予測に関する実證的研究
——選挙予測の方法論——

林　知己夫
高倉節子

(1964 年 10 月受付)

Statistical Methodology for Prediction of Election Polls

Chikio Hayashi
Setsuko Takakura

The authors present a statistical methodology for prediction of election polls based on the large scale sample surveys done before the election for those who are qualified to vote. The following data are used: those for the Election of Members of the House of Representatives (1955, 1958, 1960, 1963) and those for the Election of Members of the House of Councilors (1956, 1959, 1962). They give not only prediction of the number of candidates who are elected in every political party but also the probability of each candidate being elected from statislical analysis of the survey data obtained for eaeh candidate, especially using the theory of quantification of qualitative characteristics which has been developed by Hayashi. They confirm the reliability and validity of their methodology.

Institute of Statistical Mathematics

目　次

§1. 選挙予測について

予測の方法論については，すでに本彙報15周年記念号(第7巻1号，1959)の「数量化と予測の根本概念」において述べておいた．この考え方の大綱については，変更を加えるべき所はない．ここでは，このような予測の考え方に立った具体的な研究を示してみようと思う．予測に対するわれわれの狙いは「目的に対して妥当性ある知識」という観点から，「科学的方法によって正鵠を得た情報」——これには誤差の表明があり，かつその誤差が目的に応じて可及的に少ないこと——を与えることであり，この方法を鍛え上げることにある．前にも書いたことであるが，予測の方法論の厳しい鍛錬のためには——特に社会現象の予測——結果が明確に対応づけられるものをとりあげて研究することが絶対に必要である．こうしてつくられる方法論

であって始めて，明確でないものに対しても妥当性を予想できるものとなる．このような観点から，われわれは結果のはっきりする犯罪現象における仮釈放の予想——犯罪者の将来の行動予測——や選挙予測，その他市場調査のある種の問題をとりあげて研究を進めてきたのである．なお，われわれは方法論的成果はもちろん具体的な結果にも興味をよせている．

さて，選挙予測であるが，われわれは選挙投票前の調査をよりどころにする．過去の調査と選挙結果をつきあわせ分析することによって情報を得て——つまり調査と結果との確率的な函数関係を求める——将来行なう調査により予測を行ない，その予測の精度を明らかにしようとするものである．ここでは，衆議院議員の選挙，参議院議員の選挙，知事選挙を素材として，選挙予測の方法論をくみあげてみようと思う．なお，選挙予測における根本的な考え方は，林知己夫: 選挙予測と言うこと（思想12月号 1962, 50頁—62頁），選挙予測について（MBC クォータリー．No. 5, 1960/autumn）: に詳述してあるので，ここでは繰り返えさないので参照されたい．

さて，前述のように，選挙予測には有権者に対する事前調査（もちろん確率標本調査）が必要である．ある一定の調査方法によって得られたデータに従えば，ある結果が示されるが，これそのものが必らずしも効率のよい予測をあらわすものではない．調査結果がそっくり選挙結果を指すものでもない．ある時点の調査結果を用いて，選挙結果を推定するという立場に立たなくてはならないのである．このために，過去における調査——現在選挙予測のために行なった調査方法と同一でなくてはならない——とそのときの選挙結果をつきあわせ，調査から選挙結果を推定する方式をあみだしてゆかなければならない．しかも その方式自身が，いつも同一ないしは近似的に等しいとみなせる程度安定しているものでなくてはならないのである．逆にいえば，安定するような推定方法をつくり出さなければならないのである．この方法の大要を次に示しておこう．

実際の選挙では10票や20票，あるいは数百票程度で当落が決まることもあるので，これを標本調査でいいあてようとすることは根本から無理なのである．このようなとき，候補者の誰が当選するか判定できないという結論が当っている．したがって，当選者の名をすべて打ち出そうとすると必然的に予測の的中率の低下がおこってくるわけである．こうしたわけであるから，決定的な言い方ではなく，当落の確率で以てものをいわなければならなくなる．つまり各候補者の当選確率というものを推定しなければならなくなる．こうして，当落の確からしさの程度を確率をもって表現するように試みるのである．どうしても予測には誤差の表現がつけられてくることになる．誤差のない予測が的中したとすれば，調査に基く限り，偶然というほかはない．なお，ここに当選確率といったが，これは，当選確率 p をもつものが n 人あるとすれば，当選者の平均は np で，その分布が一応二項分布をするという意味に解されたい．

さて，まず調査は何時行なうべきかの問題があるが，投票日から遠い方が興味深いが，あまり遠くては意見がかたまっていないし，あまり近くては，調査実施や分析が困難である（調査の安定性をみるためにも2回程度行なうのが望ましく，1回は遠く，1回はかなり近く行なうのがよいと思われる）．さて，調査方法は衆議院選挙では各選挙区ごとに行ない——党派別当選総数の予測だけではこの必要はない，各候補者の当選確率の推定には，すべての選挙区調査が必要である——参議院選挙，知事選挙では各県の調査が必要となる．これについては，各論のところを参照されたい．このようにして抽出された有権者に，誰に投票するかについて面接法によって質問するのである．このときいろいろな調査法があるが，ここでは，候補者のリストを見せないで質問することにする——候補者のリストを見せて調査する方法もあるが，結果は甚だ異なったものになる，いずれがよい調査法であるかは爾後の分析によってきまるのである——．こうして，名前をあげたものだけをとりあげ（名をあげるものの比率は，選挙何日前

という調査時点が一定であるならば，かなり安定しているものである），それぞれの候補者の調査での支持率はどのくらいであるかを算出するのである．これがそのまま選挙の得票率となるものではないのである．ここで得られた調査での支持率は，候補者の勢力をあらわす特性値と考えるのである．われわれとしては，個々の有権者を問題にしてはいないのである——個人は誰に投票しようとも，それはどうでもよい．候補者の支持され方がわかればよいのである．つまり，誰々に投票するといった有権者は本当にその候補者に投票するかどうか等のことは問題にしていない．この回答の真偽は確かめることはできないのである．いくら質問しても本当をいっているか否か実際に確かめることはできない（知っているのは本人だけである）．したがって，われわれがこの個人の回答の真実性を仮定したうえでの分析を問題にしないところに，われわれの方法の良さがある．よく昔いわれた調査前の調査で名をあげた集団が投票者集団に連がるか否かといった議論は不必要なのである——なお，投票・棄権だけは事実として名簿によって確かめることができるのであるから，これに関する限りは明確にわれわれの研究対象にしうるのである．ここらの考え方は，きわめて操作的な立場によるものである．われわれはデータとデータとの関係をつけることが狙いであって，一方のデータは選挙での当落（ないしは得票率．当選確率）であり，一方は操作的に得られる候補者の諸要因の様相パタンであるのである．この両者を関係づけ，そこに当選確率を見出す方法・機構をつくりあげるのである．この機構を用いて，当面の調査情報を用いて予測を行なうことになる．

もう一度まとめてみよう．選挙予測の目標としては，各人に与えた当選確率（前述の意味）が，よく実際（その確率をもつものの当選者の割合）と確率的意味でよく一致していること——これだけでは十分ではない，なぜならすべての候補者の当選確率が$\frac{\text{定員}}{\text{立候補者数}}$であるとしておけば必らず当るが，予測の立場からは無意味である，37年の参議院選挙の全国区の場合なら2分の1に近いので，すべて当落線上といっておけばよかろうが，これでは話になるまい——，さらに当選確率の分布が1と0とにあつまるようにすること，すなわち，極端にU字型の分布をするようにすること，また当選確率が政党別にみてその意味が同一のこと——A党候補者の当選確率が p_r であるのも，B党候補者の当選確率が p_r であるものも，それぞれの党派別にみても同じ確率 p_r の意味をもたねばならない．例えばA党の p_r は実は $p_{r'}$ であり，B党の p_r は実は $p_{r''}$ であり($p_r'\neq p_r'$)．この $p_{r'}$ と $p_{r''}$ が平均されて全体的には p_r となっているということはまずいということである．いま当選確率 0.80 としたとき，その当選確率をもつと判定された自民党の候補者の当選確率は0.70であり，0.80の当選確率をもつと判定された社会党候補者の当選確率が0·90であり，両党候補をあわせれば，0.80の当選確率となるというのでは党派別当選者数のよい予測は得られない——，党派別当選者数の予測の精度の高いこと——以上の条件が充されていれば，予測値の分散が小さく，高い信頼度で予測が的中することになる——，方法に安定性のあること（安定性があるように予測方法をつくりあげるのである）——過去のデータの分析がこれからの予測に対しても妥当性をもつこと——が要求されるのである．

最後に述べた安定性の検討であるが，選挙回数が多ければ，時系列的に——単にマクロ的にみるのではなく，各候補者を単位にミクロ的に味細く見て行かなくてはならない——モデルを構成して行くことが望ましいのである．しかしながら，総選挙はそう多くあるものではなく，古いデータはないし——あったにしても戦後に近いものとは環境が変りすぎている——そのようなモデルをつくることはできない．そこで，われわれとしては，各年次の選挙調査・結果の分析を通して，そこに安定性を二標本理論，三標本理論に胚胎する考え方にもとづいて検討することにしている．現状に立つ限り，これは合理的な思考であろう．

なお，以下の分析では，数量化理論を，核心となるところで随所に用いてあるが，これにつ

いては次頁の脚注* の文献を参照されたい.

なお，衆議院選挙・参議院選挙は，その性格からいって予測方法が異なるので，以下の記述において，その記載様式を変えてみた.

§2. 経　緯

選挙予測の研究を始めたのは，かなりふるく，統計数理研究所の当時の第三部が行なった都知事選挙の予測に始まる．このときは「予測は如何にして可能か，その方法論は如何にすべきか」を中心に考えて研究が進められた．この時のメンバーは水野坦兄，青山兄，石田，木村，西平，橋爪，座間，多賀，堤の諸士であり，ディスカッションを重ねて事が運ばれた．成果は「数量化と予測，選挙予想調査を素材として」(水野坦，林知己夫，青山博次郎，丸善出版，昭和28年）にあるが，ここでは，予測のほかに選挙調査に関係するいくつかの傾向というものが分析され，記述され，以後の選挙調査の基本的知識（選挙調査の知恵といった方がよい）を得ることができた．しかし，予測という点では，一回の知事選挙であるために，不十分なことは止むを得なかった．それ以後，都知事選挙をわれわれのところで調査したが，あまり発展はなかった．その他，予測の研究としては，仮釈放の研究（西村克彦——岡山大学法学部）を行ない，予測の方法論を確めて行った（西村・林，仮釈放の研究，東大出版会，昭和30年)．これらと並行して，林と朝日新聞社世論調査室との間で選挙予測の研究の地固めが行なわれ，選挙調査も随時行なわれ，データが蓄積され分析されて行なった．これらが発展してきたのが，ここに記述する論述である.

衆議院議員選挙予測・参議院議員選挙予測・知事選挙予測が主なものとして，とりあげられたが，これらの研究は，全く朝日新聞社世論調査室との緊密な協同作戦によって成り立ったものである．私の方は，予測方法論確立のために方法論を研究し，世論調査室の方は，調査分析と選挙予測の内容を享受するのである．このため，協同のディスカッションを重ね，それにもとづいて調査は世論調査室で行ない，水平的分析を行ない，この結果をもとにして予測方法論にもとづいて予測を行なうことになる．それまでの小規模のものを経て昭和30年の衆議院選挙において大々的に行なわれ始め，この繰返し過程はそれ以後今日に及んでいる．衆議院4回，参議院3回，随時知事選挙ということになる.

以下の解析のもとになっている，前二者の調査を次頁に示してみよう.

