

⑩ 鶴岡市言語調査に於ける サムフリング計画

林 知己夫

此の調査は前年度の白河市言語調査につゞくものであり、国立国語研究所、統計数理研究所との協同の研究である。調査研究費は昭和25年度文部省科学試験研究費によるものである。統計数理研究所からは林の外、青山博次郎、西平重善、田熊雅子が参加している。

研究の目標は言語生活の実態把握、特に共通語化の程度の把握、mass communication の現状把握と云ふことである。

調査の方法は熟練せる調査員による面接調査である。但しあらかじめ準備された調査票にもとづいて行われるのである。

このため調査員は現地にあらかじめ出向、各人の規準を一定にし、耳の訓練をほどこしてある。

調査票は準備調査により十分検討されたものであり、且つ白河市に於ける調査の知識をもとにし、さらに mass communication の事項をもつたものである。内容は大別すれば共通語化の要因に関するもの例へば個人の生活歴、生育歴、生活情況、社会生活の仕方、言語意識、等、及び共通語化の程度

測定に関するもの、例えば音韻現象について、アクセントについて、語彙、語法についてであり、その他新語に関するものが盛られてある。

各被調査者の此の調査票に対する反応を総合分析して所期の目的を達しようとするものである。

なお、今度は、被調査者の一部にロールシャッハテスト (Projective method) にほどこし性格調査を行い、性格、共通語化に及ぼす影響 (及ぼし方) をもあわせ調査し、調査内容をよりダイナミックなものとしようと試みられた。

鶴岡市の選ばれた理由としては東北方言地方で、白河市と條件が似ており、さらに北にある事、調査の便宜の得られること等であつた。

この調査もやがて行なうべき大調査の一つの準備的研究調査である。

鶴岡市の人口は43858人である。此のすべてについて調査を行ふのは無駄な事である。したがつてサンプリクによつて選ばれた一部の人を調査し鶴岡市全体の様態を推定することにした。以下この計画をのべてみようと思う。

なおこの準備調査は昭和25年10月11日～16日に行われたものであり、サンプリク計画は此れ以前に立てられたものである。

§ 1 一般的なこと

調査の目的は、鶴岡市に在住する人々によって織りなされている言語現象を一面から捕捉し、言語の共通語化に及ぼす要因を見出し、あわせて新語の滲透してゆく様相をとらえようとするものである。

このためには白河市の調査計画の所に於てものべた様に熟練した調査員を必要とする。この調査は現地では得られないと言ふ事は白河市の調査で明らかになった。爲に調査員はすべて研究所の Member を以て充てることにした。費用、日時などあわせてみると昨年度同様サンプルは500程度と言うことになる。

調査員は7人、一日の調査数は昨年の経験により一人約10人(調査実施の時間は一人につき約40分以内、調査票は此の意味の adequacy あるものにしてある!) 調査日数は8日、こう考えると

$$7 \times 10 \times 8 = 560$$

となり、follow up (Call back) を勘案するならば被調査としては500人程度が精々の所になるのである。

以上の様な制限を考へて入れて能率のよいサンプリング計画をたてねばならない。

§ 2. 調査対象と母集団

調査対象は白河市調査の場合と同様、昭和25年10月から11月の間鶴岡市に在住し正常な社会生活を営む明治15年から昭和11年の間に出生した男女(数之年69才—15才)すべてであるとする。

なお鶴岡市と言つたが、新に市に編入された農村地帯、道形、文下、茅原の部落はのぞいてある。

この人々の標識は調査票に対する各種の反応のさまであるとする。各人に等しい抽出確率をあたえて母集団を構成する。

この時、各種の母集団平均(特に共通語化の程度)を推定するものとしてサンプリング計画をたてようと思う。

500のサンプルがあるならば、どのくらいの精度が得られるかは白河市のサンプリング計画に於てのべた通りである。

例えば、母集団の変異係数が0.05, 0.03であれば、サンプル平均の変異係数は95%の信頼度の下では4.4%, 2.6%となる。

§ 3. サンプリングの方法

さらに精度をあげるためには層別をほどこせばよい。共通語化の程度に強い影響をあたえるのは、生活歴、父母の出生地、学歴、年齢男女、職業、新聞の利用度、等であった。しかし、これらのすべてのものはあらかじめ解つてはいない。前以て利用できる資料は、男女年齢、職業、居住地である。なお特に年齢は種々の分析を行うとき基本となるものであることが判明しているので(生活、生育歴、生活様式、新語の滲透の相相等々の点で)層別の規準として採用することは絶対必要である。従つて年齢を第一としこれらのものを層別の目じるしとすることにした。

