⑤ 小学校の算術能力調査に於けるサムプリング計画
（横浜市に於ける）

林　知　己　夫

此は国立教育研究所所員久保舜一氏の協力の下に行つたものである。

久保氏は現在小学校生徒の算術の能力がどの程度であるか又昔の生徒の能力とくらべてどうであろうかと言ふことの調査を企画された。このために一まづ小学校6年生を対象とし，昭和の初めに行はれたテスト問題（当時教科書を中心としたもの）を現在の6年生に行つて比較してみることにした。　なお学力規準がことなるため、高学年用の他，中学年用（3、4年生用）のものもあわせ行つてみることにした。問題の用語なども現在に比べて甚だ難しいので，これも現在知られてゐる用語，かなづかひに訂正する様に注意が拂はれてゐる。

調査対象としては種々の便宜から横浜市にある全小学校の6年生が選ばれた。如上の目的を達するためには勿論すべての6年生を調査する必要はない。　サムプリングによつて何名かの6年生を調査すればよいのである。以下にこのサムプリング計画を具体的にのべてみようと思う。

§ 1. 調査対象と母集団

調査対象は横浜市にある全小学校に在籍する6年生すべてである。

各生徒の標識は，課すべき算術の問題の正答数である。即ち算術の能力点数であると考える。

今各生徒に等しい抽出確率をあたえて母集団を構成するとしよう。

我々の目標は標識の母集団平均を推定することになるのである。

小学校は86校あり，全市に散在してゐるのであるから直接に小学生を等しい確率で抽出することは困難であり、且又その場合各小学校を煩はせることになるので手間や費用がかさんでくることにもなるのである。

従つて抽出は一まづ小学校をいくつか抽出し，しかる後抽出された小学校から学生を抽出すると言ふまたぬき抽出法を採用することにした。

またぬき法を採用する以上精度が落ちるので深酷に層別をほどこし学校間の外分散を小にする様に心かけねばならない。

§ 2. 学校の層別

第一次抽出單位たる学校を層別する時に、第一次抽出單位をどの位抽出するかの見当をつけておき、層の数の目安をきめておかねばならない。　種々の事情 ―― 費用、調査員の事情、時日等 ―― を考慮し大体10位の見当をつけておくことにする。

学校の層別は何を目印に行えばよいであろうか。

横浜市に於ては幸なことがある。昭和24年に久保氏の行われた算術、国語、知能テストに関する調査がある。

此は勿論すべての小学校について行はれたものでなく、サムプル校について（ランダムサムプリング；昭和24年度、横浜市教育調査報告横浜市小学校校長会発行参照）行われたものである。

此の結果を分析して今の場合の層別に利用してみることにする。

横浜全市の学校については所在地区、二部教授の状況、全生徒数，6年の　、一学級当り平均生徒数，校格が判明してゐる。

これらの資料を組合せて層別の規準とすればよいのである。

此の資料は第一表の通りである。

さて，このためには前の調査でサムプルされた学校について，それらの各種の資料と各学校の算術の平均点数との関係をしらべておく必要がある。

今度行うテストはその性質から言って算術の成績と関係が深いと考えられるので算術の成績を分析して層別に用うるのが有利であろうと考えられる。

私は重相関関係の方法を援用してゆこうと思う。

各学校の特性を示す各項目のカテゴリーは大略A，B，C，D，E，にわかれてある。　各学校について算術のテスト（6年生のみを対象とする）成績はテストによつて求められてある。

こうゆうわけであるから点数と各項目のカテゴリーとの関係をしらべることが出來る。

このためにまず各項目のカテゴリーをそのカテゴリーの属するものの平均点数を以て数量化し各項目、点数間の相関係数を求めてみる必要がある。求めてみると次の様になる。

	点	x_1 地区	x_2 二部教授	x_3 校格	x_4 一学級当り生徒数	x_5 全生徒数
点	1	0.795	0.780	0.703	0.594	0.666
x_1		1	0.546	0.620	0.635	0.847
x_2			1	0.832	0.700	0.561
x_3				1	0.788	0.661
x_4					1	0.773
x_5						1