註 C. Hayashi: On the Quantification of Qualitative Data from the Mathematico-Statistical Point of View [Ann. Inst. Stat. Math. Vol. II, No. 1, 1950], On the Prediction of Phenomena from Qualitative Data and the Quantification of Qaualitative Data from the Mathematico-Statistical Point of View, [Ann. Inst. Stat. Math. Vol. III, No. 2, 1952], Multidimensional Quantification I, II, [The Proceedings of the Japan Academy, Vol. 30, No. 2, No. 3, 1954], Multidimensional Quantification—with the Applications to Analysis of Social Phenomena [Ann. Inst. Math. Vol. 5, No. 2, 1954], Sample Survey and Theory of Quantification [Bull. of I. S. I. Vol. XXXVIII, Part. IV, 1960], 数量化と応用例 I, II, III, IV, V, VI [統数研彙報, Vol. 2, No. 1, 1954, Vol. 4, No. 2, 1956, Vol. 5, No. 1, 1957, Vol. 5, No. 2, 1957, Vol. 8, No. 2, 1960, Vol. 9, No. 2], 社会的態度の測定と数量化 I（池内，水原，大塩，佐野と協同）[統数研彙報, Vol. 1, No. 2, 1954], 態度数量化の一方法 II（高倉，牧田，斎藤と協同）[統数研彙報, Vol. 6, No. 1, 1958]; 高倉，数量化による分類の問題——数量化と応用例 VII [統数研彙報, Vol. 9, No. 2, 1962]: 及び統数研養成所講義録，その他，成書では林知己夫・村山孝喜；市場調査の計画と実際（日刊工業新聞社，1964)，また解説としては，林；やさしい数量化 1～12（朝日新聞広告部刊，広告月報，1964年2月—1965年1月)，統計における数量化の問題（日本数学教育会誌，第45巻，第7，9号，1964）を参照されたい．また数量化の根本的態度を述べたものには，林：数量化と予測の根本概念（統数研彙報，第7巻，第1号，1959)，心理学における数量化の問題（討論）（心理学評論，1964, Vol. 4)，社会行動の数量化のコメント（年報　社会心理学，第5号，1964)，現象解析と数量化の問題（計量国語学，第28号，1964）を参照されたい.

衆議院議員選挙予測の調査

年　月	調査選挙区数	調査時期	備　考	
30年2月	49	序盤(14～15日前) 終盤(7～ 8日前)	約25地点—サンプル数 400 約50地点—サンプル数 800	
33年5月	117	序盤(14～15日前) 72 終盤(7～ 8日前) 117	約25地点—サンプル数 400 約50地点—サンプル数 800	各選挙区での調査地点とサンプル数
35年	117	序盤(14～15日前) 77 終盤(7～ 8日前) 117	サンプル数300—400(全県1区のところは 500) サンプル数600—800(全県1区のところは 1,000)	
38年	117	終盤(7～ 8日前) 117	約60地点—サンプル数 800 (全県1区のところは1,000)	

参議院議員選挙予測の調査

年月	全国区	地方区	調査時期	備考
31		終盤10県	地方区のみ(7～8日前)	約50～70地点—1県につきサンプル数800～1,200
34	序盤46県 終盤46県	序盤18県 終盤46県	序盤(15～16日前) 終盤(6～ 7日前)	約 75 地点—序盤1県につきサンプル数400～500 終盤1県につきサンプル数800～1,000
37	序盤46県 終盤46県	序盤46県 終盤46県	序盤(19～20日前) 終盤(6～ 7日前)	約57～60地点—序盤1県につきサンプル数600 終盤1県につきサンプル数800～1,000

この研究では，世論調査室の方々，とくに前主査の今村誠次氏，現主査の堀田邦美氏，近見賦道氏，二上信爾氏，馬場正人氏に御世話になったところが多く，その他昔から御世話になっている木村定氏，後藤恒道氏（現在大広）にも感謝をささげなければならない．またわけても，方法論的なことでも近見氏と協力して行なうことが多く，氏のアイディアに基いて行なってみたことも，この論述に一部入れてある．以上の方々の協力なくしては，この研究は成立していない．御礼を申しあげる．

また，統計数理研究所では，常に興味をもって励まして下さる所長の末綱恕一先生，選挙予測にかぎらず常に同じ釜の飯を食いつつ苦労を共にしている石田正次士，この論述を仕上げるのに非常に努力をされた佐藤敬子嬢，雨宮多賀子嬢，数量化の計算で御世話になった計算機室の駒沢勉君らに深く感謝の意を表する．

第１編　衆議院議員選挙における選挙予測の方法

頭初において述べた方法論・構想をふまえて，実際に選挙予測をいかに行なうかを，衆議院議員選挙について，述べてみよう．参議院選挙，知事選挙とは選挙の性格も異なり——例えば人々の政党支持がここで非常に強くあらわされるのである，参議院となる と これ が大分薄らぎ，知事選となるとさらに薄くなってくる——同一に論ずることはできない．概念的な方法論をそのまま現実にうつそうとするときに必らず問題とすべきところが起る．これをいかに切り抜けるかを考察するところに本物の方法論が生れてくるのである．その気持を以下の第1，2，3編の分析でくみとられたいのである．全体の見通しをよくするために，衆議院選挙の場合の分析目次を次にあげておこう．

第1部　序　　論

選挙調査の結果から，各候補者に対する調査支持率および各候補者の特性がわかる．これらとその候補者の経歴というものから，当選・落選を予測することを考えるのである．調査は前述のように，一選挙区ごとに行ない，調査は原則として投票日7日前に行なわれている．一選挙区で地点数約60，標本は800が原則で，全県一区のところでは1,000（地点数は約75）となっている．調査は，層別副次抽出法による．まず，過去の選挙結果で，自民支持率，社会支持率，民社支持率（或は保守支持率，革新支持率）という政治率によって市区町村を層別する．このとき，原則として郡の区分けをくずしていない——議員の地盤は，県会議員の関係で郡をその領域としているものが多いから，これをくずすことは，調査によって候補者の支持率を推定するに当って精度がわるくなる——．こうして，第一次の層をつくる．次に投票区を単位にして，この特性——事前にデータがとってある——によって層別する．このときは，行政区劃の区分けを外すのである．こうして，投票区単位の層別ができることになる．原則として，各層の大きさはほぼ等しくなるようにする——調査の利便さから，すなわち調査員の負担を同一にする——のであるが，交通事情その他地域特性の特殊な性格のため，層が小さくなることはある．標本の割当ては比例割当法（有権者総数に比例さす）であり，一地点の標本数は一般に12～14を目標にしてあるが，上記のような特例のときは，7程度になることもある．まず，各層から一つの第一次抽出単位たる投票区を確率比例抽出法によって抽出，次に投票区の有権者名簿から，指定された標本数だけ，等確率で最終標本たる有権者を抽出することになる．こうした方法によって抽出されたものに対して調査を行なう．まず，候補者名簿等を示さず，宙の形で誰に入れるかをきめているか，それでは誰かをきき，きめていない人には，今投票するとすれば誰に入れるか，をきく．これでも名の出ない人は，立候補者中誰が一番よいと思うか，また，ぜひ当選させたい人は誰か，とたづねる．こうして名前のあがってきたものが調査における候補者の支持票となるのである．この間に強度の差があるが，これ全体をとりあつかうのと，第一番目のきめていて名をあげたもののみをとりあつかうのとでは，殆んど大局的には変わりはないので，数を安定させる意味で上述の支持票を用いるのが有利であると考えられた．このほか，性，年令，職業，学歴，支持政党をたづねるのである．分析においては，性，年令，職業，学歴別の分布，支持政党別の分析，地域層における分析——地盤につながる——を行ない，候補者の特性分析の縁由としているのである．

なお，このように調査を行なったとき（比率推定）の誤差であるが，副次抽出による誤差はこれまでの研究によれば——この程度の標本抽出法であれば——外分散の寄与は，内分散と同じにおけば十分であることが確かめられてあり，また調査不能にもとづく偏りであるが，これも殆んど無視できる程度であることが類似の調査で知られている（例えば「日本人の国民性」

(統数研国民性調査委員会，至誠堂，昭和36年）の調査誤差の項参照）ので，この知識をもとにする．

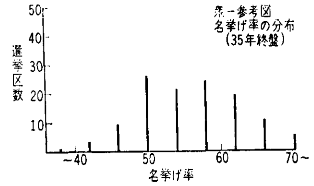
さて，こうして出た調査から，まず候補者の推定得票率を算出，さらにまた選挙区の特性（候補者乱立，激戦区等のこと）や各選挙区での調査支持率の様相を考慮に入れ，これらを総合して，各候補者の当選確率を計算することを考えるのである．もちろん，これらの分析は，定員の人区別に行なうのはいうまでもない．以下に順を追って説明を加えてみよう．分析の順序内容の大綱は下の目次によって了解されるであろう．なお，これからの解析は，一部の選挙区に対して調査した30年データを参考とし，33年，35年調査データを詳しく分析し，この結果を用い，38年の予測でこれを再検討するという形で進められている．

第2部　各候補者の得票率推定

まず，調査データから各候補者の推定得票率を計算する方法を検討してみよう．これは，予測に対する一つの要因と考えるのである．

§1.　調査支持率と選挙結果

まづ，各候補者の調査における支持率（名をあげたもの t を100とする．{各候補者の支持されている数}/t のこと）と選挙結果との関係を明らかにしてみよう．名をあげたものの比率の分布を参考のため衆一参考図にあげておくが50～65%のものが多い．さて，以下これらからの分析で，調査支持率が0.5%以下のものは，当選に全く関係がない——過去において全くなかった——ので，これを省略することにする．さて，まず第一に調査支持率はただちに選挙得票率を示すものではないことを示してみよう．



衆一参考図

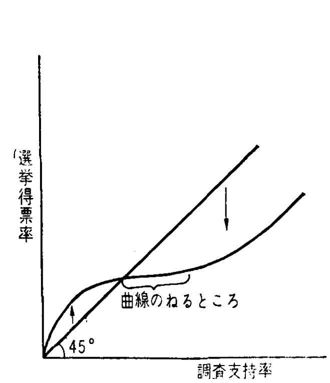
区切りは38.0～41.9，42.0～45.9のごとく，4%区切りとする．

人区別に，調査年別に，これらの関係を検討してみる．

(Ⅰ)　33年のみ
(Ⅱ)　35年のみ（終盤調査，C，D）
(Ⅲ)　35年のみ（終盤調査，B，C，D）
(Ⅳ)　33年＋35年（C，D）
(Ⅴ)　33年＋35年（B，C，D）

B，C，Dは調査時期をあらわし，Bは10日前，C，Dは7日前である．時期のズレは調査実施の都合によっておこっているものである．

調査支持率と選挙結果とを関係ずけてすぐ気のつく面白いことは，調査支持率の少ないところでは，選挙得票率が調査支持率を上回り，調査支持率の大きいところでは，選挙得票率が下回る傾向のあることがわかり，調査支持率が中位のところでは曲線がねてくる（調査支持率が多少増加しても，その割に選挙得票率は増加しない）ことである．このことから，調査支持率と選挙結果は大局的にみて，3次曲線によって関係ずけられそうなことが解る．なお，45°の直線との交点（調査支持率と選挙得票率が平均的に一致するところ）は，3人区では調査支持率22%附近，4人区では18%附近，5人区では14%附近とみられるのである．



衆一参考図

概念的に画けば，衆―参考図図のような関係がよみとれるのであるが，実際にあてはめてみた3次曲線は次頁に示す通りである（調査支持率 x，選挙得票率 y，ともに%）．

次に，こうして，あてはめられた3次曲線の安定性について考察しよう．まづ，(I)，(II)，(III) が差異の出そうなところであり，とくに (I)，(II) および (I)，(III) の問題であるが，調査支持率の特に大きいところをのぞいて――これはまれにしかおこらない――それらの間にいちぢるしい差のないことが了解される．そこで安定性を得しめるため，(IV)，(V) のいづれかの方法をとることにした．38年の予測調査ではB調査がないので，(IV) 法を採用することにした。(衆―2図，4図，6図参照)．なお，この場合，きわめて支持率の大きいところ（めったにおこらない）では，3次曲線がはね上るので，これを修正することにした。修正線は衆―1表に示してある限界点から発し，3次曲線から滑らかにそれさせ，データのあるところを通るように引いておいた．したがって，(IV) の曲線は，人区別にみるとき，衆―2, 4, 6図のようになる．修正線といっても，図にみられる通り，はげしいものではなく，限界点附近では殆んど変らないのである．