我々の調査対象地(旧市内)の人口は41817であり、対象内の比率は白河市の場合、63.7%であるから、此の市でも此の程度と考えるならば(残念乍ら、鶴岡市での対象の比率は判明していなかつた)26650となる。

此の人々は物資配給台帳の中に記入されてあるのであり、我々の直接実際の調査の対象とするの北この台帳に記載されている人々なのである。

この26650人の人々を台帳の中から拾い、カードをつくり層別することは調査員三人二日間では全く不可能である。

したがつて重ねぬき法を用いて層別の効果をあげねばならないと考えられる。

つまり第一次サンプルとして m 人抽出し、その人々を判明した資料によつて層別し、結局のサンプル n 人を抽出する方法である。

即ちまづ現地で台帳から m 人抽出し、その人々のカード(くしかるべき事項記入)をつくり、此を東京へもちかへり層別し、サンプルを決定す

るのである。

それでは第一次サンプル m 人をいかに定めたらよいであろうか。

なお鶴岡市に於ては男女別年齢別人口は調査企画の時期に判明してお
らなかつたので、自河市の場合の様な層別の効果（第一次サンプルから
第二次サンプルを抽出するも、層の母集団人口比率が判明しておれば完
全な層別の効果を得ることが出来る）を期待することは出来ない。

註、準備調査に出向した時は、昭和23年度男女別、人口別、比率が
集計されたものがあつた。然し市としては資料が古いし——
転入転出が相当あるから——対象外の地域も加えられてある。
この古い資料を用いて自河市の様な方法を採用するとき現在の母
集団に対する推定をつくるならば偏りが入りこむことになる。
この偏りは修正のしようがない。したがつてこの資料は参考程度
につかつてみることにした。

さらに又共通語化の程度は居住歴に大に關係してくるので、要因分析
の立場からは若年よりも壯年老年のサンプルが多く必要とされてくる。

したがつて、層別サンプルリングの場合の比例割当ての立場が多少変更
されるのが望ましくなる。しかし全体の集計が比較的簡単にされる様に
心掛けねばならない。

註、自河市の重ねぬき法の場合も第一次サンプルの層別後比例割当て
法によりサンプルが抽出されていた。此の方法を少しく改良す
ることが考へられねばならない！

以上の様なことを考へて m を決定してゆくのである。

今若年の比率を p_1 と考へよう。 p_1 を推定するには m 個のサンプ
ルとるとき平均

$$m p_1, \text{ 分散は略 } m \sqrt{\frac{p_1(1-p_1)}{m}} \quad \text{となる様な比率 } p_1 \text{ で推定}$$

される。

今若年のサンプルは比例割当て法で得らるべきサンプルの $\frac{1}{2}$ しか割
当てぬものと考えよう。

比割割当法でなくともこの程度ならば集計にはさして困難を來さぬであらうと考えられる。

大き m の第一次サンプルを層別してから一まづ第二次サンプル n' を抽出するものとする。こうしてから n' 中の若干のサンプルを n' にすることを考ふる。こうして結局調査すべきサンプルを 500 程度におさめることを考ふるのである。此は調査実施能力から言つて重大な條件である。

母集団で p_1 の比率をしめるものが p_1' で推定されるものとする。さうすると調査すべき数は

$$\frac{n' p_1'}{2} + n' (1 - p_1') = n' \left(1 - \frac{p_1'}{2}\right) \quad \text{で推定}$$

される。信頼度 99% とするとき、此の数は

$$n' \left\{ 1 - \frac{1}{2} \left(p_1 + 3 \sqrt{\frac{p_1(1-p_1)}{m}} \right) \right\},$$

$$n' \left\{ 1 - \frac{1}{2} \left(p_1 - 3 \sqrt{\frac{p_1(1-p_1)}{m}} \right) \right\}$$

と推定される。

m と n' とはついでグラフを書いてみると次の様になる。なお若年とは数え年 15才 — 24才までのものを考ふることにした。此の比率 p_1 は鶴岡市では豫めわかつて居らぬので白河市の資料を用いた。 p_1 は 33% である。

次に、この関係をさらにみやすくするため

$$n' \left(1 - \frac{p_1'}{2}\right) = A \quad A = 450, 490, 510, 520$$

$$p_1' \text{ として } p_1 \pm 3 \sqrt{\frac{p_1(1-p_1)}{m}}$$

とにおいて m と n' とのグラフをかいてみると次の様になった。

此等のグラフをながめ、又サンプリング実施の時日、精度の計算をながめ

$$m \doteq 1700$$

$$n' \doteq 600$$

と一応定めてサンプリングを行うことにした。

§ 4. サンプリングの精度

この場合は全くの重ねぬき法によるのである。まづ、大きさ m の第一次サンプルを抽出、此を R 個の層に分つ。 i 層に属する第一次サンプルの大きさを m_i とする。

