

母集団の縮図としての度数分布図

勝間昭一郎・橋本咲子

Frequency Distribution Diagram expressed as Epitome of Population

Shoichiro KATSUMA and Sakiko HASHIMOTO

keywords : Frequency Distribution, Population, Moment

I. は し が き

凡そ物事を実証的に考察し、科学的に処理しようとするとき、統計学の助けを借りないでできる場合は極めて稀である。それなればこそ、自然科学は言うに及ばず、社会科学・人文科学の方面においても、その必要性が年と共に増大しつつある。

ところで、統計学はこれを大きく記述統計学と推測統計学の二つに分類することができるが、その何れの場合においても、データの収集・データの整理・データの解析という過程をとる。そこで、ここではデータの整理に焦点をあて、データをどのように縮約すれば母集団の適切な縮図としての度数分布図が得られるかといった基礎的な問題について考察する。なお、度数分布図の描き方には i) データの分布の概形としてのもの と ii) 各階級の度数(頻度)に着目したもの の二通りの方法があるが、以下では i) の立場から考察する。

II. 問題の設定

統計とは、一定の共通な性質をもつものの集まり——母集団——における変異の有様を記述する数学的事実、すなわち、我々の思惟の対象である母集団の分布・特徴・傾向を数量的に調べることである。

以下、実例に対して統計的処理を行ってみよう。表1は聖徳学園女子短期大学の新設商経学科第1期生(平成4年度入学)132人の身長(単位cm)である。

表1 聖徳学園女子短期大学 商経学科 平成4年度1年 132人の身長

160.0	157.8	160.8	155.0	159.3	163.8	157.0	162.6	157.9	153.2
166.4	167.2	156.8	162.5	165.2	152.9	163.1	157.2	162.5	151.4
164.5	155.0	159.1	156.5	146.6	157.6	153.5	159.1	157.1	162.2
153.7	153.6	154.9	158.3	162.0	156.8	144.5	165.0	152.0	154.4
154.3	153.4	160.0	156.2	162.8	158.2	151.7	160.0	148.7	168.8
157.1	157.7	159.5	162.3	166.0	155.2	162.8	154.3	154.8	155.1
152.6	150.0	157.5	156.2	149.6	165.1	157.8	154.4	158.8	153.4
165.3	165.7	167.0	160.7	154.0	164.0	156.3	158.7	158.8	159.6
150.8	156.2	146.6	159.8	149.2	164.4	156.7	159.4	164.3	156.0
150.2	168.8	155.4	146.0	162.8	154.0	144.2	150.0	168.2	155.2
154.6	166.7	160.5	155.1	160.0	153.6	162.3	162.0	156.6	164.5
160.6	160.5	156.8	156.5	156.4	166.4	157.7	161.6	159.4	160.8
166.3	152.5	158.5	155.8	153.4	159.3	161.5	154.7	159.0	158.0
167.8	151.8								

III. 基本統計量

まず、表1に対する基本統計量を求めるため、大小の順に並び換えると表2のようになる。

表2 並び換え (小→大)

144.2	144.5	146.0	146.6	146.6	148.7	149.2	149.6	150.0	150.0
150.2	150.8	151.4	151.7	151.8	152.0	152.5	152.6	152.9	153.2
153.4	153.4	153.4	153.5	153.6	153.6	153.7	154.0	154.0	154.3
154.3	154.4	154.4	154.6	154.7	154.8	154.9	155.0	155.0	155.1
155.1	155.2	155.2	155.4	155.8	156.0	156.2	156.2	156.2	156.3
156.4	156.5	156.5	156.6	156.7	156.8	156.8	156.8	157.0	157.1
157.1	157.2	157.5	157.6	157.7	157.7	157.8	157.8	157.9	158.0
158.2	158.3	158.5	158.7	158.8	158.8	159.0	159.1	159.1	159.3
159.3	159.4	159.4	159.5	159.6	159.8	160.0	160.0	160.0	160.0
160.5	160.5	160.6	160.7	160.8	160.8	161.5	161.6	162.0	162.0
162.2	162.3	162.3	162.5	162.5	162.6	162.8	162.8	162.8	163.1
163.8	164.0	164.3	164.4	164.5	164.5	165.0	165.1	165.2	165.3
165.7	166.0	166.3	166.4	166.4	166.7	167.0	167.2	167.8	168.2
168.8	168.8								