このようにして調査支持率から選挙得票率を推定する第一の仕組みとして，採用すべき3次曲線は，(IV) 法により決定されることになる．

§2.　候補者の特性に関する分析

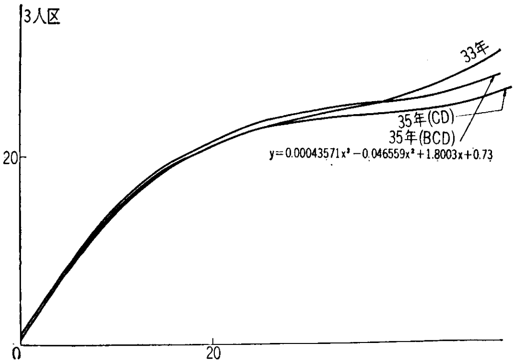
各候補者の調査支持率から§1でのべた3次曲線を用いて選挙得票率を第1次的に推定し，さらに候補者の特性を加味して，各候補者の得票率推定をつめて行くことを考える．そこで，終盤調査（C，D）による3次曲線と選挙結果とのずれ d を要因からつめることを考える．このため取りあげる要因を衆―2表のようにする．なお，このとき注意すべきことは，民社候補者を抜いてあることである．民社については別に考慮する（後述）．

衆―2表

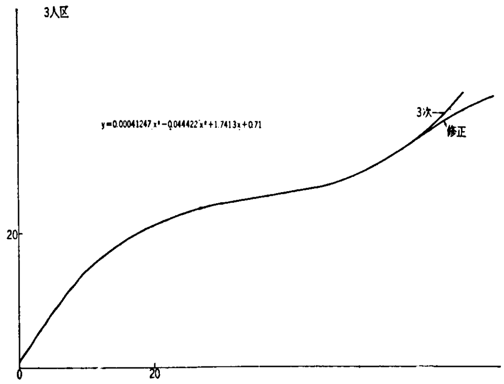
要因＼カテゴリー	カ　テ　ゴ　リ　ー	カテゴリー数
政党・議員歴・調査支持率	自元新，自前 社元新，社前 共，調査支持率5%未満，共，調査支持率5%以上 諸・無，調査支持率5%未満，諸・無，調査支持率5%以上	8
調査支持率の順位	1，2，3，4，5，6以上（3人区） 1，2，3，4，5，6，7以上（4人区） 1，2，3，4，5，6，7，8以上（5人区）	6或は 7或は 8

とくに，共，諸・無については，5%を境としてわけたのは，共，諸・無で5%以上をとるものは，選挙で強い傾向があり，また，調査支持率の低いものはむしろ弱い傾向（特に諸・無について）があると思えたからである．この傾向は d を要因別にグラフ化することによって容易にその妥当性が了解されるのである．その他いろいろの要因をとりあげて分析してみたが，行きついた結果は上記のようなものになった。ここに到るまでの検討をのべておこう．

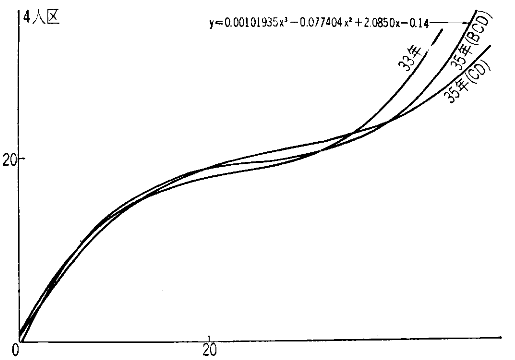
(イ) d と調査支持率，3次曲線の読み等の関係が全体的に関係するように思えるので，これらの関係を目盛ってみた（政党別，30，33年別）．すなわち，読みの特に大きいところの d が負であれば，いわゆる有名人（我々の言う調査支持率が非常に大きく出るが選挙得票率はたいして大きくないのが常である）がますますマイナスの要因であること，その他3次曲線の最小二乗法にもり切れぬものを見出そうとしてみた．この分析においては政党別にみることは非常に大切なところである．これを政党別にみるとき，自民でも d_i は3次の読みと殆んど関係な



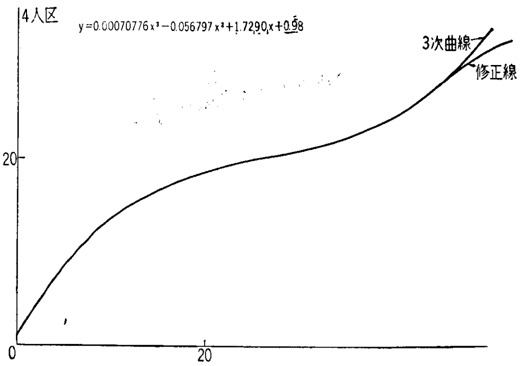
衆—1図



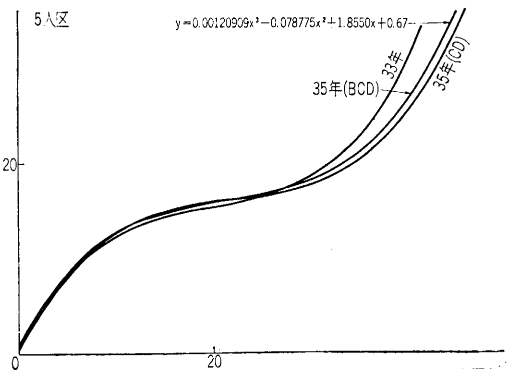
衆—2図



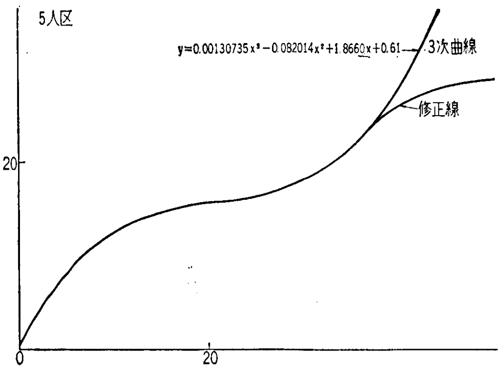
衆—3図



衆—4図



衆—5図



衆—6図

衆—1表

調査支持率	三次曲線	修正線の限界点	修正線以上にあった33年と35年の人数（支持率）
3人区	60%未満	60%以上	1人（65%）
4人区	46%未満	46%以上	2人（47.6% 49.9%）
5人区	37.5%未満	37.5%以上	4人（37.9 39.2 46.1 50.4（B調査の分））

く，社会でも同様に3次の読みに殆んど関係はなく，ただし，共，諸・無所属においては，大別するに，3人区では8%，4人区では8%，5人区でも8%(いずれも調査支持率5%程度)前後で d が異なり，それ以下では d は負であり，それ以上であると d が大きくなる傾向が見受けられる．そこで，上記の様な要因をとりあげることにした．

(ロ) d とその候補者の所属政党支持者からの支持割合（調査支持票のうち自党支持者からの支持割合）33年，3人区，5人区について検討してみたが，殆んど関係を見出せなかった．

(ハ) 調査支持票のうち支持政党なし・わからぬからの支持割合と d との関係

33年の3，4，5人区と30年の5人区について検討してみたが，このときもまた積極的な関係を見出すことはできなかった．

(ニ) 自党からの支持率の伸びと d との関係

終盤の d のつめのとき，序盤調査結果とくらべ伸びて来ている場合は強く（＋となる），減少してきているときは弱い（－となる）と考えられそうなので検討を加えてみた．つまり自党からの票がにげているか，票をまとめてきているか等を示す所の伸び加減が d に影響をあたえるのではないかとの考え方である．これについて33年3人区，5人区のデータを用い分析したが，相関係数はきわめて小さく（例えば自民党についてみると，それぞれ $\rho=0.08$，-0.23）問題にしない方がよいと考えられた．

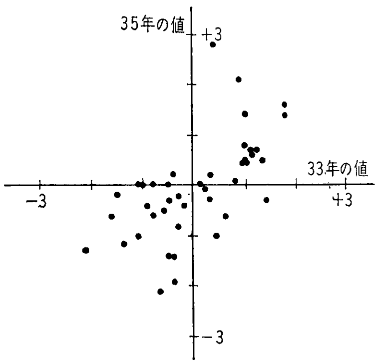
(ホ) 支持なし，わからぬものからの支持率の伸びと d との関係

これも，浮動票をあつめてきているか，浮動票がはなれてきているかのインデックスと d との関係である．33年データ，3人区，5人区について分析してみたが，このときも相関係数も低く（例えば自民党についてそれぞれ $\rho=-0.08$，-0.26），強いて取りあげる強い関係とは考えられない．

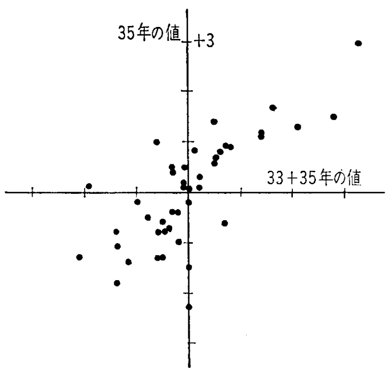
(ヘ) 調査支持率の伸びと d との関係

こんどは，まとめて，全体的にみた調査支持率の伸び d との関係をみた．このときも，33年データを用い検討したが，積極的な関係はみとめられなかった．

以上の検討で，d をつめるのに考えた要因による情報は，すでに調査支持率や3次曲線の読



衆—7図



衆—8図

衆—3表

	33　年	35　年	33＋35
3人区	0.43	0.37	0.39
4人区	0.42	0.44	0.39
5人区	0.49	0.38	0.38

みの中に含まれてしまっていると考えられるのである．たとえば伸びてきたものはすでに「伸びてきている数字そのもの」の中にその後の情報をつくしていると見做せるのである．調査支持率はすでに多くの情報を含んでいると見做せるのである．

要因カテゴリーがきまったので，自民，社会の全候補者（立候補者1人のところも含む）のデータを用い，外的基準が数量である場合の数量化によって，各要因カテゴリーを数量化してみた．なお，要因カテゴリーに与へた数値について33年，35年における安定性をみるために，数量化した結果を二次元的に目盛ってみたのが衆—7図である（点はカテゴリーに与へた数値をあらわす）．これをみると大局的には 45° の線上にのって，かなりの安定性を示していることは，興味のあることである．しかし，細かくみるとやはり，傾向が異なるものがあり，そこでひとまず33年と35年は別個に取扱ってみることにした．

なお，参考のため33年，35年のデータをあわせ総合（3次曲線は33+35年のもの）して，数量化によって要因カテゴリーの数量化をも行なってみた．この数量と前記35年データとの関係を目盛ってみた。(衆—8図)

なお，これら数量化による d との相関係数は衆—3表のようになった．

以上のようにして，調査支持率から3次曲線の読みと要因からする d の推定とを加え合せ，候補者の第2次得票率の推定が可能となる．なお，民社についての分析は後にのべることにする．

§3. 選挙区別にみた政党得票率の推定（袋の中味の推定）

次に所属政党の候補者の得た得票率を加え合わせたところの政党得票率を推定することを考えよう．この得票率は，精度よく推定できることは，30年調査，33年調査の分析で明らかになった．これを示してみよう．政党別選挙得票率と政党支持率（政党名をあげたものを100とする）との関係，及び調査での候補者支持率との関係をみたが，政党支持率による方が相関が高く（$\rho=0.90$(自)，0.89(社)），将来の利用性が高く，かつ簡単なので政党別選挙得票率の推定にはこれを用いることにした．この関係は選挙区単位にみたとき，一次式，得票率$=ax+b$，(xは政党支持率）によってあらわされることを知った．この (a，b) は自民・社会でそれぞれ異なっているが，ともに33年，35年を通じほぼ安定していることは，有力なものであると考えられる所以である．こうして，選挙区別に政党得票率がはっきりするならば，これを適切に個人に配分することを考えれば，個人の得票率をよりよく推定できると考えられる．そこで，さらに，政党得票率たる袋の中味を精度よく推定する方法を考えてみることにする．