$$m = \sum_{i=1}^R m_i$$

第二次サンプル n を抽出するのは i 層から n_i 本のサンプルを抽出するものとし

$$\frac{m_i}{n_i} = k$$

と考ふるならば、母集団平均 \bar{x} (共通語化の程度をあらわす点数の母集団平均) の推定としてつくる偏りのない推定値 \bar{x}

$$\bar{x} = \sum_{i=1}^R \bar{x}_i \frac{m_i}{\sum_{i=1}^R m_i}$$

\bar{x}_i は i 層のサンプル平均

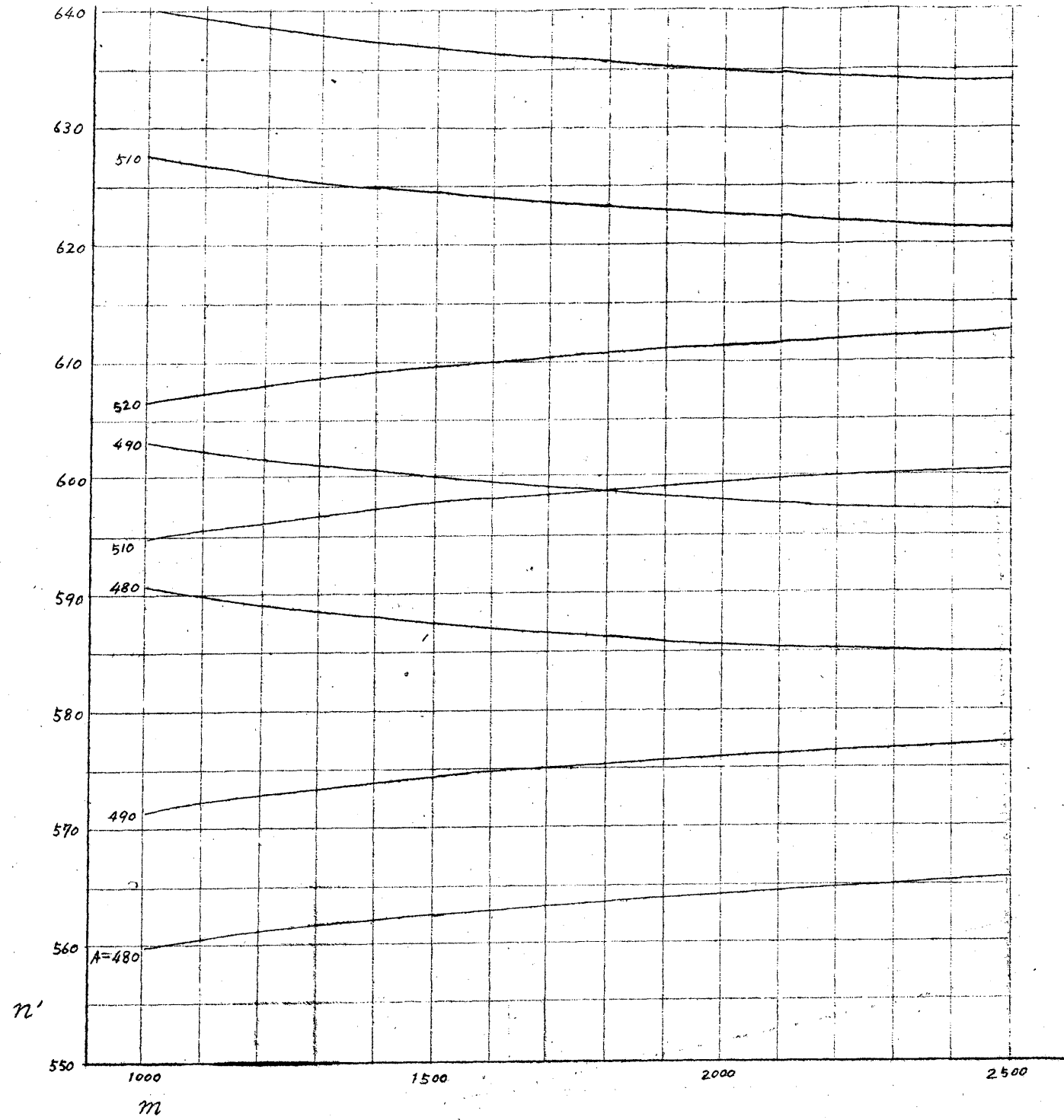
の分散 $\sigma_{\bar{x}}^2$ は、近似的に

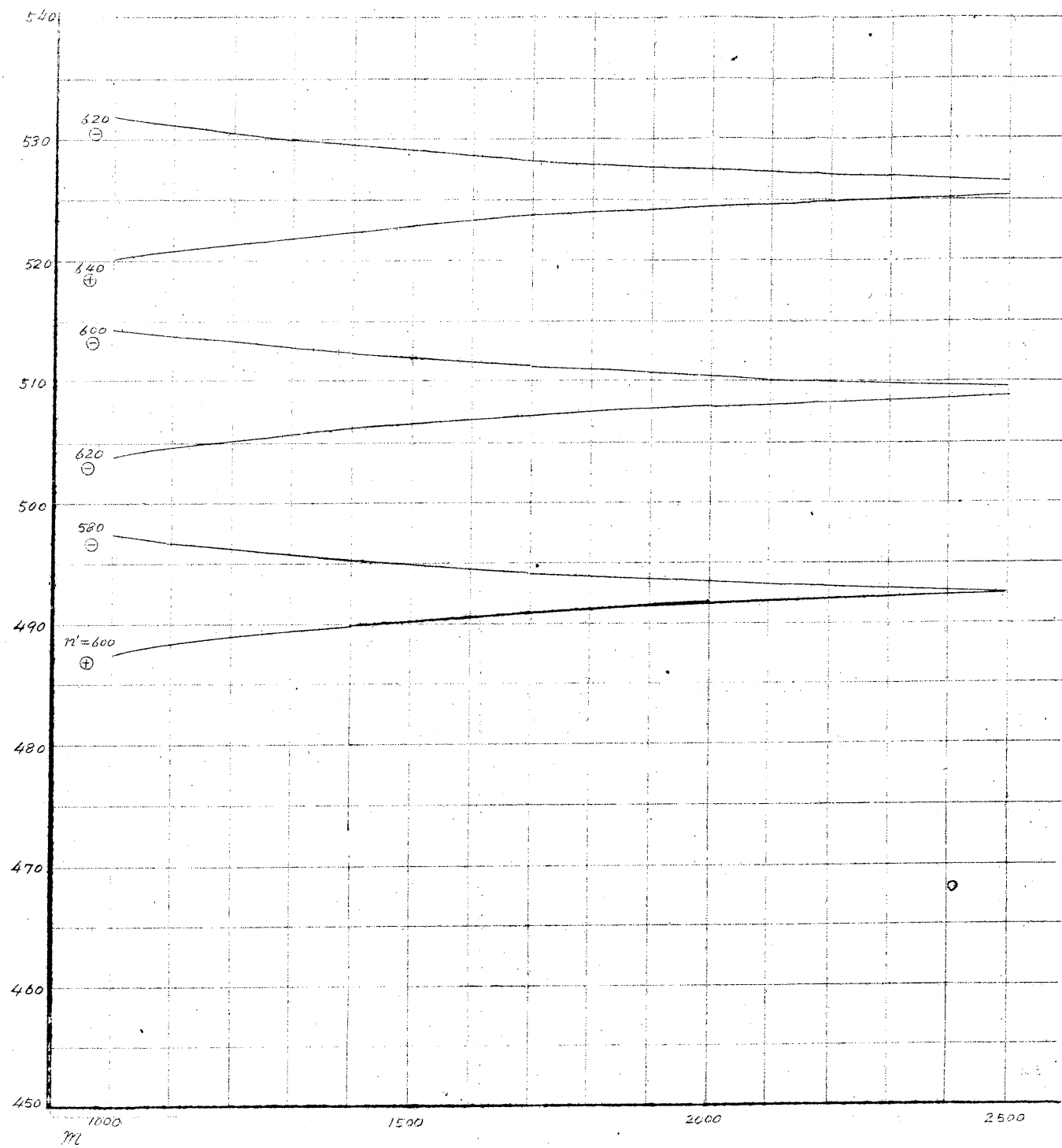
$$\sigma_{\bar{x}}^2 \doteq \frac{\sigma^2}{n} - \sigma_b^2 \left(\frac{1}{n} - \frac{1}{m} \right)$$

となる。ここに

$$\sigma_b^2 = \sum_{i=1}^R p_i \bar{x}_i^2 - \bar{x}^2$$

折込 (115頁/次) 外





\bar{x}_i は、第 i 層の平均

N_i は、第 i 層の大きさ

$$p_i = \frac{N_i}{\sum_{i=1}^R N_i}$$

である。

さて我々の場合は

$$\frac{m_i}{n_i} = k \quad (\text{であり})$$

$$m_i = a_i n' \frac{m_i}{\sum_{i=1}^R m_i} \quad \text{であるから}$$

$$\sigma_{\bar{x}}^2 \doteq \frac{1}{n'} \sum_{i=1}^R \frac{1}{a_i} \sigma_i^2 p_i + \frac{\sigma_b^2}{m} \quad \text{となる。}$$