此を用いて各学校の点数 $\tilde{y}$ の推定をつくつてみると

$$\tilde{y} = 0.831x_1 + 0.764x_2 - 0.020x_3 - 0.264x_4 - 0.146x_5 - 7.279$$

をうる。

なお，得点と他項目との重相関係数は 0.906 となり，項目よりする得点との推定は極めて精度のよいものとなる。

此によつて推定した各学校の成績は第一表に示すとおりである。勿論サンプルの学校数はさう大でなかつたので信頼度はさう高くはないと考えられるが，重大な目安となるであらう。

此の推定された成績からみて各第一次抽出単位たる学校間の外分散が小となる様に層別をほどこせばよいのである。

このために各点数の分布をとつてみた。第二[illegible]通りである。

学校のサムプリングは各層から一つの学校をその大さに比例した確率を以て抽出する方法をとるのであるから，抽出すべき学校数をまづさだめ、次に層にぞくする生徒数をにらみ合せ try and error の様な方法で学校数（層の数）及びサムプルすべき6年生の数をきめてゆくのである。

此等の思考経過をたどり調査実施上の制限をも考へ合せ抽出学校数は10校（層の数10）と定めた。

さうして属する生徒数の大さを考へに入れて層をつくつたのであるが此の層は第二表に示した区切りを以て示される。　次に各層に属する学校の外分散を求めてみた。　此は第三表に示す通りである。

なお，此の表で $\overline{X}_i$ は，i 層の平均，σ_{bi}^2 は i 層の分散であり

$$\overline{X}_i = \sum_{j=1}^{M_i} \overline{X}_{ij}\, p_{ij}$$

$$\sigma_{bi}^2 = \sum_{j=1}^{M_i} (\overline{X}_{ij} - \overline{X}_i)^2 p_{ij}$$

$\overline{X}_{ij}$ は第 i 層に属する j 学校の平均点数

p_{ij} は第 i 層に属する j 学校の生徒数の大さの i 層全体の大さに対する比

M_i は第 i 層に属する学校数

$p_i = \dfrac{N_i}{\Sigma N_i}$ ， N_i は第 i 層に属する6年生の生徒数である。但し，この場合 $N_i = k N_i'$

N_i' は第 i 層に属する学校の全生徒数，k は常数と考えてある。（此は無理のない仮定であろう、§5．参考資料参照）

以上の様にして層別を終り精度の計算に移ろう。

§3．精度の計算とサムプルの大さ

前にも述べた様に我々は各生徒のテスト成績点数の母集団平均を求めたいのであつた。

今各層から一つの学校をその生徒数に応じた確率で抽出、次に各学校から所定の生徒数を等しい確率で抽出、母集団平均の推定をつくるものとしよう。

こうしてつくつた推定平均を $\overline{x}$ とする。

$$\overline{x} = \sum_{i=1}^{10} \overline{x}_i p_i$$

$$\overline{x}_i = \frac{\sum_{j=1}^{n_i} x_{ij}}{n_i}$$

但し n_i は i 層に配当せられたサムプル数　x_{ij} は，抽出された学校から抽出された j なる生徒の示す点数

勿論，$E(\overline{x}) = \overline{X}$（母集団平均）であり $\overline{x}$ は $\overline{X}$ の偏りのない推定量となつてある。

さて，$n_i = n p_i$（比例割当法），n はサンプル の大さと云う様な方法をとるものとする。

さて，$\bar{x}$の分散 $\sigma_{\bar{x}}^2$ を求めてみると

$$\sigma_{\bar{x}}^2 = \sum_{i=1}^{10} p_i^2 \sum_{j=1}^{M_i} \frac{N_{ij}-n_i}{N_{ij}-1} \frac{\sigma_{ij}^2}{n_i} p_{ij} + \sum_{i=1}^{10} \sigma_{b_i}^2 p_i^2$$

