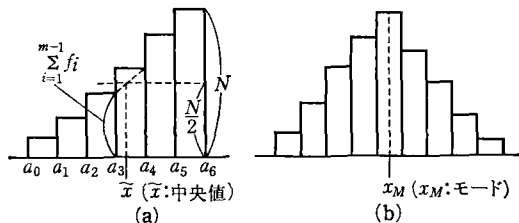


図 1.3 (a) 柱状累積分布図 (b) ヒストグラム