我々の場合は

$$a_1 = \frac{1}{2}$$

$$a_i = 1 \quad (i \neq 1)$$

であるから

$$\begin{aligned} \sigma_{\bar{x}}^2 &\doteq \frac{1}{n'} 2 \sigma_1^2 p_1 + \sum_{i=2}^R \sigma_i^2 p_i + \frac{\sigma_b^2}{m} \\ &= \frac{1}{n'} \sum_{i=1}^R \sigma_i^2 p_i + \frac{\sigma_b^2}{m} + \frac{1}{n'} \sigma_1^2 p_1 \\ &= \frac{\sigma^2}{n'} - \sigma_b^2 \left(\frac{1}{n'} - \frac{1}{m} \right) + \frac{1}{n'} \sigma_1^2 p_1 \end{aligned}$$

となる。

$$m = 1700, \quad n' = 600 \text{ とすれば}$$

$$\sigma_{\bar{x}}^2 \doteq \frac{1}{n'} \left(\sigma^2 - \frac{2}{3} \sigma_b^2 + p_1 \sigma_1^2 \right) \quad \text{である。}$$

唯共通語化の程度の推定と言ふ場合には、この様なサンプリングを行うとき $q_i = 1$ ($i = 1, \dots, R$) のサンプリングと比較して、精度の上でどうなるであろうか。

$\sigma_{\bar{x}}^2$ と $\sigma_{\bar{x}'}^2$ との比較である。

$\sigma_{\bar{x}}^2 - \sigma_{\bar{x}'}^2 > 0$ ならば得をすることになるのである。

$$\frac{\sigma^2}{n} - \sigma_b^2 \left(\frac{1}{n} - \frac{1}{m} \right) - \frac{\sigma^2}{n'} + \sigma_b^2 \left(\frac{1}{n'} - \frac{1}{m} \right) - \frac{p_1 \sigma_1^2}{n'} > 0$$

今 $\frac{n'}{n} = 1 + \beta$ とおくと

$$\beta > \frac{p_1 \sigma_1^2}{\sigma^2 - \sigma_b^2} \quad \text{ならば有利と言うことは}$$

なる。

$$\beta > \frac{p_1 \sigma_1^2}{\sum_{i=1}^R p_i \sigma_i^2}$$

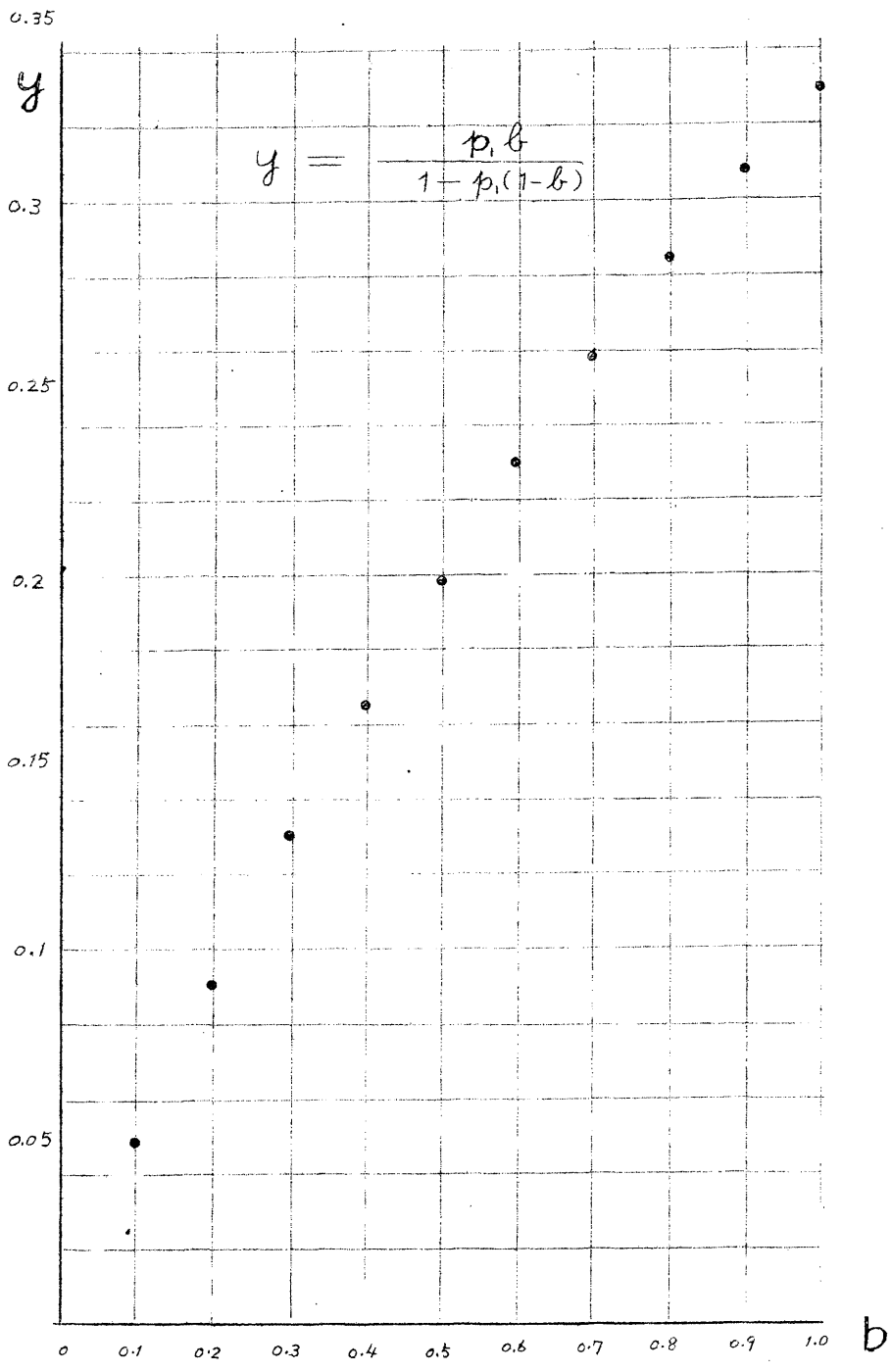
$$\begin{aligned} \text{今もし} \quad \sigma_1^2 &= b \sigma^2 \quad b < 1 \\ \sigma_i^2 &= \sigma^2 \quad (i \neq 1) \end{aligned}$$