表2を基に、母集団の特性値である基本統計量を計算すると表3を得る。

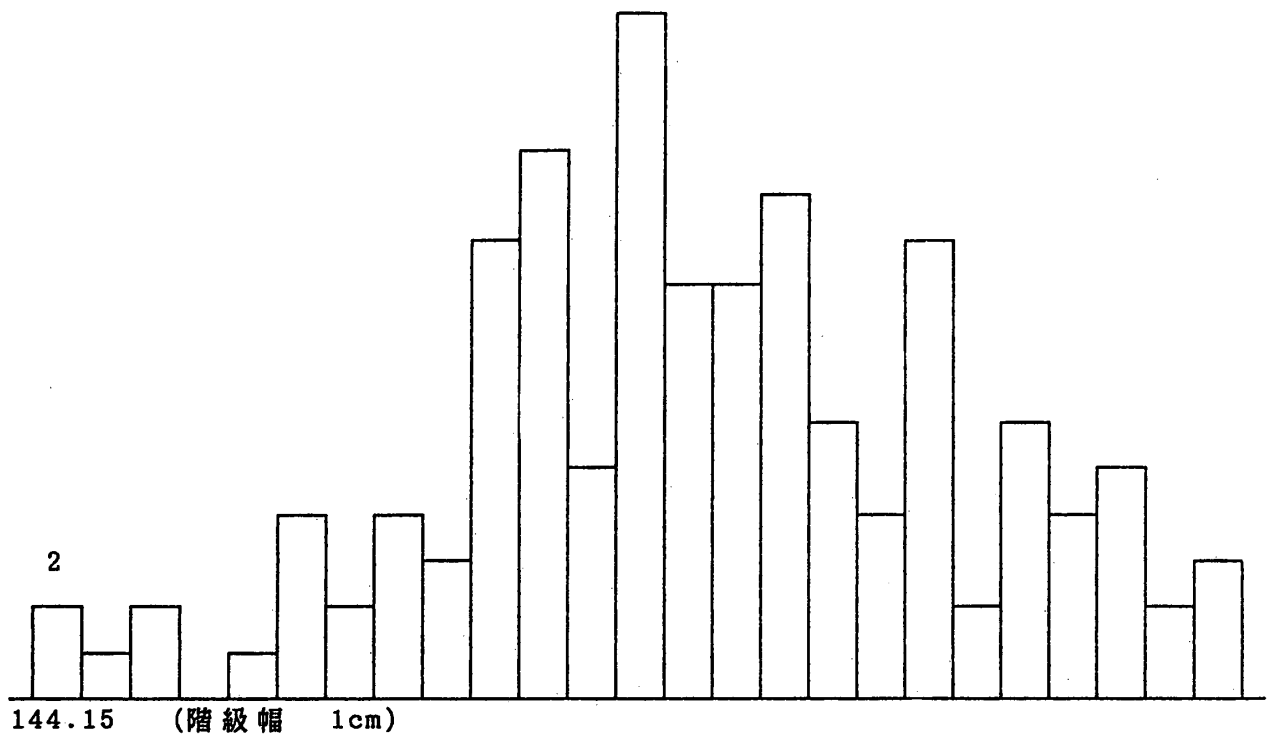
表3 基本統計量

データ数	132
最大値	168.8
最小値	144.2
範囲	24.6
平均値	157.98
分散	27.87
標準偏差	5.28
中位数	157.75
第1四分位数	154.4
第3四分位数	162

IV. 度数分布図

次に分布の特徴を調べるために、最小値が144.2cmであることに留意しながら、1cm及び5cm刻みの度数分布図を描くと図1及び図2のようになる。

(図1)



(図2)