但し　N_{ij} は，第i層第j学校の大きさ

σ_{ij}^2 は，第i層第j学校の生徒の間の分散

$\sigma_{b_i}^2$ は，第i層の外分散

となる。

今　$\sigma_{ij}^2 \fallingdotseq \sigma_i^2$

$$\frac{N_{ij}-n_i}{N_{ij}-1} \fallingdotseq \left(1-\frac{n_i}{N_{ij}}\right) = a$$

と考えて $\sigma_{\bar{x}}^2$ の評価を行えば

$$\sigma_{\bar{x}}^2 = a \sum_{i=1}^{10} p_i^2 \frac{\sigma_i^2}{n_i} + \sum_{i=1}^{10} \sigma_{b_i}^2 p_i^2$$

$$= \frac{a}{n} \sum_{i=1}^{10} p_i \sigma_i^2 + \sum_{i=1}^{10} \sigma_{b_i}^2 p_i^2$$

さらに又　$\sigma_i^2 \fallingdotseq \sigma^2$ と考えると

$$\sigma_{\bar{x}}^2 \fallingdotseq \frac{\sigma^2}{n} a + \sum_{i=1}^{10} \sigma_{b_i}^2 p_i^2$$

となる。　$\bar{x}$ の変異係数を求めてみると

$$\frac{\sigma_{\bar{x}}^2}{\bar{X}^2} = \frac{\sigma^2}{\bar{X}^2}\frac{a}{n} + \frac{1}{\bar{X}^2}\sum_{i=1}^{10} \sigma_{b_i}^2 p_i^2$$

ここで実際の数値を入れてみる

$$\frac{1}{X^2}\sum_{i=1}^{10}\sigma_{bi}^2 p_i^2 = \frac{1}{(43.6)^2}\sum_{i=1}^{10}\sigma_{bi}^2 p_i^2 = 0.0000595$$

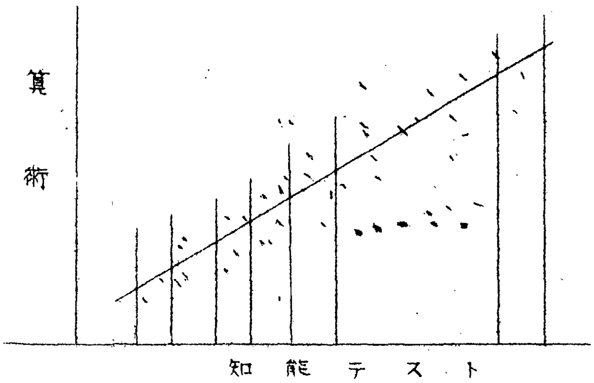
となる。 重相関係数は 0.9 であるが，散ばりの見当をつけ外分散を大き目に見積り上の値の２倍をとっておけば，まづ安全な推定となるであろう。（「日本人の読み書き能力」「サムプリング調査は如何にすべきか」東大協同組合出版部の二書の中，読み書き能力調査に於ける‘層別の効果’の項参照）

さうすると $$\frac{1}{X^2}\sum_{i=1}^{10}\sigma_{bi}^2 p_i^2 \fallingdotseq 0.000319$$

となる。

さて次に第一項の見積りになる。

簡単のため，学校内の生徒間の分散はみな等しく σ^2 であるとした。我々は抽出された学校から生徒を抽出するのに次の様な工夫をすることにする。 即ち，各生徒の知能テストと算術能力との相関関係のあることを利用し，知能テストの点数で生徒を層別し，n' 人の生徒を抽出することを考へるのである。