とおくと、右辺は

$$\frac{p_1 b}{1 - p_1(1-b)} \quad \text{となる。}$$

このグラフをかいてみると次の様になる。

但し、 $p_1 = 0.33$ とする。



白河市の共通語化の程度の資料を用いて β の見当をつけてみる

年 令	\bar{x}_i	σ_i^2
15—19	20.9	15.527
20—24	21.5	13.293
25—29	20.2	17.326
30—34	20.0	16.127
35—39	19.0	15.440
40—44	19.3	13.760
45—49	18.8	16.028
50—54	16.9	24.094
55—59	16.7	20.298
60—69	17.4	22.360
全 体	19.6	19.85

比より β の右辺を求めてみると

$$\beta > 0.290$$

の時有利となるのである。

$$n = 500 \quad \text{ととるとき}$$

$$n' > 645 \quad \text{の時有利となるのである。}$$

我々の場合のサマリングは共通語化の程度の測定のみが目的でなく、複雑な要因分析（此のために総サンプル一定の下に壯年老年のサンプルを比較的多くとる必要がある）を行うためにも好都合でなくてはならぬ様に社組んであるのである。分析の立場を重んじてもなほ且共通語化の程度の推定でも損をしたくないのである。このためには n' として 645 以上を必要とするのである。

しかし、645 の n' をとると

$$n \left(1 - \frac{p_0}{2} \right) \doteq 539$$

となり調査能力を超える心配がある。

以上の様な考察からさう損をしないでしかも我々に都合よいサンプルを抽出するには

$$m = 1700, \quad n' = 600$$

程度が適当であろうと考えられた。

なお、第一次サンプルからの抽出サンプル n' とするとき、結局の調査サンプル数の平均は $n' \left(1 - \frac{p_1}{2}\right)$ 、

標準偏差は $\frac{n'}{2} \sqrt{\frac{p_1(1-p_1)}{m}}$ となるから

$$n' = 600 \quad \text{とすると}$$

平均は 510、標準偏差は 3.6 であり、調査能力の範囲内にある。

§ 5. サンプルリングの実施

以上の理論的考察によつて重ねぬき法が有利である（此の程度は数量的には求めてはないが）と言ふ保証を得、且又壯年若老年者の分析により豊富な結論を出すことが出来るのでそれらのものについてしめるべきサンプル数を確保する意味からも重ねぬき法を用ふるのが我々の場合妥当であろうと思われる。

このため第一次サンプルとして約 1700 を抽出することに決定した。

抽出は例により物資配給台帳からのランダムスタートの等間隔抽出法によろうとした。

前述の様に村界内の母集団総数は 26650 と推定されるので、1700 のサンプルを抽出するためには 15.5 の間隔で抽出する必要がある。なお、市の平均家族人数は 4.6 人であるから此の間隔で抽出しても同調することなくゆがみのない抽出が可能であろうと思われる。