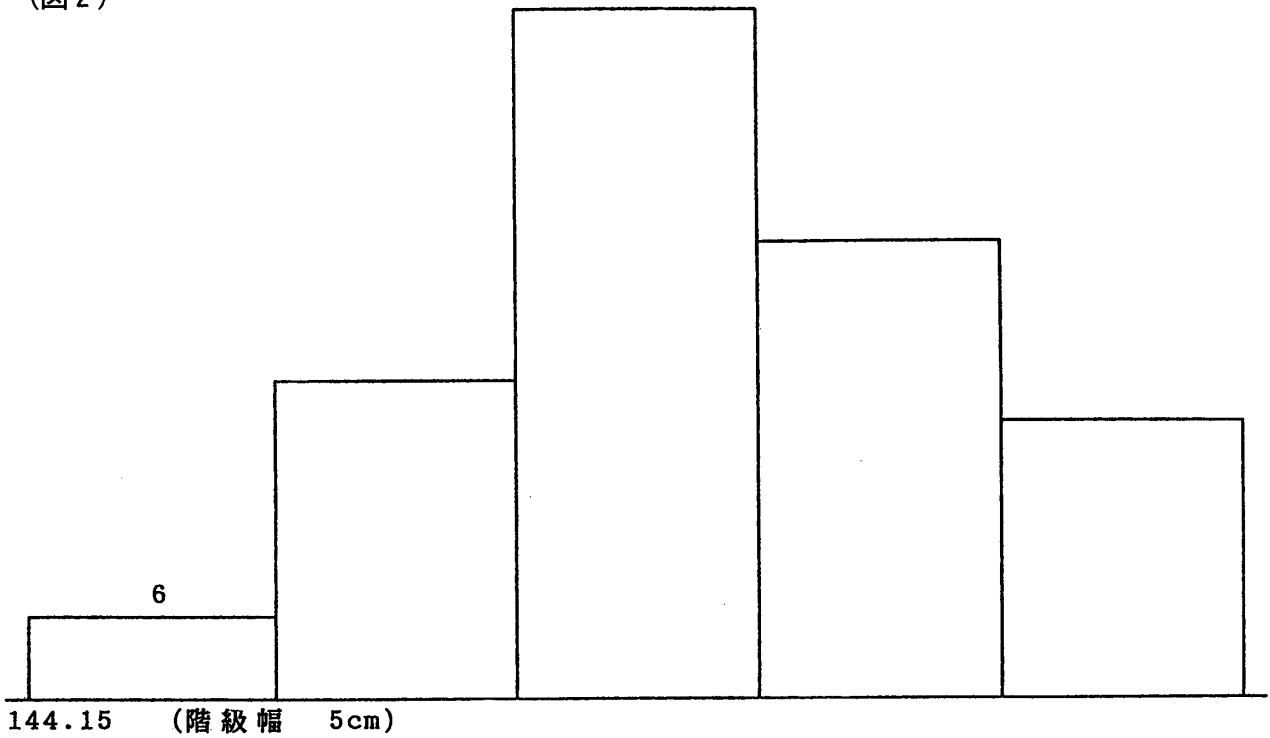
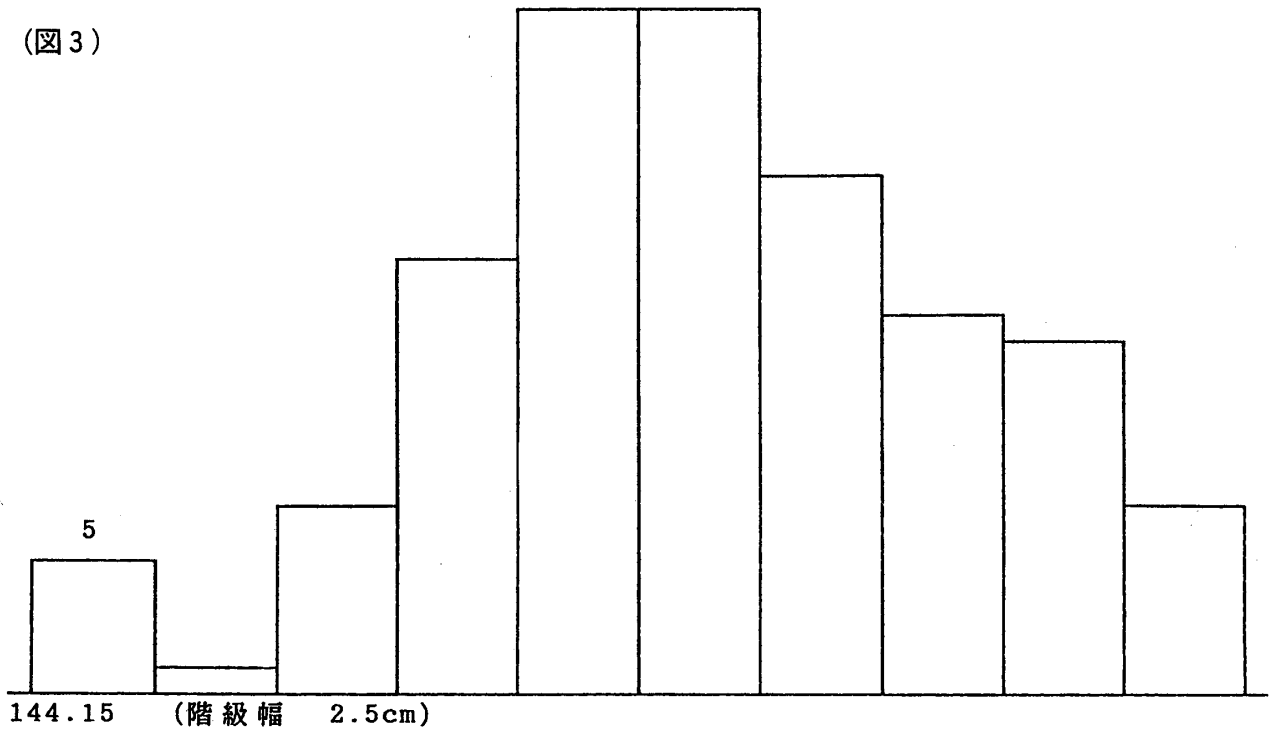