n'人サンプルを抽出することは各生徒を n' の層に分けることと考へる。

さうして各層から一つのサムプルを抽出することと考へる。

今，大局的に計算するため，知能テスト一定とするとき算術成績の分散は一定であるとすると~~すると~~，知能テスト一定のときの算術の分散は

$$\sigma^2 \ (1-\rho^2)$$

となる。

ここに ρ は算術と知能テストの間の相関係数であり，横浜市の６年生の場合 0.70 と推定されてゐる。（前稿の報告参照）

又，各層の中の知能テストは略同一の点をもつものと仮定しよう。

n' が相当大のとき此の仮定はさう無理なものでなく，第一次近似としては認められよう。

こう考えてくると n' のサムプルを抽出してつくった平均の分散は

$$\frac{\sigma^2(1-\rho^2)}{n}$$ となるのである。

以上の様な考え方に従い，知能テストによつて層別しサムプルを抽出するならば（実際的には知能テストの高いものから低いものへとならべ，等間隔抽出法によつてサムプルを抽出すれば所期の目的は達せられると考えてよいであろう。）

第一項に相当するものの評価は

$$a\cdot\sum_{i=1}^{10} p_i \frac{\sigma_i^2}{n_i} \sim a\sum_{i=1}^{10} p_i \frac{\sigma_i^2(1-\rho_i^2)}{n_i} \doteqdot a \ \frac{\sigma^2(1-\rho^2)}{n}$$

となる。

$$a\cdot \frac{\sigma^2}{\overline{X}^2} \ \frac{(1-\rho^2)}{n}$$ の数値を求めると

$$\sigma^2 = 529.00$$
$$\overline{X}^2 = 2141.838$$
$$\rho^2 = 0.4914$$

であるから

$$\frac{a}{n}\left\{\frac{529.\times 0.5086}{2141.838}\right\}$$ となる。

$n = 800,\ 700,\ 600$ とれば a は平均的に言って（各学校の平均的にみた６年生総数と各学校に対する平均サムプル数との比であると見做す！）

$$a \fallingdotseq 0.39,\ 0.46,\ 0.54$$

となるから

$$\frac{a}{n}\ \frac{\sigma^2(1-\rho^2)}{\overline{X}^2} \quad \text{は夫々}$$

$$0.000061,\quad 0.000083,\quad 0.000113$$

となる。 但し，６年生の総数は全生徒数の $\frac{1}{7.6}$ と考えてある。
種々の分析，学校の負担を考えて $n = 700$ と定めた。

こうすると，結局のところ，抽出学校数１０，サムプル数７００と言うことになる。 このサムプルを各層に比例割当てした数は第三表にある。

このときの推定平均の精度は

$$\frac{1}{\overline{X}^2}\sigma_{\bar{x}}^2 = 0.000229 \qquad \text{となる。}$$

$$\frac{\sigma_{\bar{x}}}{\overline{X}} = 0.015 \qquad \text{となる。}$$

つまり，信頼度９５％とするとき，信頼巾の相対精度は約３％と言うことになり，相当満足すべきサムプリングとなるであろう。

§４. サンプルの抽出

各層から夫々生徒数の大きさに比例した確率を以て学校を抽出した。勿論，サムプル割当ての場合と同様６年の生徒数は全生徒に比例するものとしてある。 抽出校は第三表にある。 此の学校から割当てられたサンプルを抽出するのであるが抽出巾は略々第三表の様になる。
参考のため各層別に平均的な抽出巾を出しておいた。