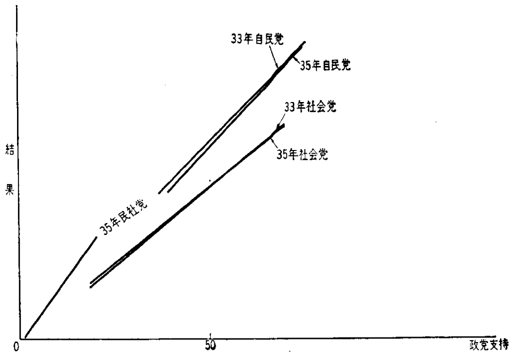
なお，ここに政党別得票率は，自系，社系の諸派無所属（7％以上のもののみ）はそれぞれ自，社の得票率として計算した．自系，社系の判定はエキスパートの判断によった．しかし，今後は候補者の支持基盤の多い政党（支持票の50%以上のものを自民あるいは社会の支持層から得ているもの）によって定めるのが適切と思われる．

さて，要因としては，上述の様にまづ政党支持率（支持政党の名をあげたものを100とした比率）が考えられる．各候補者の支持率の合計よりも政党支持率を用いる方が能率的であることは，上にものべた通りである．政党別にこれをみよう．データの完備している33年，35年別に検討してみる．衆—9図に示してある．この曲線をみると33年と35年できわめてよい一致を示していることがわかる．図式的には参考図のようになる．

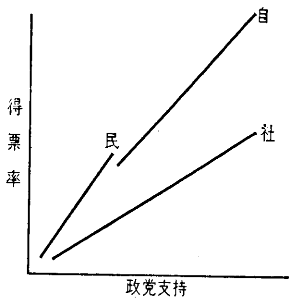
自民党は殆んど45° の線で，政党支持率と得票率とはよく一致するが，社会党は政党支持にくらべて得票率は低くなる．一方，民社は45°より，すなわち政党支持より得票率は高い．これらは，支持なし，わからぬ，というものからの得票率が影響するためではないかと予想される．

次にこの直線からのずれ D をつめることを考える．要因としては，衆一4表のようなものが考えられる．

これらをとりあげるに到った根拠を示してみよう．まづ立候補者数の関係をみよう．33年，



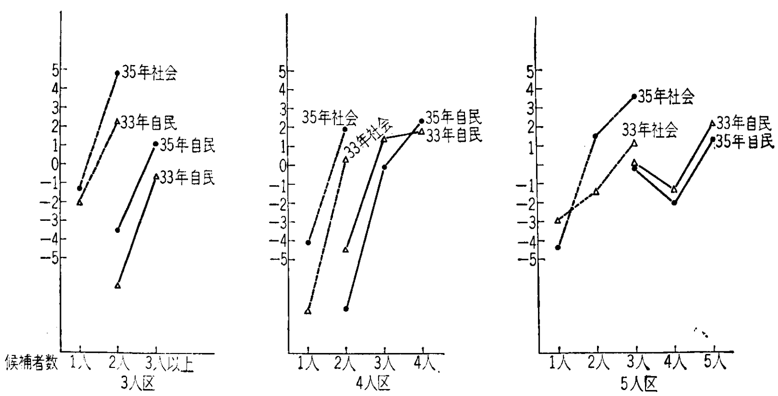
衆一9図



衆一参考図

衆一4表

	33年 カテゴリー	35年 カテゴリー
立候補者数	自 3人区 2人，3人以上 自 4人区 2，3，4以上 自 5人区 3，4，5以上 社 3人区 1人，2人以上 社 4人区 1人，2人以上 社 5人区 1人，2人，3人以上	左に同じ
支持なし，わからぬからの調査支持率	自 40以下 40.1←→60 60.1←→80 80.1←→ 社 10以下 10.1←→20 20.1←→40 40.1←→	左に同じ
民社支持率	(33年に民社党になかった)	立たぬ 5%以下 5.1←→10.0 10.1←→15.0 15.1以上



衆一10図

35年を衆—10図に図示してみたが，相当いちぢるしい関係が見られ，候補者の増加が D に及ぼす影響（得票率に及ぼす影響）をみることができる．なお，各政党の立候補者といったとき，前述の定義に従い，自民系，社会系の無所属をも含めてあることに注意されたい。

次は支持政党なし，わからぬからの支持率との関係である．33年，35年との相関図を検討したところ，かなりの関係が見受けられ，特に社会党に著しいが，当然直線的ではない．そこでさきに示したような区分を用い，数量化を考えることにした．

次は，35年度における民社の支持率との関係をみてみた．なお，これは政党支持率との関係であるので，すべての選挙区が入るが，この中には民社の候補者の立っていないところも含まれている．しかし，実効的には候補者が立つ立たぬが大事なので，要因のカテゴリーとしては「立たぬ」を第一にとりあげ，立っているところを政党支持率によって衆—4表のように区分した．社会党においてはかなりの逆相関がみられるのである．すなわち，民社支持が多いところでは，社会党の D は少ないというのは注目すべき関係である．自民党にたいしては，さしての影響はみられない．なお，この要因決定に到るまで次のような検討も加えてみた．

(イ)（前回の推定－前回の選挙得票率推定）と D との関係

前回について，我々の行なった推定より選挙得票率が多かったものは，いつもその傾向にあるのではないか（出にくい選挙区），推定より少なかったものはいつも少ないという傾向にあるのではないか（出すぎる選挙区）ということのための検討である．

ここで前回を33年として35年の D との関係をみた．積極的にとりあげるほどの要因とは考えられなかったので取止めた．

(ロ) 相手党の支持率と D との関係

自民党の D では社会党の政党支持，社会党の D では自民党の政党支持を入れてみてはどうかとの疑問がおこる．D が他党の支持率に影響されるのではないかとの心配である．検討を重ねたが積極的な関係を見出すことはできなかった．

以上の検討によって，とりあげる要因は，さきに示したようなものになった．

さて D を推定するのにとりあげた要因を用いて数量化によって要因カテゴリーの数値を決定した結末を次に示そう．33年，35年別，人区別に行う．まず見通しをよくするために相関係数をあげておく．

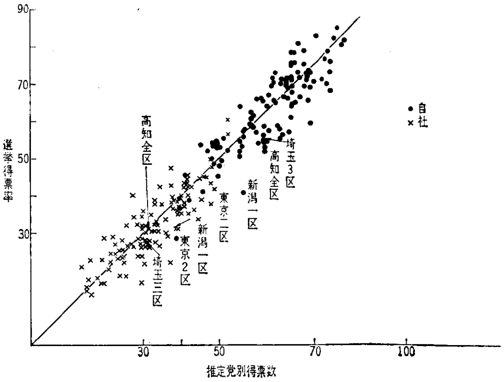
衆—5表

	33 年		35 年	
	自	社	自	社
相関係数	0.50	0.56	0.52	0.72

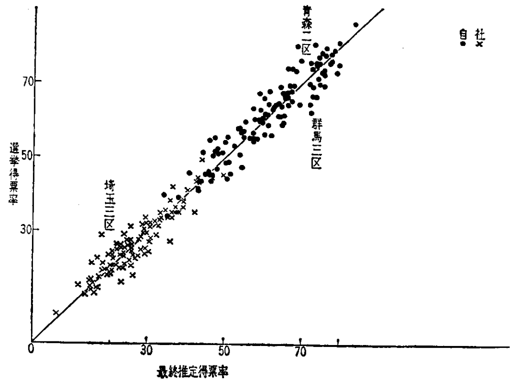
相関係数が相当高い値を示しているのは推定の有効性を示していることになる．

こうして数量化によって与えられたカテゴリーの数値をみると興味深いものがあるが，特に立候補者の数のところは非常に面白い．なかでも，社会党についてみると面白い．35年データの結果によると，平均的にみて，3人区では2人立てれば1人のときより6％増加になる．4人区では1人より2人の方が約5％増加する．5人区では1人のときより3人の方が約7％増加するといった関係がみられるのである．

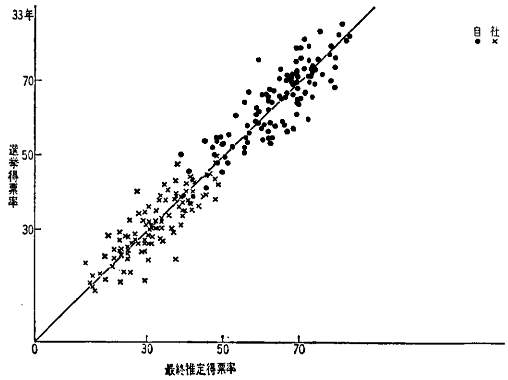
選挙区ごとに，政党支持率による推定をまず行ない，次に，要因の数値を用いて D を推定し，両者加え合わせて，政党別得票率を計算した結果次のようになった．衆—11図，衆—12図参照．35年については，埼玉2区，群馬3区，青森2区をのぞいてよい推定が得られている．



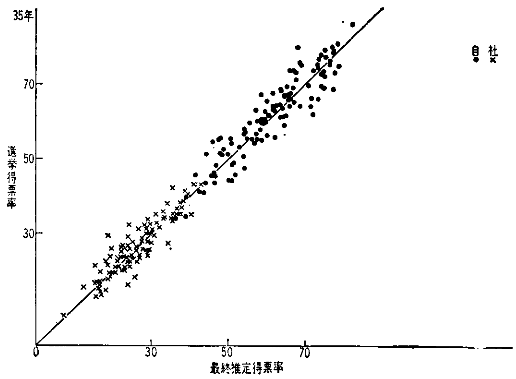
衆—11図



衆—12図



衆—13図



衆—14図

しかし，33年については少しちらばっているが，これは，自系，社系の判定が，35年にくらべて，きびしすぎたためであると思われる．

さて，もう少しつめられないかと考え，これまでのべてきた推定と結果との差 D' をとり，さらにこれをつめる事を考えた．なぜ一度にこれをしなかったかというと，相手党の推定得票率を加えたいためである——これは，一度にすることは不可能である，ここにいう相手党の推定得票率とは，いまここで行なおうという推定以前の段階までで得られたものであるからである——．

自民党の D' と社会党の推定得票率との関係，社会党についてはその D' と自民党の推定得票率の関係を検討した．35年について，傾向が見られたので，これをとりあげることにした．つまり，相手党の推定得票率の大きいところでは，D' が少な目に出ており，相手党の推定得票率の小さいところでは D' が大きく出すぎているという傾向があるのである．

この関係を利用し，上述の推定に D' の推定を加へ，最後の袋の中味の推定を出した．これと選挙得票率とをつきあわせたところ，衆—13図，衆—14図のようになった．

このようにして，自民，社会両政党の最終推定得票率，袋の中味を決定することができた．なお，民社については別角度からいろいろの検討をほどこしてみた．

民社の政党支持率と得票率(袋の中味)との関係は，前の衆—9図に示しておいた．しかし，民社党については，敢て，民社の支持率を用いずに——その他の新政党出現を予想して，何等かの既存のものから推定したいと考えた——推定することを考えてみた．

まず自民，社会の一次式の係数の算術平均をとってみて関係を求めてみたが，あまり好ましいものではなかった．したがって，民社に関しては，袋の中味を出して各候補者の推定得票率に加味する方法をとることを躊躇せざるを得なかった．共，諸無については袋の中味を考えないことにした．なお，民社・共産については一般に立候補が1人なので個人の推定得票率を出す方が有利の様に思われた．これは，新しい政党の出現にさいしても同様であると思われる．

§4. 各候補者の推定得票率

これまでに3次曲線，各候補者の特性によるつめ，袋の中味についてのべてきた．これらを総合して，個人の推定得票率を出すことを考える．

(イ) 3次曲線の読み（全候補者，ただし，調査支持率0.5％以下をのぞく）

調査支持率から3次曲線を読む，これが第1次推定値となる．

(ロ) d のつめ（民社をのぞく候補者）

各候補者の特性要因から，それらに与えられた数値を用いて推定値を算出し，3次曲線の読みを加え第2次の候補者の推定得票率を求めた．

(ハ) 袋の中味にもとづく推定

さらに，これに袋の中味を加味して，最後の推定得票率を出すことを考えるのである．前にも述べたように選挙区別の政党別推定得票率は，推定精度が高いので，個人の得票率の選挙区での積みあげが，政党別推定得票率に等しくしておくことが望ましいと考えられるのである．