図1及び図2によると、前者は出入りが激しく、また、後者はおおざっぱすぎて、共に全体としての特徴がつかみにくい。そこで、階級幅を決める一つの目安として

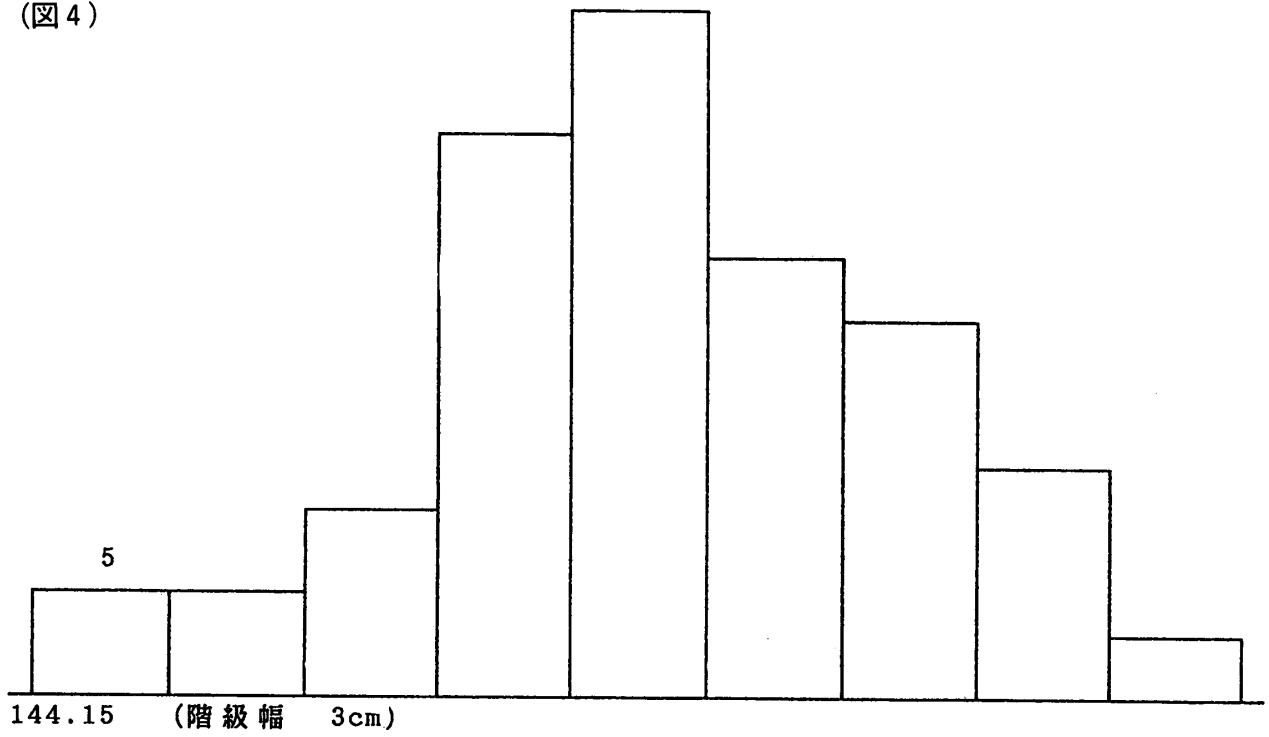
最大値	168.8	最小値	144.2
範囲	24.6	データ数	132

