㉘ リテラシイ調査にあらはれた分布の型など

［平均と標準偏差との関係］

林　　　知己夫
丸　山　　文　行
石　田　　正　次
西　平　　重　喜

全国から層化任意標本副次抽出法に依つて２１００８の標本を２７０個所の地点（市区町村）へ割当て集合テストを行つた結果１６８１４の被調査者を得た。（此等のデザインの結果の分析は別に発表をする）

各人の読み書き能力をテストに依つて調べ１００点満点で表す事にし，こゝではその分布の型のみを問題にしてみよう。

総得点の分布構造は多くのものについて亅字型分布をなすことが知られた。（高得点に最も人数多く，又低得点に於て之稍人数が多い。）［図Ｃ参照］　此は読み書き能力を調査する為に作られた問題の性質に依るものである。更に易しい問題，難しい問題を出す時，おそらくはこの分布の型は、正規分布に近い密度函数の和（出来る群、と全く出来ない群）として表はせるのではないかと思はれる。

しかし，その様に更に易しい又，更に難しい問題を出すことは，「読み書き能力調査」の主旨にそはない。

この様な問題に対して被調査群が亅字型分布をなすのは却つて面白

い事であらう。

さて，この様にして得られた分布の平均と標準偏差の間に一定の興味深い問題（抛物線的な関係）が見出された。これは次の表に示されてゐる通りである。

◎ 表「各調査地点別平均と標準偏差の表」

この様な平均と標準偏差との関係は，分布の型から由來するものではなからうか。

これを考へるに当つて分布の型が $y = ax^m + b$ に依つて示されるものとしてみよう。

$a,\ m,\ b$ は常数

x は得点

$$\int_0^{100} y\,dx = 1 \quad \text{即ち} \quad \frac{a}{m+1}\cdot 100^{m+1} + 100b = 1$$

の関係は当然満たされねばならない。

この分布函数の平均 S, Variance σ^2 を計算してみると

$$S = \frac{(100)^{m+2}}{m+2}\cdot a + \frac{100^2}{2}\cdot b - x_0^2\,\frac{m}{2(m+2)}\,b.$$

但し $\quad x_0^m = -\dfrac{b}{a}$

$$a.\ \frac{100^{m+1}}{m+1} + b\cdot 100 - \frac{m}{m+1}\,bx_0 = 1$$ が

満足せられてゐるものとする。

$$\sigma^2 = \frac{(100)^{m+3}}{m+3}\cdot a + \frac{100^3}{3}\cdot b - \frac{m x_0^3}{3(m+3)}\cdot b - S^2$$

とする。

今，もし

$$100\,\frac{m+1}{m+2} \geqq S \geqq 50\,\frac{m+1}{m+2}$$

であるならば S と σ^2 との間に次の様な簡単な関係が満されることが容易に證明せられる。

$$\sigma^2 = (S-50)\cdot 100\cdot \frac{4(m+2)}{3(m+3)} - S^2 + \frac{100^2}{3}$$

なほ　$m=1$　の時は

$$50\,\frac{2}{3} > S,\ \text{又は}\ S > 100\,\frac{2}{3}$$

にある S に関しては

$$S = \frac{1}{\sqrt{2}}(100-m)$$ なる直線となる

この様にして

S と σ との関係を夫々の m の価に関して描いてみると次の様になった。

◎　表挿入

この結果と全国の各 test spot の平均と標準偏差との関係を示すグラフとを重ねて見る時，理論式が極めて良く現象を記述して居る事が判る。

しかし実際の分布の型が $y = ax^m + b$ に依うて表現せられて居るであらうか。（若し表現されて居らねば上述の結論は意味が無い）

この答は肯定的であった。

一例を全国の得点分布について示してみよう。全国の平均と Variance とから a, m, b の常数を計算してみた処

全国の分布は

$$y = 3.402 \times 10^{-15} x^{6.607} + 0.2679 \times 10^{-2}$$

に依つて表現せられる。

此の理論式と実際の分布とを比較してみると次の如き極めてよい一致を得た。

◎ 表C図

点数	調査結果(%)	理論式(%)
0 — 14	4・1	4・3
15 — 19	1・3	1・5
20 — 24	1・3	1・5
25 — 29	1・5	1・5
30 — 34	1・7	1・5
35 — 39	1・6	1・6
40 — 44	1・9	1・7
45 — 49	2・0	1・9
50 — 54	2・6	2・3
55 — 59	3・3	3・0
60 — 64	4・1	4・1
65 — 69	5・7	5・9
70 — 74	8・1	8・6
75 — 79	11・7	12・5
80 — 84	18・0	18・2
85 — 90	31・1	29・9
計	100・0 %	100・0 %