15.5の間隔で抽出するために最初は15の間隔，次は16の間隔
 次は15の間隔と言ふ風に抽出してゆくのである。

註 此の方法は実施上極めて困難を感じしめるのであつて，抽出
 能率は等間隔抽出法によるときよりもはるかに落ちることが
 見出された。多数サンプルを抽出するときは極めて単純な
 機械的の操作が望ましいのである。

さて，抽出はその様な方法で行い，指定された番号のものが対象内
 のも即ち昭和11年—昭和15年のものであれば，サンプルとして
 抽出，さうでないものはそのままにして更に次の間隔で数えてゆく様
 にした。サンプルについては性，年齢，住所，職業（家の職業），
 移動年月日，氏名，を次の様なカードに書き込んだ。

男 / 女	明治 大正 昭和	年	昭和	年	町	方		
			職	業			移	動
特調(50)-1						世帯主の職業		
<div style="border: 1px solid black; width: 150px; height: 60px; margin: 0 auto;"></div>								
被調査者 No.						家族人数		

§ 6. 抽出した結果

抽出した結果第一次サンプルとして、男789、女1021、計1810を得た。

予定したサンプル数より多く抽出されすぎた様に思われるが、此は白河市よりこの市の方が対象内の母集団が多いためであることが判明した。

なお、抽出後昭和23年の男女別人口が入手されたので参考のため、サンプルとその比率とを比較してみよう。

註。この資料は古いのもその後の変化を考えてサンプルの割当てについては用いなかった。不測の偏りを恐れたためである。まづ、男女比についてしらべてみる。

	サンプル	昭和23年
男	56.4	55.1
女	43.6	44.9

左表の様になり χ^2 検定を行うも

$$\chi^2 = 1.2 \text{ (自由度 1)}$$

となり有意な差はみとめられない。この市は女の比率が多いのは注目すべきである。

次に年齢別の比較を行ってみる。

数え年	サンプル	昭和23年
15—24	28.0	30.4
25—34	22.5	22.0
35—44	20.9	18.9
45—54	14.4	14.9
55—69	14.2	13.8

左の様により一致を示しており前述の様な抽出法は偏りを興へないとの一つの保証を得たことになる。

さて前には $p_1 = 33\%$ として種々の計算を行ったのであるが、昭和23年の資料をつかうと、 $p_1 = 30\%$ であり、サ

ンプル抽出の結果、第一次サンプルの大きさは1810となつたのであらためて n' の計算を行つてみることにした。このため

$$n' \left(1 - \frac{p_1'}{2}\right), \quad p_1' = p_1 \pm 3\sqrt{\frac{p_1(1-p_1)}{m}}$$

$$p_1 = 0.3$$

ととり n' と m とのグラフをかくと次の様になつた。(別葉)

此のグラフからみて $n' = 580$ 見当が適當であるうと考へられた。
かうすると調査すべきサンプルの平均は

$$n' \left(1 - \frac{p_1}{2}\right) = 493,$$

$$\text{標準偏差は } \frac{n'}{2} \sqrt{\frac{p_1(1-p_1)}{m}} = 3.1$$

となり、調査能力を超えぬ様になる。

なおこの資料を用いて、 β を計算してみると

$$\beta > 0.27$$

の時、重ねぬきの方が有利となる。従つて我々のサンプリングによる
とき母集団の共通語化の程度推定に對してはやゝ精度は落ちると結論
せられる。然し他の有利の点を考へるとこの程度でも満足すべきもの
と思われる。