A政党候補者（R 人いる）の選挙区での推定得票率を ${}^Ax_1$, ${}^Ax_2$, $\cdots {}^Ax_R$ としよう．$\sum_i^R {}^Ax_i={}^Ax$ を求める．この x がその選挙区での政党の推定得票率 ${}^AU$ と等しければ問題はないが，等しくないとき調整しなければならない．ここに調整法が問題となってくる．

自民党については ${}^{自}U'-{}^{自}x$ の ${}^{自}x_i$ による比例配分方式が適切であると考えられた．なお，この ${}^{自}U'$ に関しては注意しなければならないことがある．袋の中味としての ${}^{自}U$ には自系無所属の分が入っているのである．これをさし引いておかなければならない．自系無所属の推定には，後述するように(イ)と(ロ)とを加味した個人の推定得票率を用いるのである．こうして ${}^{自}U$ から，この推定の分をさし引いたのが ${}^{自}U'$ である．これを純自民袋の中味とする．

しかし，社会党についてはいろいろ考えられた．つまり，選挙戦術の問題があるのではないかとの配慮である．

社会党の袋の中味を社会党の候補者数で割ったものが，自民党の候補者1人当り袋の中味を下回れば比例配分，上回れば次の3法を考える．

a. 袋の中味の残りを全部，社会党候補者の推定得票率の最低のものに与える．あとは0，ただし，このとき得られた得票率が上位のものよりかへって上になれば，上のものと同じにする．すなわち，袋の中味を候補者に均分する．

b. 袋の中味の残りを候補者に等分する．

c. 比例配分

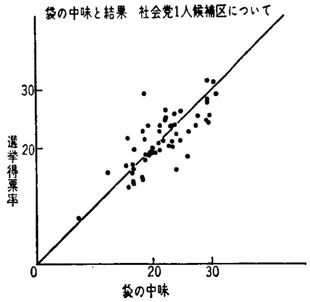
これをくわしく説明するまでもないが，うまくやれば（或は相手方が下手であれば）候補者全員を当選させることができる可能性がある場合には，うまく票を分けようとする努力があらわれるかも知れないとの配慮である．

35年について検討を重ねてみたが全体的には著変はないが，結局は比例配分が穏当であると考えられた．

なお，社会党の1人候補者については，袋の中味と個人の推定得票率と一致すべきものであるから，中味からせめるか，個人からせめるかの問題点がある．

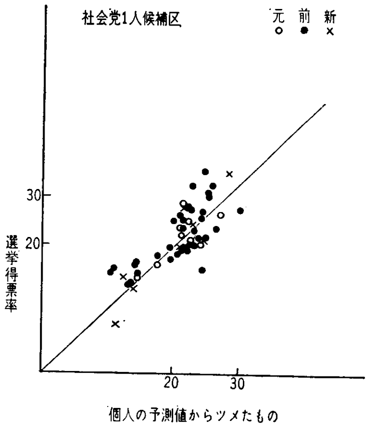
袋の中味と選挙得票率との関係をみると衆—15図のようになる．（35年）

社会党1人候補区



衆—15図

第2次推定



衆—16図

衆—6表

要因	カテゴリー
議員歴	元，前，新
社会党政党支持(%)	15以下，15.1—20，20.1—25，25.1以上
3次の読み(%)	16以下，16.1—20，20.1以上

一方，個人の得票率を推定する方からやってみた結果を示すと衆—16図のようになった．このときの推定方法は d を次のような要因でつめた．

つめる方法は，もちろん外的基準が数量である場合の数量化によった．こうして衆—15図，衆—16図を検討した結果，大差はないが，比較的得票率の多いところで，個人による方法がバラツキかつ袋の中味による方法の方の誤差が一様なところから，むしろ優れていると見て——さらに推定の繁雑さをも考慮に入れて——袋の中味による方法を採用することにした．つまり1人候補者であるからといって特別扱いはしないことになった（1人であるから配分の問題はなく，すべて袋の中味はその候補のものとなる）．

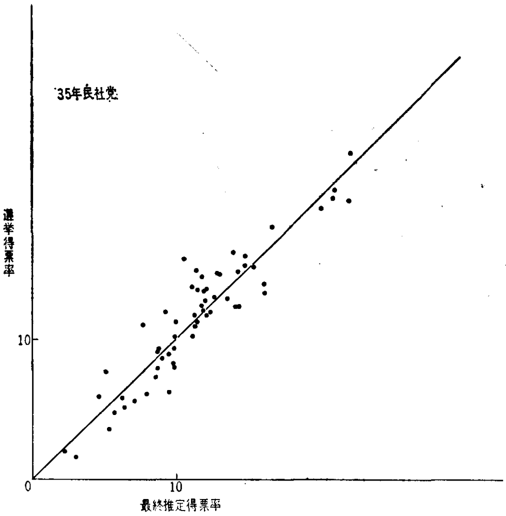
民社については袋の中味の方法がうまくないので，袋の中味の方法は用いなかった．

(ニ) 民社党について

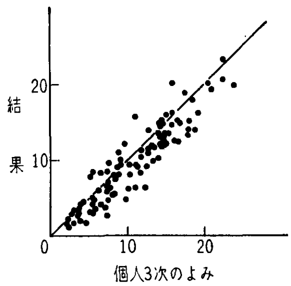
民社については，これまで特別扱いしてきて，d のつめの時にも除外してあった．そこで，ここでは民社のつめをのべよう．前にもいったように袋の中味を考えず，専ら個人の得票率を問題にすることにする．

衆—7表

	カテゴリー				
議員歴	新，前，元				
民社支持率(%)	5%以下，5.1—10.0，10.1—15.0，15.1%以上				
調査得票率		a	b	c	d
	3人区	6%以下	6.1—12.0	12.1—20.0	20.1%以上
	4人区	4%以下	4.1—8.0	8.1—12.0	12.1%以上
	5人区	4%以下	4.1—7.0	7.1—11.0	11.1%以上



衆—17図



衆—18図

まず全体の3次曲線を出し *d* をよむ．*d* をつめる要因を検討した（前述の(ロ)にはぬけている）．要因には，次のものをとりあげた．

これをとりあげた根拠は，図表による検討によって得られたものである．これらを数量化によって数量化し，その値を用いて35年を推定した結果は衆—17図に示すようになっている．

なお，3次の読みだけで35年を推定した結果は衆—18図に示すが，たしかに系統的に下に出ている傾向がみられる（約2.5％下まわる）．

なお，つめるのに他要因を問題にしてみた．これを示しておこう．支持政党なし・わからぬからの支持率と *d* との関係をみると，関係ははっきりしていない．なお，支持政党なし・わからぬと民社の歩留り率（支持票のうち民社支持層からの支持割合）との関係をみてみると——民社の歩留りの高い人は支持政党なし・わからぬからの支持も多いであろう——かなり，この予想は肯定されるが，*d* のつめにこの関係を用いるほどは有効ではない．

次に強固な支持率（きめていて投票するもの，及び今投票すれば誰にするかで名の出た数の比率）と *d* との関係をみたが，積極的な関係は見出せなかった．また，*d* と社会党支持率との関係を調べてみたが，関係は見出せなかった．以上によって民社の候補者の得票率は，前記の要因による推定方法をとることに決定した．