にSturgesの公式を適用し、得られた数値を参考にして、階級幅を例えば簡潔な数値2.5cm、3cm、3.5cm及び4cmにとると、分布図として図3、図4、図5、図6を得る。

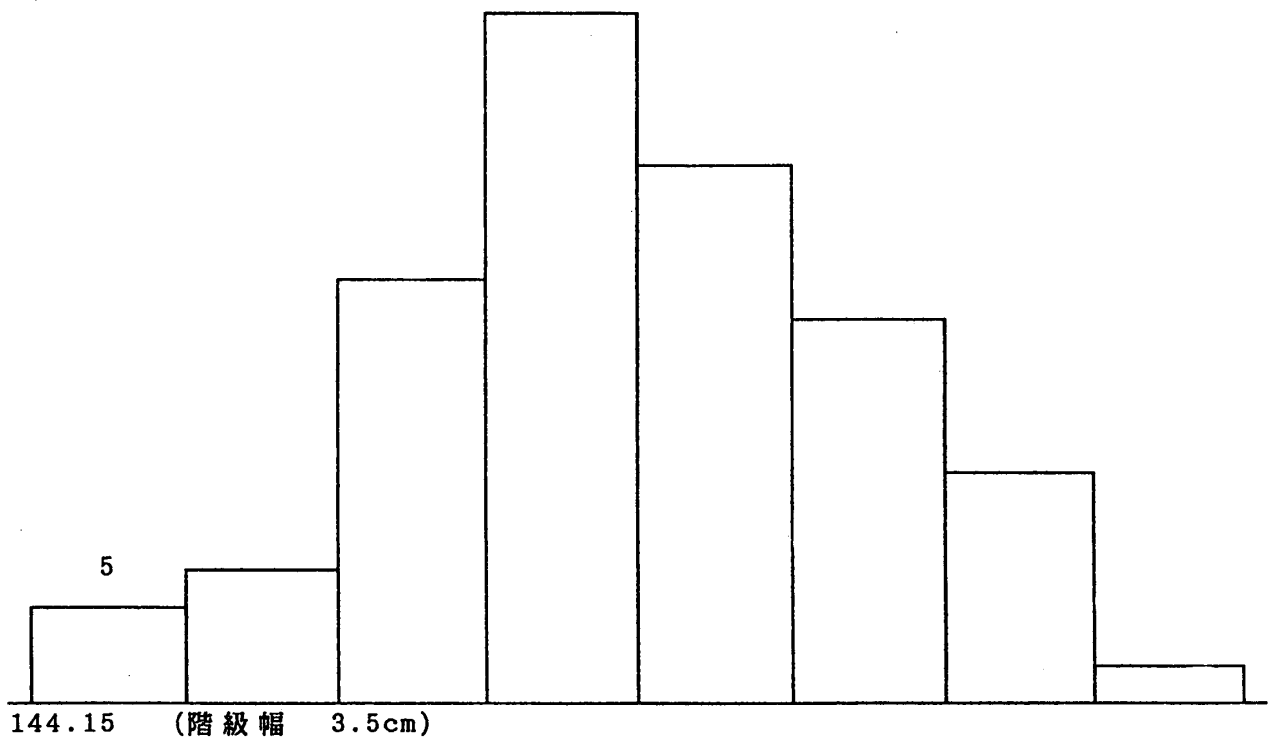
(図3)