第　一　表

学校名	地區	二部教	全生徒数	一学級生徒数	校格	推定による数学点	Size	学校名	地區	二部教授	全生徒数	一学級生徒数	校格	推定による数学点数	Size
1	E	D	C	B	D	40	1274	26	E	D	D	A	B	40	615
2	E	C	D	A	B	51	658	27	E	C	A	A	C	50	2687
3	D	D	D	A	D	36	1454	28	D	D	D	B	B	37	622
4	D	D	C	A	B	35	1099	29	D	C	C	A	C	46	1177
5	E	C	D	A	C	51	880	30	D	C	D	B	B	48	603
6	E	C	C	A	D	50	1335	31	E	C	C	A	B	50	1264
7	E	C	D	A	D	51	781	32	C	D	C	A	B	34	1184
8	D	C	A	A	B	46	1896	33	C	D	D	A	B	35	608
9	B	D	A	A	C	29	1540	34	B	C	A	A	C	40	2403
10	D	C	A	A	D	46	1806	35	A	D	A	A	A	25	664
11	B	D	C	A	B	30	1285	36	A	C	C	A	B	36	1326
12	C	C	A	A	B	45	1608	37	A	D	D	A	B	26	706
13	E	C	A	A	D	50	1753	38	A	D	B	B	A	28	221
14	E	C	A	A	C	50	1737	39	A	D	B	A	B	27	313
15	D	C	A	A	D	46	2409	40	A	C	D	A	B	37	692
16	E	C	A	A	D	50	2692	41	A	D	D	A	B	26	713
17	D	C	A	A	B	46	1586	42	B	C	A	A	B	41	1741
18	C	D	D	A	B	35	581	43	D	C	A	A	C	46	1748
19	C	D	B	B	B	37	383	44	B	C	D	B	C	42	808
20	A	D	B	B	A	28	386	45	B	B	A	A	C	49	2711
21	D	B	C	A	D	55	1428	46	B	C	D	A	C	41	880
22	D	C	A	A	D	46	2062	47	B	C	A	A	B	41	1739
23	E	C	A	A	B	50	2135	48	B	B	A	A	B	49	2416
24	E	C	A	A	C	50	2030	49	C	C	D	A	E	46	671
25	E	C	A	A	C	50	2043	50	E	B	A	A	D	59	2079

39・40頁間折込

第一表、つゞき

学校名	地區	二部教授	全生徒数	一学級生徒数	校格	推定による数学点数	Size
51	B	C	C	A	C	41	1090
52	C	B	C	A	B	54	1386
53	E	C	D	A	B	51	804
54	B	C	D	A	C	41	573
55	E	D	A	A	C	39	1544
56	E	C	D	A	B	51	670
57	E	A	A	A	B	46	1674
58	B	B	A	A	C	49	2435
59	D	C	A	A	B	46	2181
60	D	C	A	A	A	46	1257
61	A	D	B	A	B	27	321
62	C	C	D	B	B	47	741
63	A	B	A	A	B	45	2151
64	D	C	D	B	B1	48	825
65	A	D	D	C	B	30	660
66	A	D	D	A	B	26	646
67	A	D	D	A	A	26	896
68	D	D	D	A	B	36	686
69	A	D	B	B	A	28	399
70	A	C	B	B	A	39	309
71	C	D	B	A	B	36	312
72	A	C	D	B	A	38	563
73	A	D	B	B	B	28	374
74	A	C	D	A	A	37	645
75	A	C	B	A	A	38	320
76	E	C	D	A	B	51	621
77	C	D	B	A	B	36	303
78	D	C	A	A	C	46	1715
79	D	D	C	A	C	35	1369
80	B	A	D	A	B	38	797
81	C	C	C	A	B	45	1340
82	C	D	C	A	B	34	1204
83	A	C	C	A	A	37	1019
84	C	C	D	A	B	46	927
85	A	D	B	C	A	31	127
86	A	D	B	C	B	31	173