(ホ)　共産党について

共産党については，袋の中味を加味せず，(イ)及び(ロ)の方法のみによった．

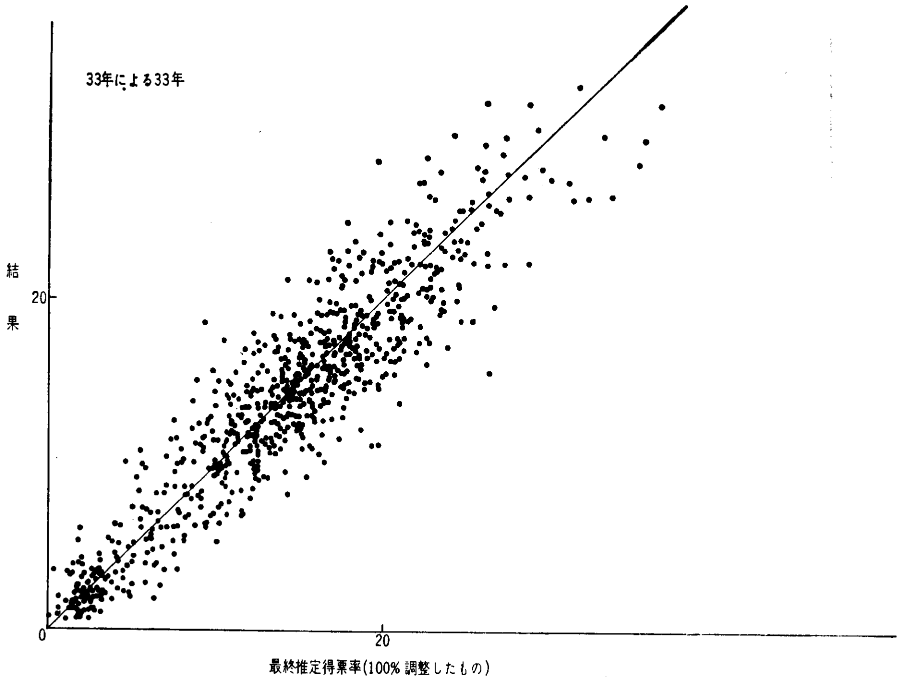
(ヘ)　諸派・無所属について

(イ)及び(ロ)の方法のみによった．

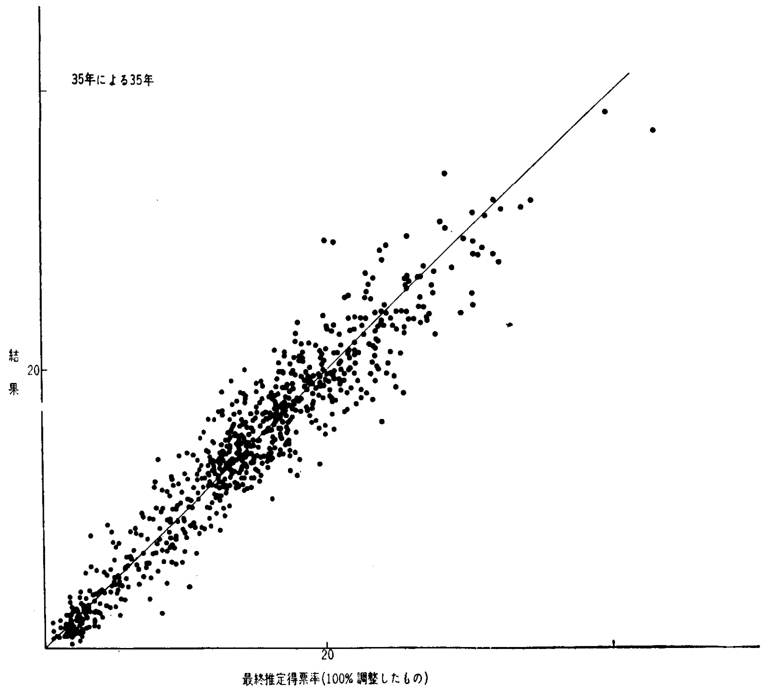
以上によって個人の得票率が最終的に決定されたわけである．これをまとめてみよう．

なお，新政党出現にさいしては，3次の読みだけで推定するのが手堅いと思われる．

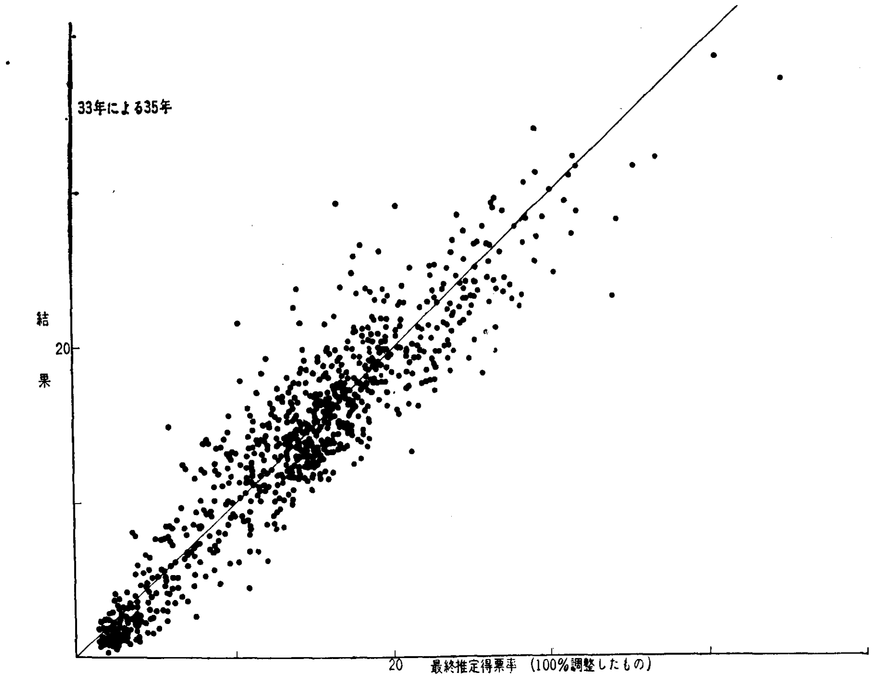
こうして推定された推定得票率——最終推定値として出された個人の推定得票率を各選挙区で100％調整したもの——と選挙結果との関係をみよう．衆—19図，衆—20図，衆—21図に示した．それぞれ33年による33年，35年による35年の平面的記述による検討及び33年データにもとづく35年の予測で，いずれもほぼ45°の直線上にあり，全体的には，かなりよい一致が示されているとみられよう．なお，33年による35年のときは民社党については3次曲線の読み


33年による33年
結果
20
0
20
最終推定得票率(100% 調整したもの)

衆—19図


35年による35年
結果
20
20
最終推定得票率(100% 調整したもの)

衆—20図



衆—21図

衆—8表

	方　　　法
自　民	(イ), (ロ), (ハ), 純自民, 袋の中味の比例配分により決定
社　会	(イ), (ロ), (ハ), 袋の中味の比例配分*
民　社	(イ), (ニ)要因による個人のつめ** (或は(イ)のみ)
共　産	(イ), (ロ)
諸・無	(イ), (ロ)

* 社会党についても社系が無所属を引いた純社会党袋の中味の比例配分を考える. これまでの分析では33年に1人あったのみである.

** (ロ)による方法にはぬけて居り, 民社のみ別個の(ニ)を用いてあることになる.

のみによる推定を用いた.

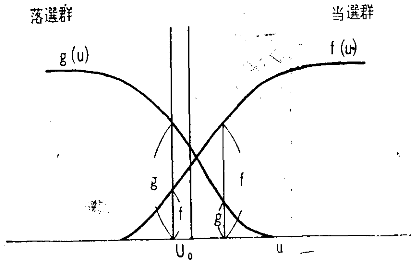
また (33+35) 年による 38 年の予測においては (33+35) 年に基く 35 年の分析結果を用いて行なうことにした. このとき 3 次曲線は (33+35) 年, d は数値を安定さす意味で (33+35) 年のデータをあわせ数量化によって計算したものを用い, 袋の中味は 35 年データ によるものを用いた. また, この民社については, 35年に基くものを用いてある.

第3部　各候補者の当選確率

§1. 当選確率推定の方法

第2部によって, 各候補者の推定得票率を出すことができた. したがって, これから, 推定誤差分布を求め, この分布を利用して, 各選挙区ごとに各候補者の当選確率（§1で定義した）

を計算することが考えられる．この方法については，参議院議員選挙の予測方法（地方区）について詳述してあるので参照されたい（後述第2編第4部．§5参照）．しかし，そこでも，少し触れたように，定員の数が多くなると計算はきわめて複雑になり実効はなく，そうだからといって全国区式に推定するのでは立候補者数がそう多くないので妥当性が欠ける．そこで，ここでは立場をかえて，当選確率を算出する方法を考えることにする．



衆一参考図

過去のデータによれば，各候補者は選挙結果の当落はわかっている．これを外的基準とするのである．調査によって各候補者はいろいろの要因をもっている．この要因はカテゴリーに分解される．すなわち，各候補者は要因パタンをもっている．この要因パタンを数量化することを考える．つまり，外的基準が分類（この場合の分類は当落の2分類である）である場合の数量化法を用い要因パタンを数量化するのである．こうして，当選群，落選群にわけて，各候補者のもつ数量（数量化法によって与えられる）の分布を考えるのである．この分布の函数（密度函数の形で書いておく）が上のようであるとする．

算出されたパタンの総合数量を u としておく．当選，落選の予測には，U_0 を求め，U_0 以上の値（数量化された要因パタンにもとづく候補者の数値）をもつものは当選，U_0 以下の値であれば落選と予測する方が考えられる．こうしたときの予測の的中率が最大になるような U_0 の求め方も数量化理論によって求められる．これは，たしかにのぞましい方法であるが，しかし，この方法には次のような不十分な点がある．

すなわち党別に u の値がかたよるとき，すなわち例えば甲党のものが U_0 の左側のすぐ近い所にあつまれば，U_0 以下のため落選と判定されて，当選数予測のとき下まわる可能性が高い．ということである．

そこで u をもつもの（u と $u+du$ の間の u をもつものといえば正確）をみると平均的にはそれが当選群においては，$N_1f(u)du$ 人あり，落選群では $N_2g(u)du$ 人だけある筈である．そうすると u の値を示すものが当選する割合，すなわち当選確率と表現さるべきものは，

$$p(u)=\frac{N_1f(u)}{N_1f(u)+N_2g(u)}$$

となる．ここに当選者の数を N_1(定員の数)，落選者の数を N_2 とする．

たとえば(33+35)年データの分析による38年の予測などのときは，(33+35)年データについて $f(u)$, $g(u)$ をきめておき，38年における N_1, N_2 を用い $p(u)$ を求めて，38年調査から算出される u にもとづく38年の当選確率の予測をつくることになる．なお，u の大切なところで，$f(u)$, $g(u)$ の形の安定性が，或は $N_1f(u)$, $N_2g(u)$ の値の安定性が要求されるのである．また，当然のことながら，これは人区別に行なうのである．

なお，衆議院選挙では N_1 は定員で一定であり，N_2 も33年，35年では安定した（調査支持率0.5％以下のものをのぞく）ものであるが，38年では減少しているのが注目されるが，とくに3人区において著しいが他についてはそれほどではない．

また，密度函数 $f(u)$，$g(u)$ の求め方であるが，候補者の数が十分でないときは，u の値をうまくグルーピングして，また当落のキイポイントがぼけないように，その形を出す工夫は必要である．

以上のような思想から当選確率を出すのであるが，要因をいろいろ検討してみる必要がある．これを検討した結果，あるときはよく利くが，他のときには利かぬと言う安定性のないも

衆一9表

		33年	N_2/N_1	35年	N_2/N_1	38年	N_2/N_1
3人区	当選 N_1	120	1.01	120	1.01	120	0.825
	落選 N_2	121		121		90	
4人区	当選 N_1	156	0.72	156	0.725	156	0.69
	落選 N_2	112		113		107	
5人区	当選 N_1	190	0.68	190	0.66	190	0.58
	落選 N_2	130		125		111	

のは除外するようにし，単純・明快なものをとりあげるようにした.

§ 2.　当選確率算出のための要因の決定

要因としては，選挙区の特性を代表すると思われる支持率パタン，候補者の最終推定得票率の順位をとりあげ，あと補足的に，経歴（前新元）と政党を加味することにした.

しかし，これをきめるまでいろいろ利きそうな要因を検討したので，それをまとめてみることは意義のあることと思われるので，それを最初に述べてみよう.

(イ)　検討された要因

(i)　地　盤

(a)　30年の選挙結果と33年の調査の結果とを選挙区別，地域区分別，候補者別につきあわせてみた．これは，30年のときの結果による地盤が33年のときくずれているか，くずれているとすれば弱いのではないかをみるためである．〔30年選挙—33年調査〕を用いたのは〔33年選挙—35年調査〕の関係を利用し予測に役立てようとしたためである．しかし，弱くなった（得票の減少した）ものはどの地域でも弱くなり，強くなった（得票の増加した）ものはどの地域でも強くなっている様相がしめされ，地盤での弱化のみが特にあらわれるという形勢は見出せなかった．すなわちこの形勢をいくつかの型に分類してみたが，B, C をより以上にわけるものを見出すことはできなかった．ここに示されている情報は殆んど調査支持率に包含されていることがわかった.

実際 ＼ 調査支持率によるもの	当	落
当	A	B
落	C	D

注　調査支持率による当落はそれによる数字の順位と定員との関係で決まったものである.
A, Dが的中，B, Cが不的中となる.

(b)　また，33年調査の地域別得票の様相つまり1地域のみで多くとるか，こまめに各地域でかせぐかなどの型を考えてみたが，やはりB, C をより以上にわけるものを見出せなかった．ここの分析では，型を中心に考えた.

(c)　33年調査結果をみて何%以上の地域が何個所あるかをいろいろ調べた（(b)の一表現）が有力なものは見出せなかった.

ここで分析したかぎりでは，地盤についての情報は全体的にみた調査支持率以上のものを見出せなかったといえる.

(ii)　支持基盤

(a) ある候補者が何党から支持されているか，を要因とすることを考えた．自・社別に，自，社，支持なし・わからぬ，それぞれから如何に支持されているかを見るために三角グラフに目盛り検討したが，やはり前述のB，Cをより以上にわけるものを見出せなかった．

(b) 職業別，すなわち，農林，商工，給料・労働からの支持を党派別に考え，やはり三角グラフに目盛ってみたが(a)と同様に有力なものを見出せなかった．

(c) 年令別＝20代，30代，40代以上にわけ，これからの支持を党派別に三角グラフに目盛ってみたが有力な要因とならなかった．

ここでの分析の目標は例えば社会党候補でも，農林・商工からの支持のあるものは強いであろう，また高年令からの支持のあるものは強いであろう，と考えたことにあったが，この情報はすでに調査支持比率の中に含まれていることがわかった（この比率の大なものは，農林・商工からの支持もあり，高年令からの支持もあることであった）．このようにして，支持基盤も強い要因たり得なかった．

(ロ) 支持率パタン

当選確率の推定で，個人の要因ばかりでなく，相手の様子をも考えて行なうことが必要であると考えられた．このため選挙区の特性を考慮に入れる必要がある．

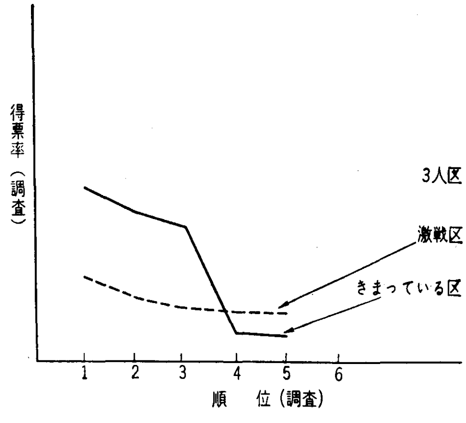
例えば，激戦区，きまりきった区などのことを考慮することを考えた．このため次のようなグラフをつくり，その型をもって，これを表現しようとした．

しかし，これを行なってみたが，なかなか，よい型を見出せないので次のような表現をもってこの型をあらわすことにした．

これには，有意差の考えを用いるのである．例えば，1位のものは2位，3位，4位とくらべ有意差があるか否か（あることを＋，ないことを－と表現する）をくらべ，＋＋＋，＋－＋，－－－などの表示をあたえる．＋＋＋は強いこと，－－－は弱いことが表現される．2位のものは1，3，4位とくらべ，－＋＋，＋－－などと表わす．－＋＋は強く，＋－－はよわい2位となる．このようなことがそれぞれ表現されれば，