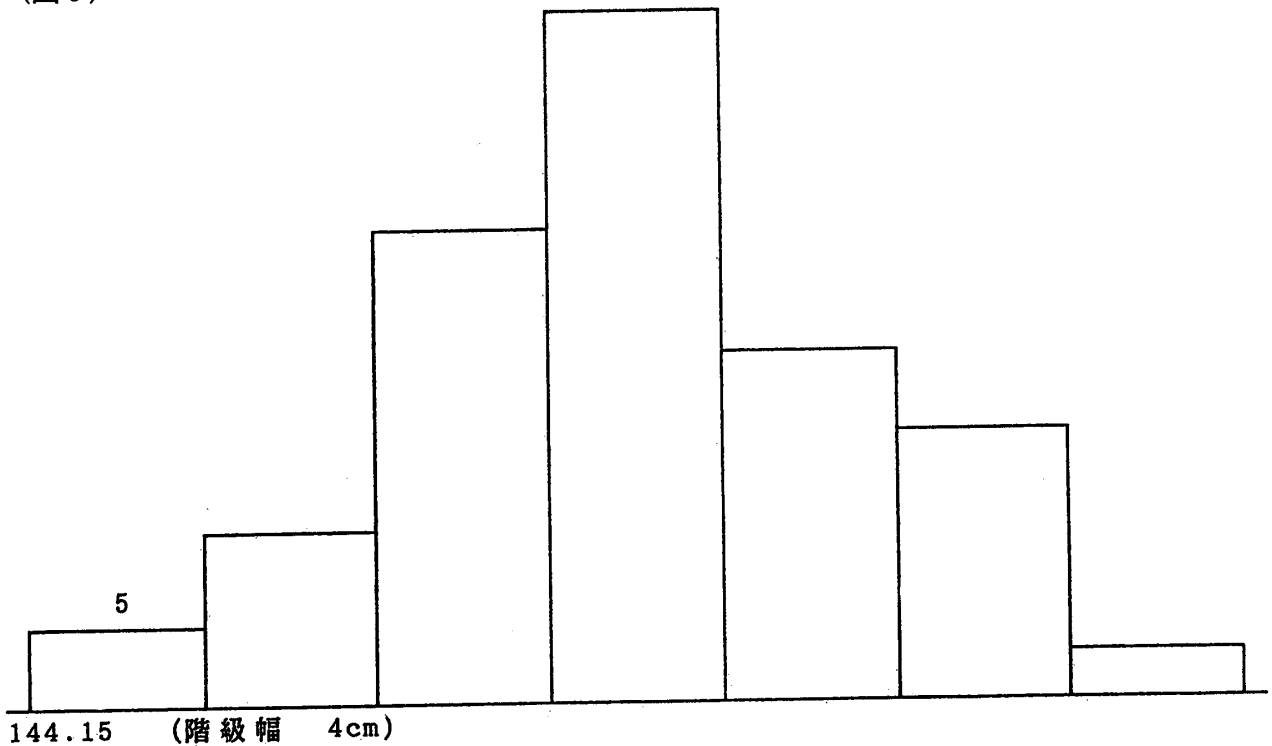
(図4)



(図5)



(図6)



V. 積 率

図3ではやや弱い複峰性が現れる。また、図4、図5では最大度数の階級の両側の階級の度数が大小反対になっており、非対称性の歪みが問題となる。そこで、偏りという考えの最も直観的なPearsonの簡便法： $(M-Mo)/S$ (ここにMは平均値、Moは最頻値、Sは標準偏差)の符号で歪みを判定すると、これは図からも概略分かることであるが、図4と図5では正となり最頻値に対し右に偏り、図6では負となり最頻値に対し左に偏っている。ところが、この方法は母集団表1には適用することができないので、平均値を中心とした偏差三乗平均法： $\alpha_3 = \Sigma (X_i - M)^3 / nS^3$ を用いて歪度を計算すると