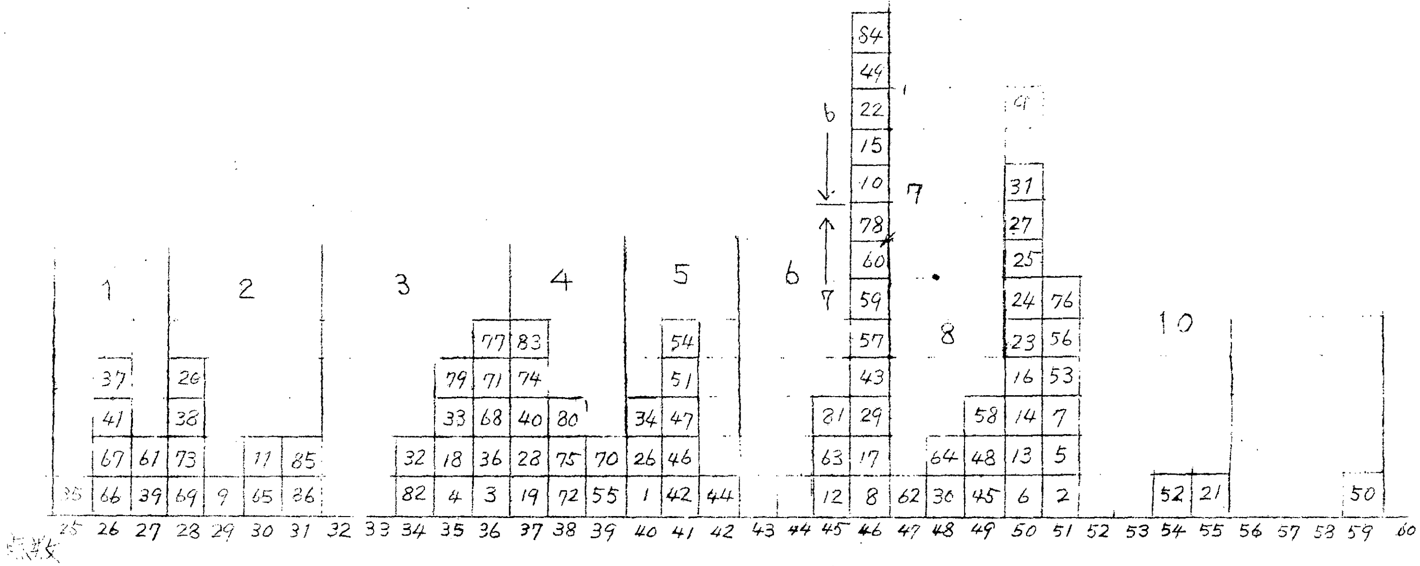
生徒の抽出法は，前述の様に知能テスト点の順にならべ，ランダムスタートの等間隔抽出法によって（或は此を少しく変更して行う。抽出巾が整数でないから）サムプルを抽出するのである。

第二表

六年算数　推定点数

distribution の数字は学校の番号　：（1～86）

区间と数字は層を表す



第　三　表

$$= \frac{\dfrac{\text{全生徒数}}{7.6}}{\text{Sample数}}$$

層番号	$\overline{X}_i$	σ_{bi}^2	p_i	学校名	$p_i^2\sigma_{bi}^2$	Sample数	6年平均生徒数	平均抽出比	学校名	全生徒数	抽出巾
1	26.0	0.30	0.043	7	0.0006	30	80	2.7	川上	713	3.1
2	29.2	0.85	0.053	9	0.0024	37	76	2.1	平沼	1285	4.6
3	35.2	0.61	0.102	11	0.0064	71	121	1.7	中和田	1326	2.5
4	37.8	0.71	0.070	10	0.0034	49	91	1.9	田奈	563	1.5
5	40.7	0.36	0.112	9	0.0045	78	163	2.1	岸谷	1090	1.8
6	45.6	0.24	0.130	8	0.0040	91	212	2.3	日枝	2409	3.5
7	46.0	0.	0.140	8	0	98	229	2.3	白幡	2181	2.9
8	48.7	0.36	0.098	6	0.0035	69	213	3.1	下野谷	2711	5.2
9	50.0	0.	0.158	9	0	111	229	2.1	間門	1335	1.6
10	53.8	9.98	0.094	9	0.0880	66	136	2.1	大島	880	1.8
全	43.6	/	1.000	86	0.1128	700	/	/	/	/	/