試みに χ^2 検定を行つてみる。　$\chi^2 = 21.76$

D.F. 15.　$P_r\{\chi^2 > 21.76\} \fallingdotseq 0.15$

Sample 数は 66820 である。この程度の大 Sample においてのこの値は，極めて良い Fit を示してゐると言へる。

この様にして全国の各 test Spot の分布構造は $y = ax^m + b$ なる分布曲線群に依つて一応表現せられ —— 此はニつの Parameter（平均と Variance）に依つて分布の型が完全に決ると云ふ極めて重要な結果である。（結論に於て多くの場合このニつを計算して論をすすめた）——— る事が了解せられた。

従つて平均と標準偏差の算出が大いに意味を持つことになつた。なほ各種の分類法に依る平均と標準偏差の関係は［圖 D］に示した通り一定の興味ある抛物型の関係を持つ様に見られるが分布が T，L 字型でないものを含んで居るので上述の様に一元的に扱ふわけにはゆかない。しかし，このグラフの示す一定関係は分布構造に関し多く示唆する処があると思ふ。

◎　図 D.　各 Break Down に応じた平均と標準偏差の関係図表

［註］　J 字型（又は T）分布をなすものに対しては，従来

$$\left(\log \frac{1+x}{1-x}\right)^{\frac{2m+1}{2n+1}}$$

の如き（もとの標識一応 -1 と 1 との間に変換した後）変換を行い正規分布となしうるか否かをみてゆく方法がとられてゐた。（特に $m = n = 0$ として）そして正規とした上で平均とか variance とかの検定がなされてゐたが此については疑問がある。変換された標識の上での平均とか variance とかは，変換されぬ標識（此には内容的に意味がある）の上で意味のつけ難い場合も多いことと思はれる。したがつて J 字型の曲線群の型を特性化して検定論をつくりなほす必要もあらうかと思はれる。

［附　録］　Sampling により得た Sample 平均と、Sample Variance との関係以上とは全く別の立場のものである次の様なことが考へられる。

ある母集団から Sampling を行い n 個の Sample を抽出これより平均と variance とを計算する。

この時、平均が大き目に又少な目に出てきたとき variance は大きく出るだらうか、少く出るであらうか、これをみるために次の様な指標を考へる。

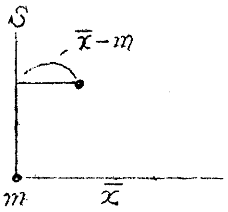
$$(\bar{x}-m)^2 = l^2$$

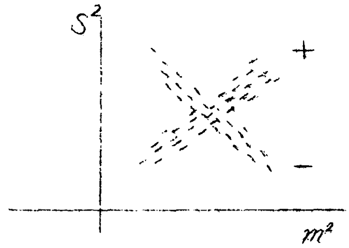
但し　m は母集団平均

$$\bar{x} = \text{Sample の平均} = \frac{\sum x_i}{n}$$

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

つまり、l^2 は母集団平均からのへだたりをあらはす一つの指標であり l^2 が大なるとき sample の平均は母集団平均との差が大なることをあらはす。したがつて S^2 と l^2 との相関係数をみることによつて l^2 が大なるとき S^2 が大か小か特別の傾向をもつかをみることが出来る。





計算の簡単のため無限母集団とする。

$$E(\bar{x}) = m \qquad E(S^2) = \frac{n-1}{n}\sigma^2$$

σ^2 は母集団 variance

$$\rho(S^2, l^2) = \frac{E\left(S^2-\frac{n-1}{n}\sigma^2\right)\left(l^2-\frac{\sigma^2}{n}\right)}{\sigma_{S^2}\cdot\sigma_{l^2}}$$

$$= \frac{E(S^2\ell^2) - \frac{n-1}{n^2}\sigma^4}{\sigma_{s^2} \cdot \sigma_{\ell^2}}$$