激選区は　1位　－－－
　　　　　2位　－－－
　　　　　3位　－－－
　　　　　4位　－－－

などと表現されることになる．



衆一参考図

この考えにもとづき，さらに一般化して，得票率パタンなるものを考えることにする．つまり，得票率の有意差パタンである．得票率といったが，ここでは調査支持率を用いるので，調査支持率パタンといった方が適切であるが，気持としては得票率パタンである．支持率を用いるのは，予測の現場において迅速性を要求されたためである．なお，パタン作成において政党議員歴を加えて検討してみたが，不安定なものが入りこむ虞れがあるので，これを除外することにした．ただし，支持率パタンをつくるとき，単純に有意性パタンのみではなく，調査支持率の％を情報として用い△％以下，▽％以上というような分類をのこしておいたところもある．これは，調査にもとづくものなので，のこしておいた．これをまとめたものを衆—10表にあげておく．

なお，ここで有意差の＋，－をつける方法について述べておこう．

有意差は一応信頼度 95 ％を一応の目安としてとることになる．x_1，x_2 を２人の候補者の調査支持率とする．このとき x_1-x_2 の分散として $\sigma_{x^2_1}+\sigma_{x^2_2}-2\rho_{x_1x_2}\sigma_{x_1}\sigma_{x_2}$ を考えるのであるが，

$\dfrac{x_1+x_2}{2}=x$ を用いれば，$\dfrac{2x(1-x)}{n}+\dfrac{2x(1-x)}{n}+4\dfrac{x^2}{n}=\dfrac{4x(1-x)+4x^2}{n}$ となる．

このとき，内分散 $\dfrac{x(1-x)}{n}$ と外分散 $\sigma_b{}^2/m$（$\sigma_b{}^2$ は投票区内の外分散，m は調査地点数）と同一に考えてある．このことは，後述する参議院予測の場合の標本誤差の検討の所でも解るように相当大きめの見積りである．さらに衆議院の場合では，選挙区内の層別が地域を加味（地盤を加味）しているので相当よくなっているので，外分散はこれより十分小さくなっていると思われるし，また，n が選挙区でことなっているなど考え合わせれば，さらにまた有意差で深刻な意味をつけるのではなく，型の作成に資するためのものであることを勘案して，n の平均的の値 $\bar{n}$ を用い，検定は $4\sqrt{\dfrac{x}{\bar{n}}}$ によった．これで一般には大過はなかろうと考えられた．$\bar{n}$ としては 600 をとった．この $\bar{n}$ は過大であったが，過去のデータを修正できないので，その情報を保存する意味で $\bar{n}$ を用いてある．$\bar{n}$ はいまのところ実際は 383 でほぼ 400 見当である．しかし，信頼度95％にこだわることはさして重要でなく，また外分散の見積り等のことを考えてみると，$\bar{n}=600$ としたものを用いても型の作成の本質をそこなうものではない（$\bar{n}$ を 600 から 400 にするとき，95％信頼度を固執すれば外分散の見積りが大体，内分散の 0.4 倍程度ということになるとみなしておけばよいのである．この外分散の見積りもそう無理ではないと判断されよう）．

このパタンで一般にAは強く，E，Fは弱いと考えられるのである．なお，後述する数量化にしたがって要因を数量化するとき，パタンをまとめる必要もおこってくる．このまとめ方は相関表を調べ，数があまりにも不揃いとなり，安定性を欠くことになるとまずいので，この観点からまとめてみたのである．

なお，ここに調査支持率で同順のために，パタンが二つにまたがる場合が出てくる．この時は，二つのタイプに 0.5 づつにわけてかぞえることにした．つまり，0.5 人とわかれたものが二つのタイプにそれぞれ属するとして計算を行なうことにした．この立場から，これから取扱うので分析相関表は必らずしも整数となっていない．なお，あとの取扱いについては別にのべる．

(ハ) 調査支持率の順位

調査支持率は予測に用いることは適切でない．人区をそろえたところで，立候補者の数によって，その「力」が異なるからである．３人区の推定得票率20％で強いこともあるし，弱いこともある（例，候補者の推定得票率が 50％，20％，20％，5％，5％ の 20 ％は強く当選しそう

衆—10表—1

3人区

	1	1	1	2	2	2	2	3	3	4	4	4	4	4	5	5	5	5	6～
1				−	+														
2																			
3	+	−	−	+	+	−	−			−	−	+	+	+	−	+	+	+	
4		+	−			+	−	+	−							+	−	−	
5										+	−	+	−	−					
6～													+	−			+	−	
パタン	A_1	B*	C	A_2	B*	B*	C	B*	C	D**	D**	D**	D**	E	D**	E	D**	E	F

*B を調査支持率によって B_0, B_1 にわける．B_0……22%以上，残りは B_1

**D を調査支持率によって D_1, D_2 にわける．D_1……8%以上，その他 D_2

衆—10表—2

4人区

	1	2	2	2	3,4	3,4	5	5	6	6	7～
1											
2											
3		+	−	−							
4							−	+	−	+	
5			+	−	+	−					
パタン	A_1	A_2	B*	C	B*	C	D	E**	D	E**	F

* B_1 については，調査支持率22%以上と B_0，それ以下を B_1 にわける．

** E については調査支持率6%以上を E_1，それ以下を E_2 とする．

衆—10表—3

5人区

	1	2	2	2	3,4,5	3,4,5	6,7	6,7	8	9～
1										
2										
3		+	−	−						
4										
5							−	+		
6			+	−	+	−				
パタン	A_1	A_2	B*	C	B*	C	D	E**	E**	F

* 調査支持率18%以上を B_0，それ以下を B_1 とする．

** 調査支持率4%以上を E_1，それ以下を E_2 とする．

衆—11表

	パタンのまとめ
3人区	A_1，A_2B_0，B_1，C，D_1，D_2E，F
4人区	$A_1A_2B_0$，B_1，C，D，E_1，E_2F
5人区	A_1，A_2B_0，B_1，C，D，E_1，E_2F

なのであるが，32%，30%，20%，18% の 20%は安心ではないのである)．そこで個人の推定得票率による順位（100% 調整しても，しなくても同様）を要因とした．これは(ロ)パタンの具合のわるいところを補正する意味が十分あると考えられるからである．ともに単独に用いるよりも相補なう面があるからである．要因としての順位のまとめは次の通りである．

衆—12表

	カテゴリー
3人区	(1)，(2)，(3)，(4)，(5)，(6以上)
4人区	(1)，(2)，(3，4)，(5)，(6)，(7以上)
5人区	(1)，(2，3)，(4)，(5)，(6)，(7)，(8)，(9以上)

順位の場合も同順である場合は，(ロ)パタンの時と同様なあつかいとすることにした．すなわち，甲，乙両候補者が1位と2位が同一のときは，甲は1位として0.5，2位として0.5の分身をもつ，乙も1位として0.5，2位として0.5の分身をもつものとして，すなわち1人でなく0.5と0.5にわかれた別個のものとして集計すべきことになるのである．したがって，パタンも順位も全く同じときは1人が0.25人づつにわかれるものとして計算するこにした．

(ニ) 議員歴，政党

以上のほかに参考のための要因として，この二つを考えておく．これは，かなり利く要因なのであるが不安定さが心配になるものである．全人区共通で，元新，前の二グループと政党は自，社，民，共，諸・無（33年では民社をぬく）である．

以上の考察によって，とりあげるべき要因とそのカテゴリーがきまった．そこで，§1 の考えにもとづいて当選確率の算出を行なってみよう．

なお，ここで要因として，政治記者の判定を加えることも検討の余地は十分あるが，判定の基準が同一でないことからみて，今後の研究問題となろう．

§3. 数量化による当選確率の算出

(ロ)—(ニ)までの諸要因を総合して当選確率を出すのであるが，要因として（パタンと順位）を主とし，不安定な要素たり得る議員歴と政党は参考にすることにした．

方法は，前にものべた外的基準が分類（$T=2$）の場合の数量化によるのである．$T=2$ では当選・落選を意味する．

まず33年のデータによる33年の平面的分析，次に35年のデータによる35年の平面的分析，次に（33+35）年による35年の平面的な分析を行なってみる．

こうした結果を35年の予測，38年の予測に用いてその的中率を検討してみることにする．

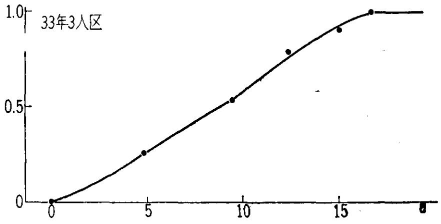
(イ) 33年データによる33年の平面的分析

この平面的分析にもとづいて，その適切性を検討するのである．これを，35年に適用したらどうなるかによって，そのモデルの妥当性が験証されることになる．

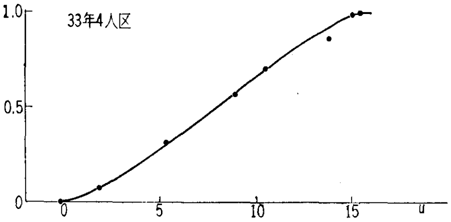
まず，要因とパタンの関係を示す相関比 η^2（予測の的中率と深い関係にある）の値と要因パ

衆—13表　η^2 の値

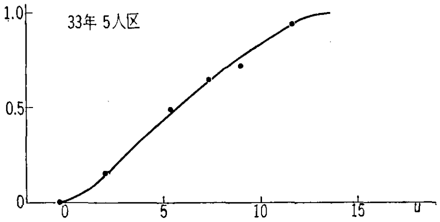
	パタンのみ	順位のみ	パタン+順位	元，新，前を加える	政党を加えた全要因
3人区	0.6318	0.5535	0.6501	0.6621	0.6666
4人区	0.5421	0.4847	0.5575	0.5693	0.5702
5人区	0.5071	0.4995	0.5495	0.5535	0.5635



衆—22—1 図



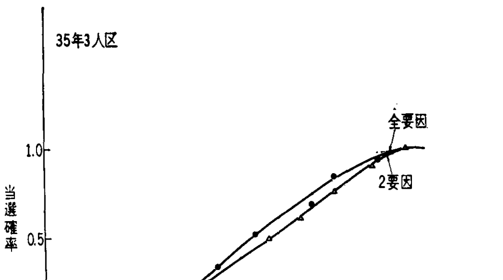
衆—22—2 図



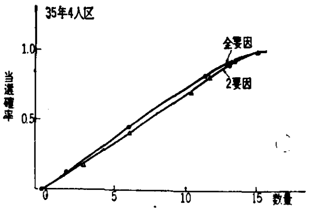
衆—22—3 図

衆—14表 η^2 の値

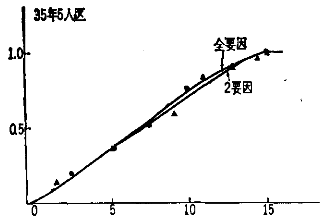
	パタンのみ	順位のみ	パタン+順位	元，新，前を加える	政党を加えた全要因
3人区	0.5629	0.6042	0.6201	0.6225	0.6387
4人区	0.5064	0.5724	0.5846	0.5891	0.6113
5人区	0.5376	0.6003	0.6195	0.6247	0.6372



衆—23—1 図



衆—23—2 図



衆—23—3 図

タンの増加との関係をみよう．

かなりの精度が期待できることがわかる．相関比でみる限り，1要因でもかなり行けるが，個別データについて観察するとき1要因だけによると，妙に誤って，当落を推定する不備があるが，これを他の要因で補なうことによって異様なものがなくなり，妥当性が高まることは見逃せないので，2要因は用いるべきである．なお，議員歴及び政党は，個別データにみるとき，特に民社党において利くのであるが，不安定ではあるし，全体的にみたとは，あまり利いているように思われないので，参考程度に用いることにした．