表1 では $\alpha_3 = -0.14 < 0$

図4 では $\alpha_3 = -0.15 < 0$

図5 では $\alpha_3 = -0.05 < 0$

図6 では $\alpha_3 = -0.01 < 0$

となり、何れも負の非対称分布となっていることが分かる。

なお、ここで尖度についても触れておくと、平均値を中心とした偏差四乗平均法による尖度測定式：

$\alpha_4 = \Sigma (X_i - M)^4 / nS^4 - 3$ を用いた結果

表1 では $\alpha_4 = -0.21 < 0$

図4 では $\alpha_4 = -0.19 < 0$

図5 では $\alpha_4 = -0.33 < 0$

図6 では $\alpha_4 = -0.28 < 0$

となり、何れも正規分布のよりやや鈍峰であることが分かる。

以上をまとめると、我々は母集団表1の縮図の候補として次の表4を得る。

表4 基本統計量

	(表 1)	(3 cm刻み)	(3.5cm刻み)	(4 cm刻み)
データ数	132	132	132	132
平均値	157.98	158.15	158.02	158.09
分散	27.87	28.93	29.22	28.60
標準偏差	5.28	5.38	5.41	5.35
中位数	157.75	157.97	157.76	157.97
第1四分位数	154.4	154.71	154.49	154.42
第3四分位数	162	162.01	161.83	161.79
歪度	-0.14	-0.15	-0.05	-0.01
尖度	-0.21	-0.19	-0.33	-0.28

VI. 母集団の適切な縮図としての度数分布図

表4より、我々は母集団表1の適切な縮図として、対称性が正規分布のより僅かにくずれた、偏差三乗平均法による負の非対称で、尖りが正規分布のよりやや鈍い図4又は図5を提示することができる。参考までに、図4に対する詳細な表及び図を示しておこう。

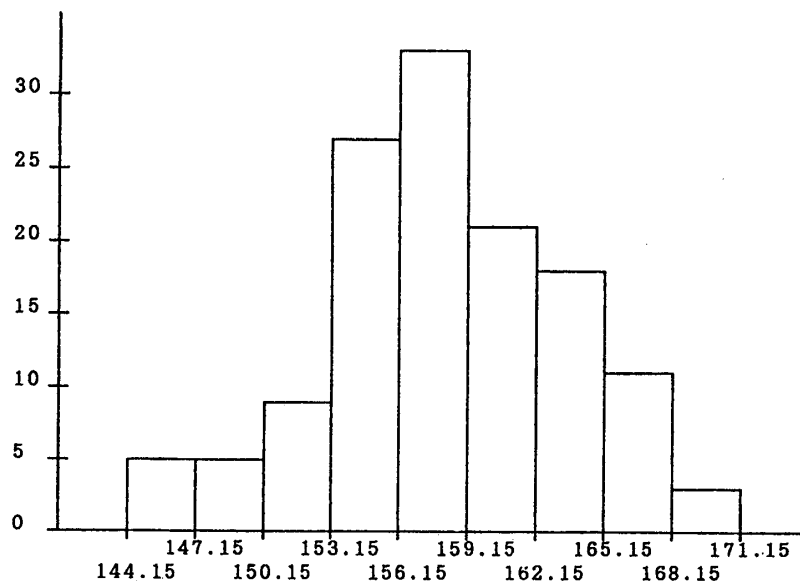
表5 度数分布

第1階級の級下限界	:	144.15				
階級幅	:	3				
階級	階級値	度	数	累積度数		
144.15-147.15	145.65	5	3.8%	5	3.8%	
147.15-150.15	148.65	5	3.8%	10	7.6%	
150.15-153.15	151.65	9	6.8%	19	14.4%	
153.15-156.15	154.65	27	20.5%	46	34.9%	
156.15-159.15	157.65	33	25.0%	79	59.9%	
159.15-162.15	160.65	21	15.9%	100	75.8%	
162.15-165.15	163.65	18	13.6%	118	89.4%	
165.15-168.15	166.65	11	8.3%	129	97.7%	
168.15-171.15	169.65	3	2.3%	132	100.0%	
合計		132	100.0%			