此をほごして一々計算すると

$$\rho = \frac{\frac{n-1}{n^3}(\beta_2 - 3)}{\sqrt{\frac{(n-1)^2}{n^3}\beta_2 - \frac{(n-1)(n-3)}{n^3}}\sqrt{\frac{1}{n^3}\beta_2 + \frac{(2n-3)}{n^3}}}$$

となる。但し β_2 は尖度をあらはす　したがつて $\beta_2 \gtreqless 3$ にしたがい

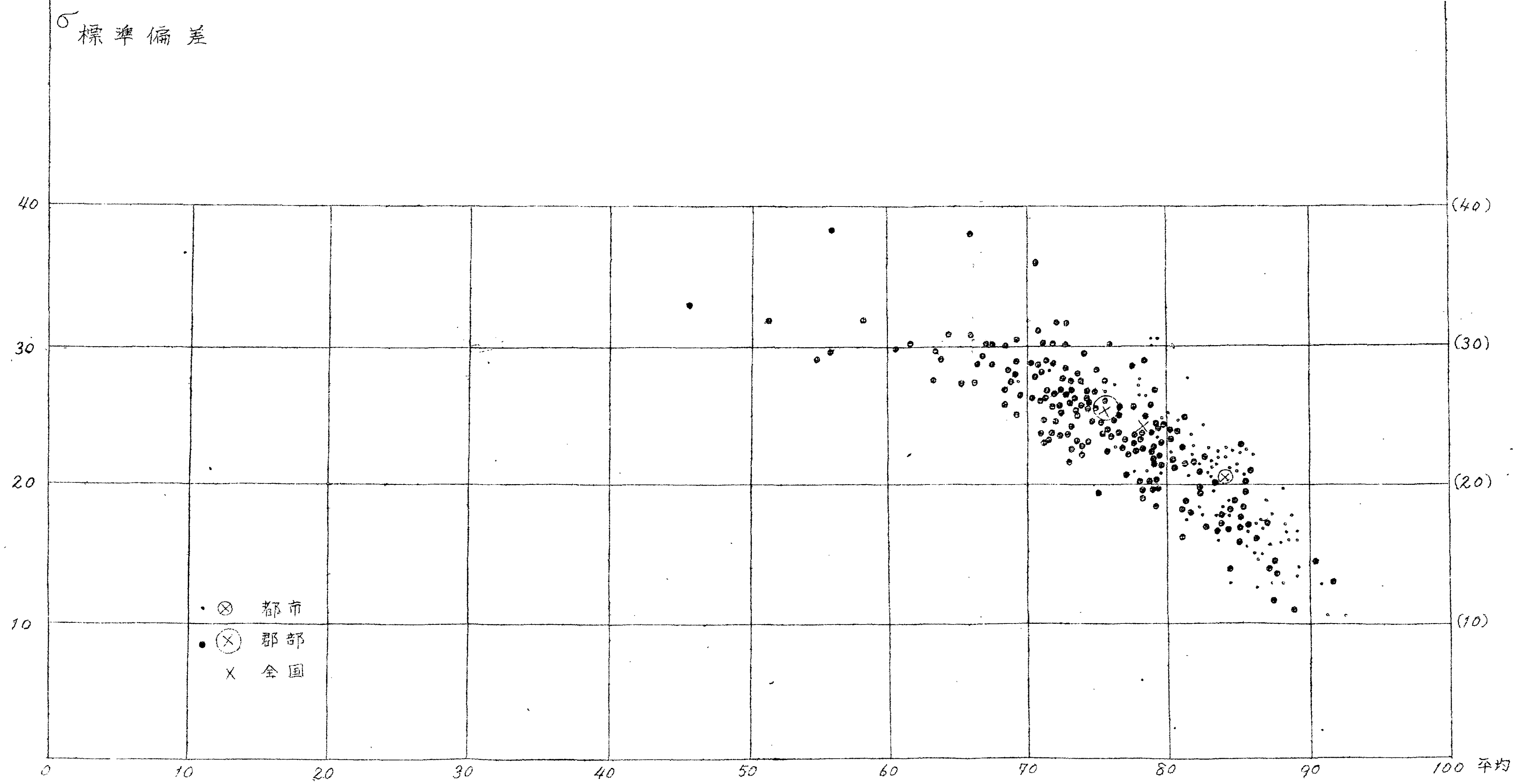
$\rho \gtreqless 0$ となることがわかる

即ち *normal* ならば $\rho = 0$　$\beta_2 > 3$ ならば　へだたり大ならば *sample variance* も大　$\beta_2 < 3$ ならばへだたり大なるほど *variance* は小となって出てくる。