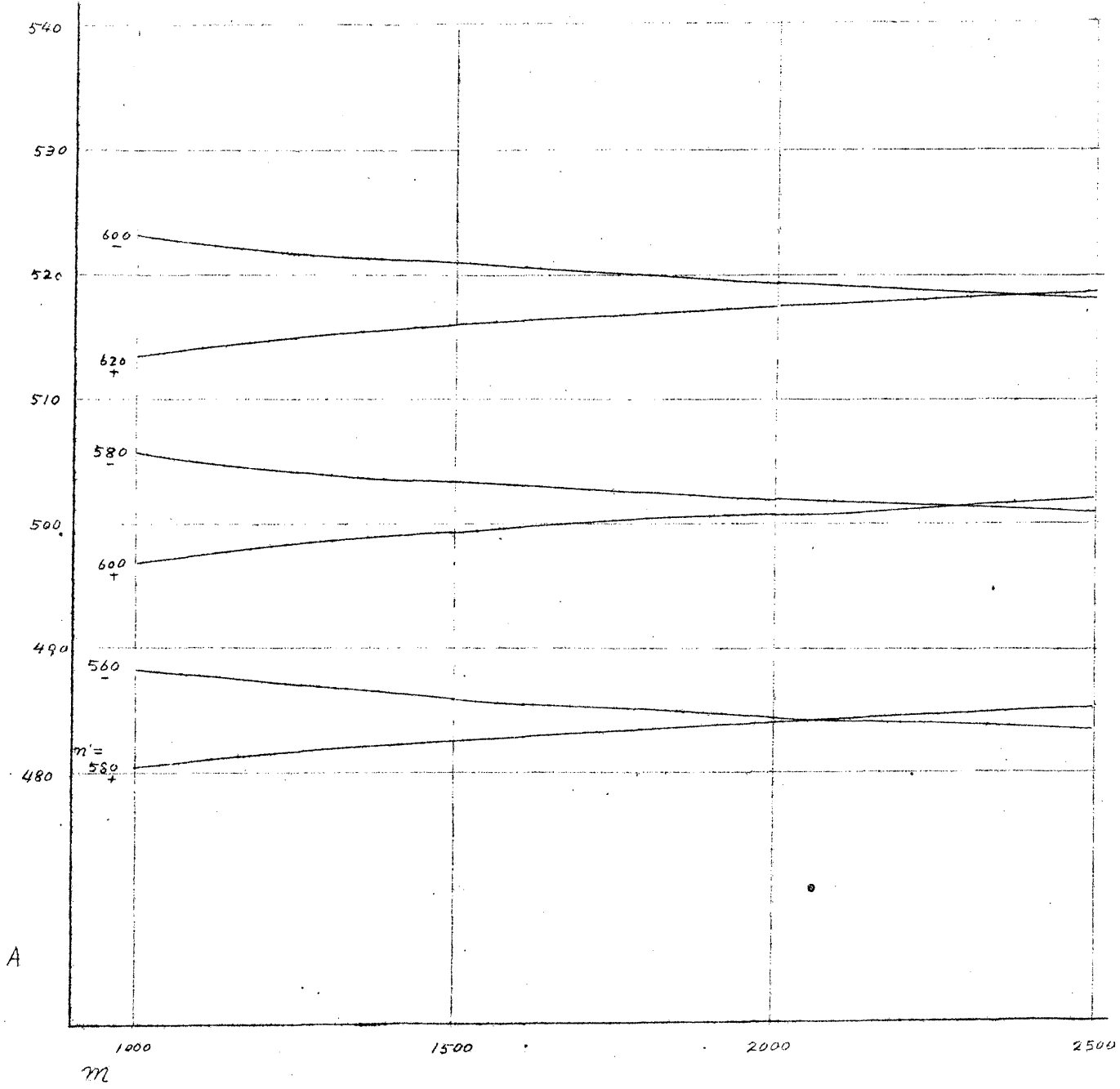
§ 17. 第二次サンプルの抽出

得られた第一次サンプルから第二次サンプルを抽出することになる。

層別は、まづ男女別、年齢別(15~24 25~34 35~44
45~54 55~59)に分類、次に最近移動してきたものを分類
する。この四段階(住宅地、商業地、小商工業住宅地、農業地)に
分けた居住地で分類さらにはその中を職業(専給生活者、商業、工員、
士、農、主婦、学生、無職)によつて分類した。かう
してから580のサンプルを得る様に1809の第一次サンプルから
等間隔抽出法の考へで抽出することにした。この場合間隔は37を
1サイクルとし、3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 4の間隔
を計37

折込 123 頁 の 況 (件 画)

$P_0 = 0.30$



の繰返しによつて第二次サンプルを抽出した。

結果は、584を得た。次に15才～24才のサンプルを $\frac{1}{2}$ とした。その結果、調査すべきサンプルとして

15～24	83	
25～34	115	
35～44	111	
45～54	83	
55～69	82	
最近の移動	27	但し15～24をのぞく
計	501	

を得た。

註. このようにして得られたサンプルを集計するときには若年の分のサンプルの調査^票を二つ同じものを作り584のサンプルとして通常の通り集計すればよい。一々ウエイトをかけるのは面倒であるからである。この様な簡単な操作で集計の方法がたてられるなら強いてサンプルの割当ては比例割当て法によりなくともよいと思われる。

§ 8. ロールシャツハテストのサンプル

ロールシャツハテスト (Projective Test) を用い、パーソナリティを刻定、共通語化の程度の変因として如何なる内奥的な働きをなすかを見出すこととした。此のテストは一人一時間程度を要し、調査員の関係からさう多くのサンプルをきめることはできない。この能力の最大限100を以てサンプルとしなければならない。

なお、この調査は共通語化の程度などをみる調査票による調査が終了してから始めることにした。

このためサンプルを、性、年齢、学年の外、居住地、共通語化の程度等によつて層別、重ねぬき法の考へはしたがつて（しかも調査前は知られなかつた重要な性質を加へることゝできる！）100のサンプルを抽出することにした。

此は調査しつゝ整理し、サンプルを抽出すると言ふ方法である。