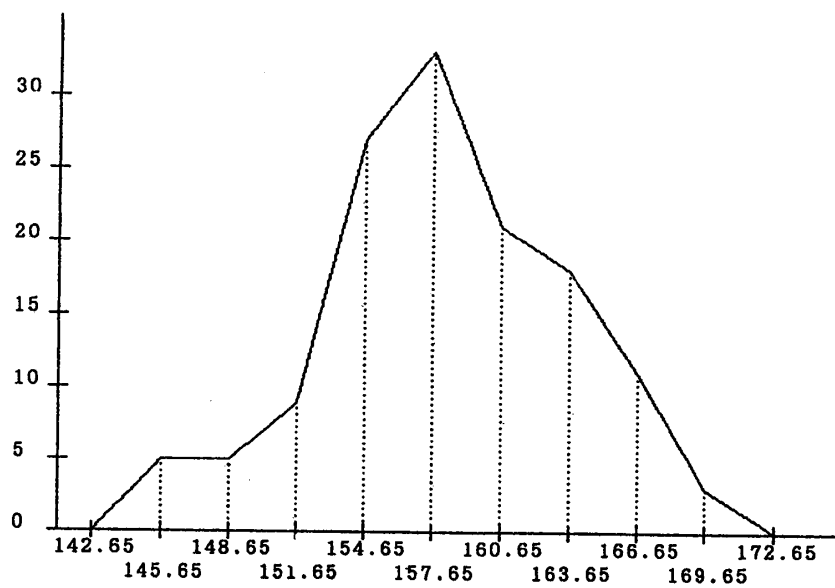
表6 基本統計量

	(表 1)	(3 cm刻み)
データ数	132	132
平均値	157.98	158.15
分散	27.87	28.93
標準偏差	5.28	5.38
中位数	157.75	157.97
第1四分位数	154.4	154.71
第3四分位数	162	162.01
歪度	-0.14	-0.15
尖度	-0.21	-0.19

(図7)



(図8)



Ⅶ. 付 記

我々は最頻値として通常使用される最大度数の階級の階級値を用いた。第2番目に多い度数の階級の側へ、比例或いは反比例方式で修正する方法もあるが、それは意見等の多義性のときに有用であり、いまの場合は全体像を考察の対象にしているので、通常の方法によった訳である。なお、第1階級の級下限界を、最小値144.2cmを考慮した可能な範囲で変化させても、略同じであることを付け加えておく。

Ⅷ. あ と が き

以上、我々はデータの縮約としての度数分布図の階級幅を、Sturgesの公式による数値だけを目安にして求めたのではなく、同時に歪度及び尖度をも考慮しながら決定した。それは、安易に階級幅を決めて度数分布図を描くよりも、母集団の特性が少しでも多く反映されたような度数分布図の作成を願ったが故である。そして、その一助として積率の活用を試みた次第である。

最後に、新設商経学科の限りない前進を祈念しながら筆を擱く。

——商経学科 情報処理研究室にて——

参 考 文 献

- (1) H. A. Sturges : The choice of a class interval ; J. Amer. Statist. Assoc. 21 (1926)
- (2) 小暮厚之 : ヒストグラムのための最適な級区間 MISE基準 ; 日本数学会 数学第41巻第3号 (1989)
- (3) H. Cramer : Mathematical Method of Statistics ; Princeton Univ. Press (1946)
- (4) A. M. Mood・F. A. Graybill・D. C. Boes : Introduction to the Theory of Statistics ; McGraw-Hill (Kogakusha 1974版)
- (5) 岸根卓郎 : 理論応用 統計学 ; 養賢堂 (1981版)
- (6) 脇本和昌・垂水共之・田中豊 : パソコン統計ハンドブック I 基礎統計編 ; 共立出版 (1989版)