此の傾向は *Sampling* による~~母~~集団平均の *Estimate* にさいし，信頼度，信頼巾の問題に関し注意すべきものであらう。

（*variance* の *over estimate*, *under estimate* に絡んで）

各調査地点別平均と標準偏差の表



㉘ 三二八・三二九頁間挿入

㉘ 三三〇・三三一頁間挿入

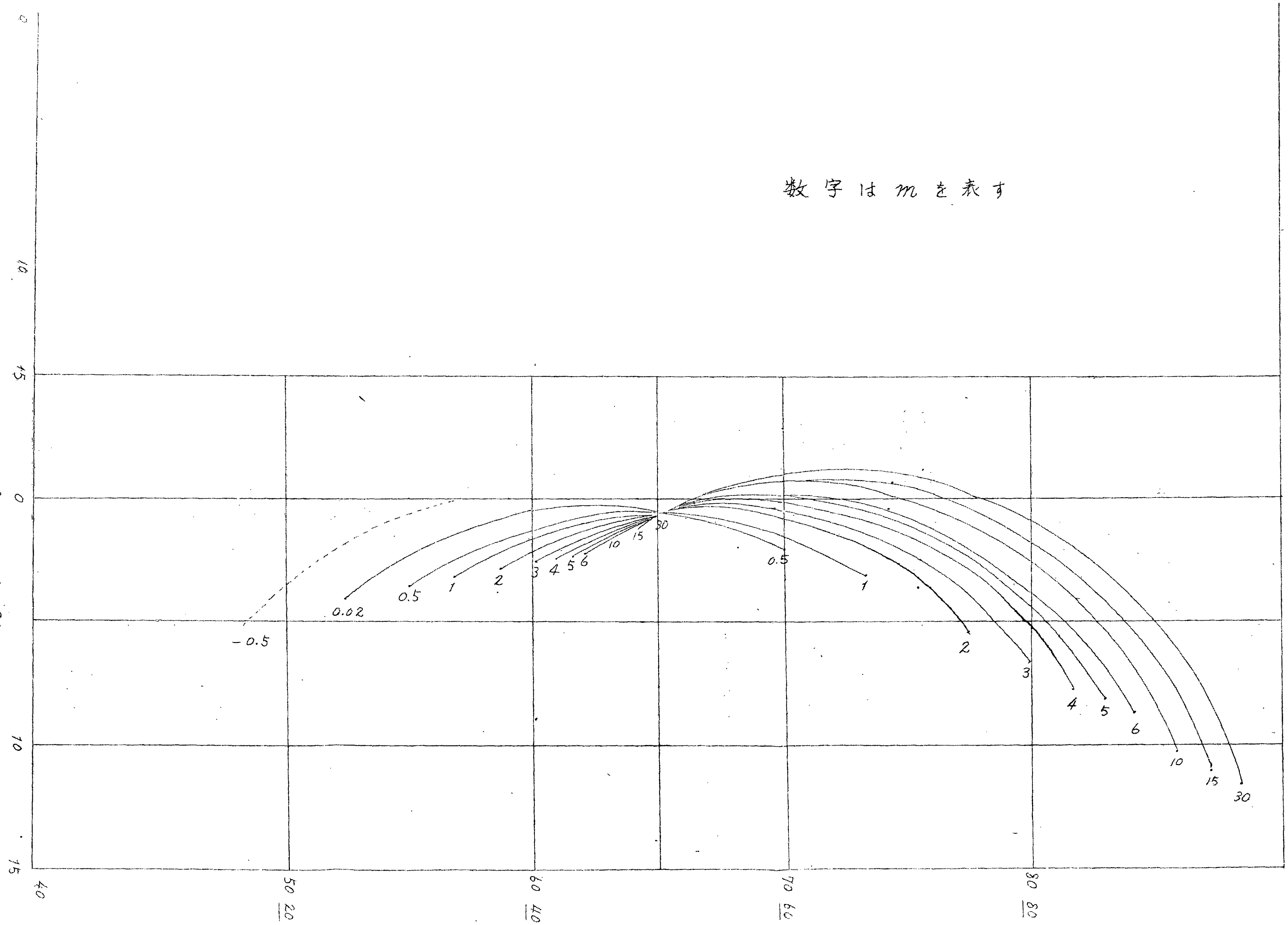
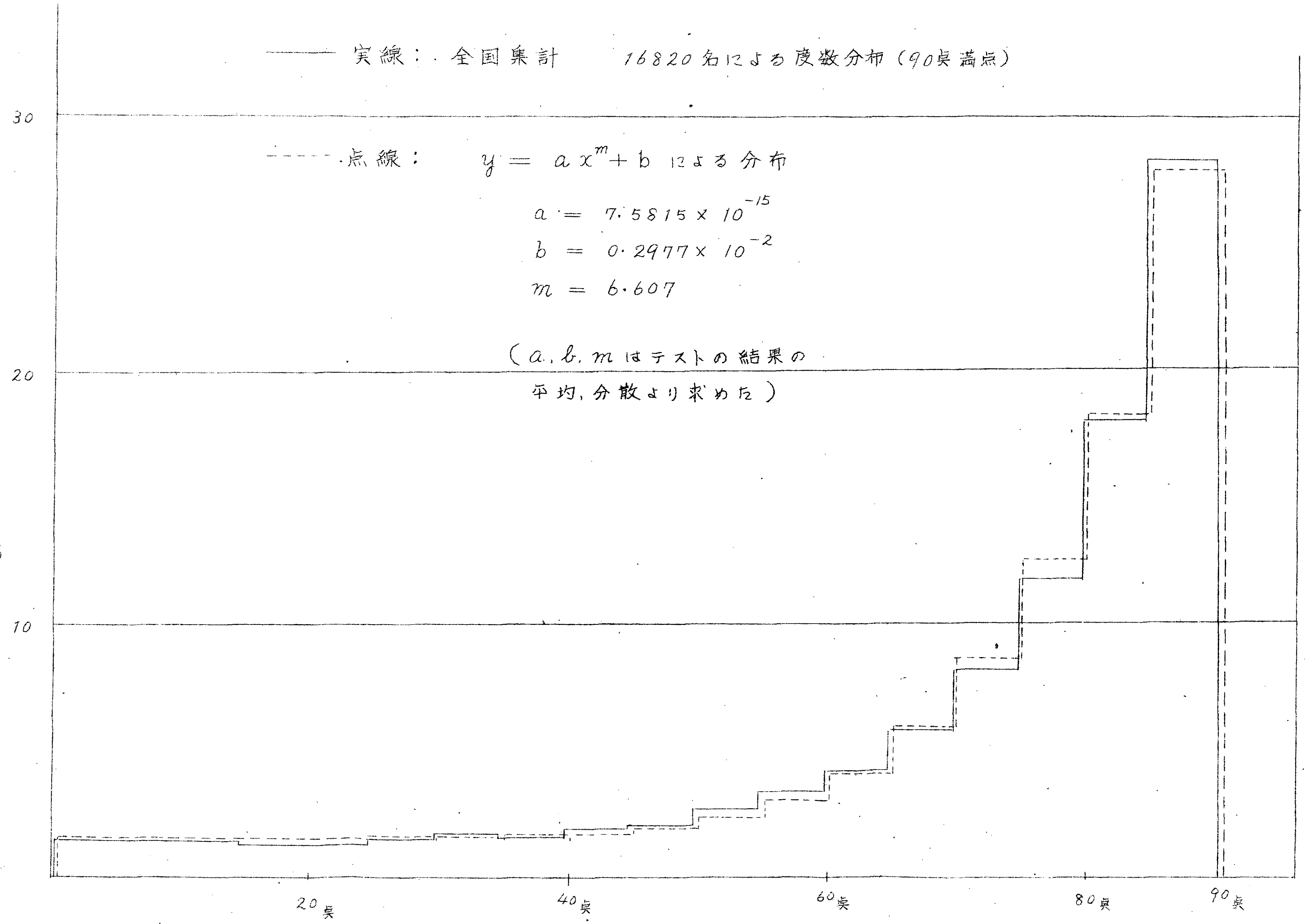