§5. 参　考

前に6年生と全生徒数の関係を比例的にみたが，参考のため両者の相関関係をとつてみると次の様になる。

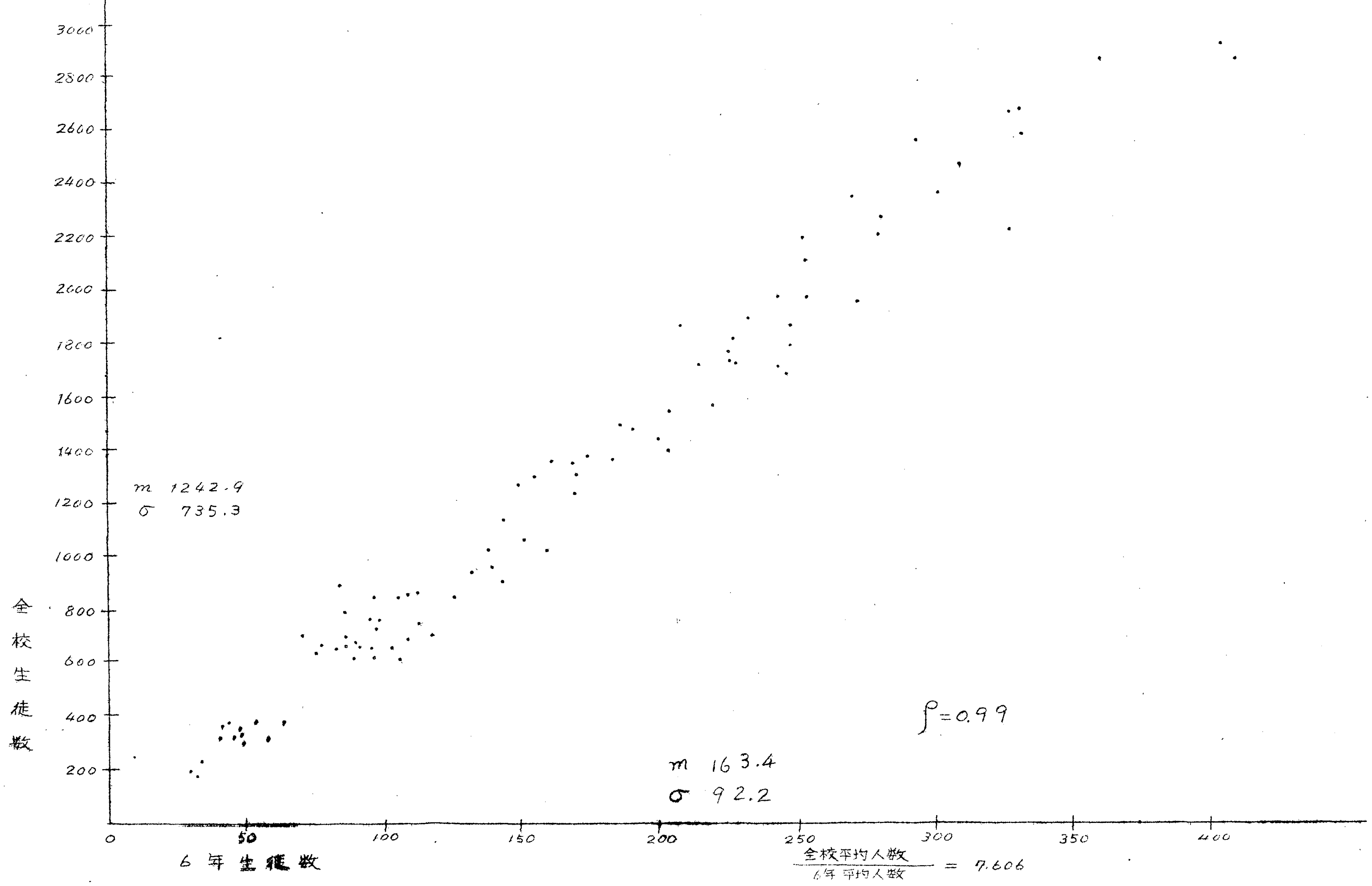
この相関係数は 0.99 であり

$$\frac{6年の生徒数}{全生徒数} = \frac{1}{7.6}$$ となる．

したがつて，前述の議論はサムプリング企画の立場では大局的に妥当なものと言へよう。

第四表

６年生と全校の生徒数相関（86校について）



（43頁ノ次ニ挿入）