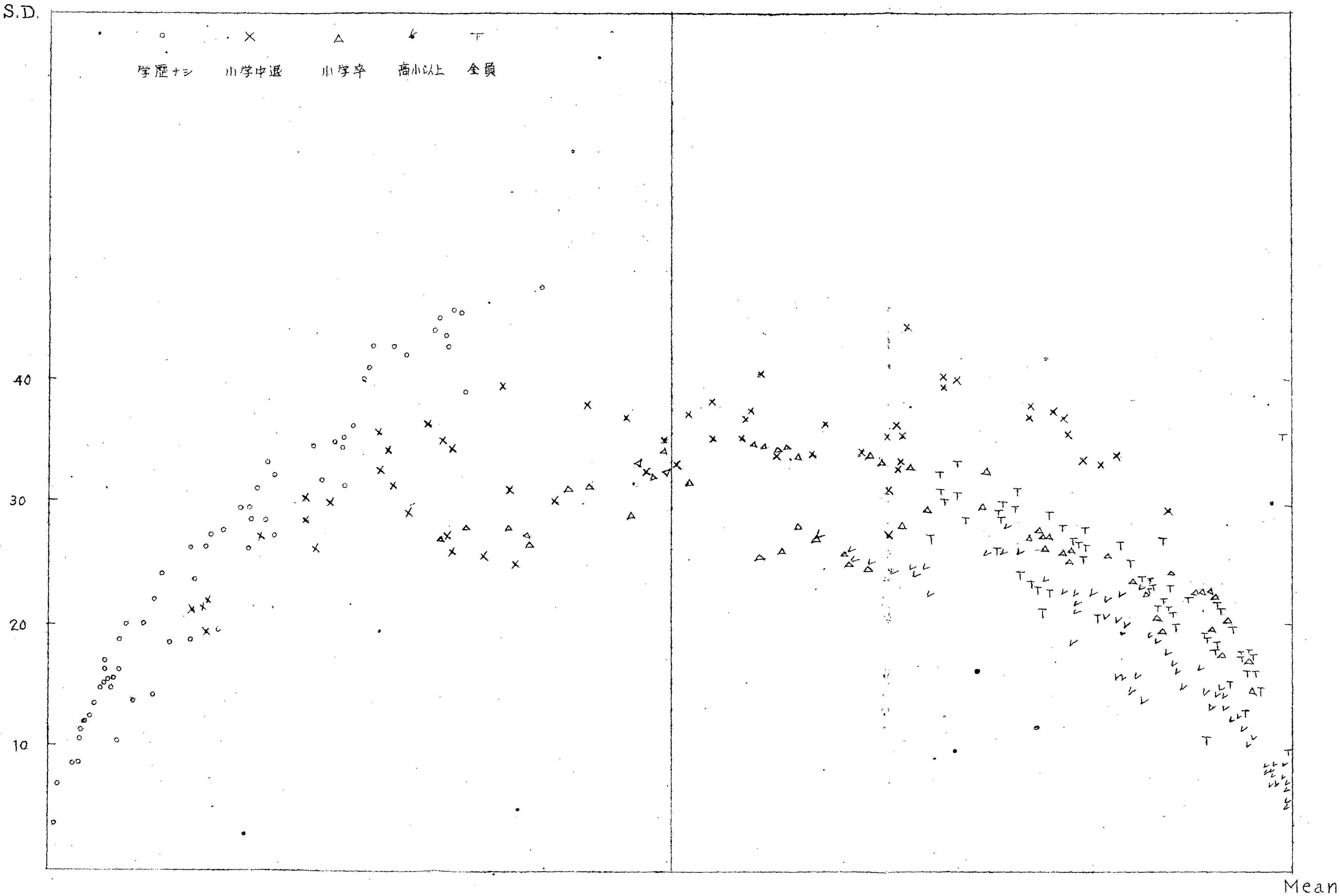


圖 C



㉘ 三三一・三三二頁間挿入

圖 D　　各 Break Down に應じた平均と標準偏差の關係圖表.



㉘ 三三二・三三三頁間挿入