さて，こうして数量化された u と当選確率との関係をみよう．（パタン＋順位）の2要因のものについて示そう．図にある点は，データにもとづく $p(u)$ である．線は，平滑したものである．u とともに $p(u)$ の増加の様相のいろいろな姿が見られよう．

(ロ)　35年データによる35年の平面的分析

33年の場合と同様に35年データを分析してみよう．35年要因による35年の当落結果を用いての分析である．まず要因と当落との相関比をみよう．

このときも，33年によるデータの分析と全く同様の考察を得ることができる．ここで，2要因，4要因のものについてグラフ化してみよう．2要因，4要因比較のため，2要因の場合の要因パタン数量のレインジと4要因の場合の要因レインジをそろえ，また，最小の点を一致させて，グラフ化してみたのが衆—23図である．u の同一の値でも二者の意味が異なることには注意する必要がある．

(ハ)　(33＋35)年データによる35年の平面的分析

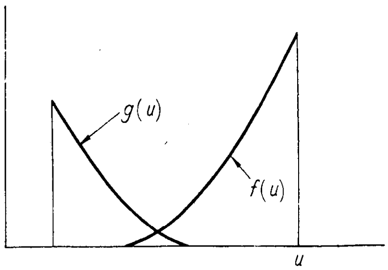
これは，予備段階で，33年と35年のデータをあわせて分析を行ない，一つの安定したものを見出そうとする試みである．

すなわち，(33＋35)年のデータによって，分析するに際し，前述のように，(33＋35)年の3次，d も (33＋35) 年によるものを用い，袋の中味は35年のものを用い，これらによって推定した要因を35年データのみについて操作した．33年を用いなかったのは，民社の問題が入ったためである。このようにして，支持率パタン及び推定順位を出し，これを用いて数量化を行った．まづ，$f(u)$ と $g(u)$ の関係を衆—24図に示した．これから求められる u と $p(u)$ との関係は衆—25図のようになる．

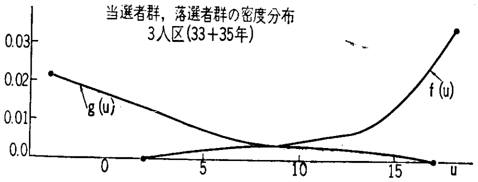
こうして求めた関係を用い予測を行なうことになる．推定順位，支持率パタンをさきに示したようにして求め，$f(u)$, $g(u)$, N_1, N_2 の情報を用い，u と $p(u)$ との関係を求め（一般に時期によって N_1, N_2 が異れば過去の u と $p(u)$ の関係を用いることは必ずしも妥当ではないことも考えられる．このときは，$f(u)$, $g(u)$, 新しい N_1, N_2 を用いて u と $p(u)$ との関係を作り直さねばならない．要は，u の大事なところで $f(u)$, $g(u)$ の意味で安定性があるか，$N_1f(u)$, $N_2g(u)$ の意味で安定性があるかにかかっている）．これから当選確率を算出する．ここ

衆—15表　η^2 の値

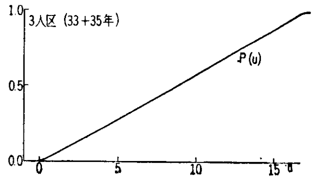
	パタン＋順位
3 人 区	0.58960
4 人 区	0.55957
5 人 区	0.59650



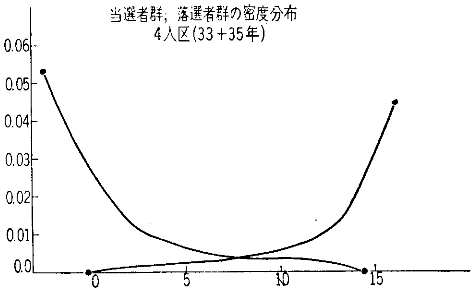
衆—参考図



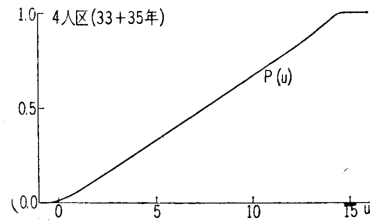
衆—24—1 図



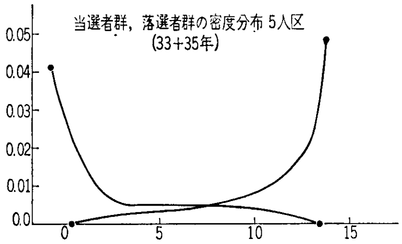
衆—25— 1図



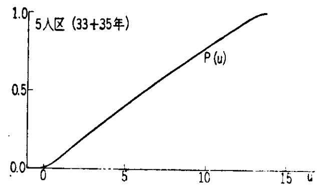
衆—24—2 図



衆—25—2 図



衆—24—3 図



衆—25—3 図

で一寸注意すべき点は，まづ，パタンと順位の同順のものである．このときは，両人は，二つのパタンの数値の平均，或は順位の数値の平均を与えて確率を出すのである．これは，近似的には（パタン＋推定順位）による当選確率表における確率——それぞれのパタン，或はそれぞれの順位としたときの確率を出し，それらの算術平均を用いる——の平均によって当選確率を与えることにする．

また，当選確率の算出において，その出し放しにした場合と選挙区別につみあげて定員数に等しくする調整法の二方法がある．

さて，ここで調整するということは，当選確率がきまり，これによって各選挙区ごとに各候補者の当選の平均値が求められる——確率0と確率1とは対象外とする——ことに呼応する．

100%調整における考え方は，当選確率にもとづくものである．今 m 人の候補者の当選確率を $p_1, p_2 \cdots p_m$ とする．このうち R 人のみが当選すべきものとする．$x_1, x_2 \cdots x_m$ を1か0をとる確率変数とし，1は当選，0は落選とすると，1をとる確率はそれぞれ $p_1, p_2 \cdots p_m$, 0をとる確率は $q_1, q_2, \cdots q_m$ $(q_i=1-p_i,\ i=1, \cdots R)$ となる．

$\sum_{i=1}^{m} x_i=R$ という条件の下での 1, 2, $\cdots$ m 人の候補者の当選する確率を計算すればよい．すなわち，R 人が当選するという事実を知った上で各候補者の当選する確率を出せばよいのである．そこで $P_r\left\{\sum_{i=1}^{m} x_i=R\right\}$ の確率を計算し，i の当選する確率（当選確率，$p_i(m, R)$）を次のよ

うに計算すればよい．$\sum_{i=1}^{m} x_i=R$ のとき当選者となることを I_R で示せば，

$$p_i(m,\ R)=P_r\{i \in I_R\} \Big/ P_r\left\{\sum_{i=1}^{m} x_i=R\right\}$$

として与えられる．なお，$\sum_{i=1}^{m} x_i=R$ となる場合の数は $\binom{m}{R}$ だけある．

例として，$m=4$，$R=2$ とする．

$$\binom{m}{R}=\binom{4}{2}=6$$

候補者 A, B, C, D の当選確率 p_A，p_B，p_C，p_D をそれぞれ 0.8, 0.7, 0.6, 0.4 としよう．

	A	B	C	D
I	○	○	×	×
II	○	×	○	×
III	○	×	×	○
IV	×	○	○	×
V	×	○	×	○
VI	×	×	○	○

○印当選
×印落選

の場合が $\sum_{i=1}^{4} x_i=2$ の場合である．

I の確率はそれぞれ当選確率のおこり方を独立と見做せるので　$p_A\ p_B\ q_C\ q_D$

II　〃　〃　$p_A\ q_B\ p_C\ q_D$

⋮

VI　〃　〃　$q_A\ q_B\ p_C\ p_D$

となる．そこで例えばAの当選する確率を出してみると，

$$p_A(4,\ 2)=\frac{P_r\{\mathrm{I}\}+P_r\{\mathrm{II}\}+P_r\{\mathrm{III}\}}{P_r\{\mathrm{I}\}+P_r\{\mathrm{II}\}+P_r\{\mathrm{III}\}+P_r\{\mathrm{IV}\}+P_r\{\mathrm{V}\}+P_r\{\mathrm{VI}\}}$$

I, II, III は A が I_2 に属する場合である．同様にして，D の当選する確率は，

$$p_D(4,\ 2)=\frac{P_r\{\mathrm{III}\}+P_r\{\mathrm{V}\}+P_r\{\mathrm{VI}\}}{P_r\{\mathrm{I}\}+P_r\{\mathrm{II}\}+P_r\{\mathrm{III}\}+P_r\{\mathrm{IV}\}+P_r\{\mathrm{V}\}+P_r\{\mathrm{IV}\}}$$

となる．III, V, VI は D が I_2 に属する場合である．

これによると，

$p_A(4,\ 2)=0.75$

$p_B(4,\ 2)=0.60$

$p_C(4,\ 2)=0.44$

$p_D(4,\ 2)=0.21$

となり，かくして $R=2$ の場合の当選確率がきまる．

さて，こうして算出された当選確率の分布を次にみよう．また，これが，党派別にどうなっているかについても検討してみよう．この一例として，35年による35年，33年による35年の予測，(33+35)年による38年の予測のものを示してみよう．

§4. 党派別当選者数の予測

さて，各候補者の当選確率が出たならば，これを各候補者の所属政党別につみあげて当選者数の予測値を求めるのである．さて，こうして出た党派別予測者数はもちろん整数ではない．これを定員数とにらみあわせ，誤差が最も少なくなるように整数にして当選予測者数を出すの

衆—16表—1
35年による35年

	未調整	調整済	実際の数
自民党	299	303	295
社会党	131	132	145
民社党	27	24	17
共産党	3	2	3
諸派 無所属	6	5	6
計	466	466	466

衆—16表—2
33年による35年

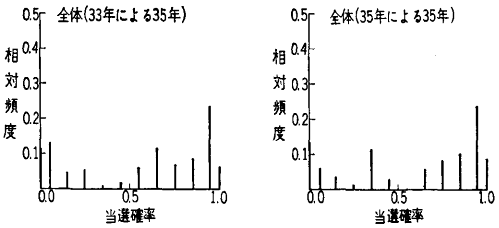
	未調整	調整済	実際の数
自民党	299	302	295
社会党	123	125	145
民社党	36	33	17
共産党	2	1	3
諸派 無所属	6	5	6
計	466	466	466

衆—16表—3
33＋35年による38年

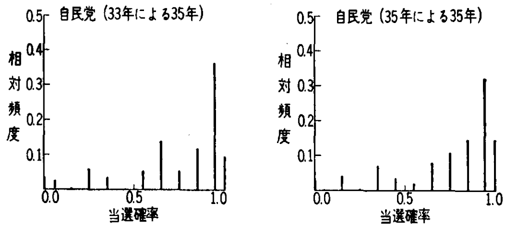
	未調整	調整済	実際の数
自民党	277	284	283
社会党	146	147	144
民社党	25	22	23
共産党	4	3	5
諸派 無所属	14	10	12
計	466	466	467

衆—17表
推定誤差（約95%）

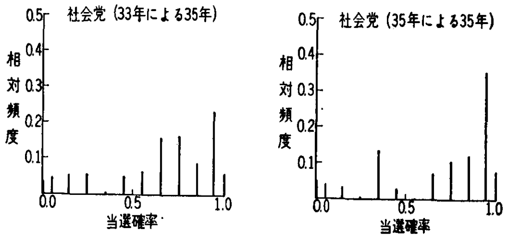
	33年による35年	35年による35年	33＋35年による38年
自民党	11	10	11
社会党	8	7	8
民社党	5	5	4
共産党	1	1	2
諸派 無所属	2	2	4



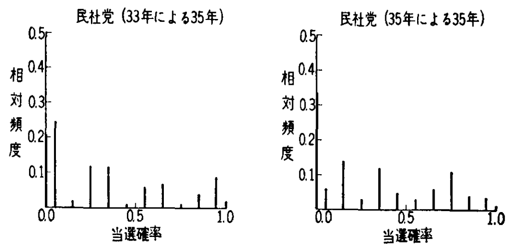
衆—26—1 図



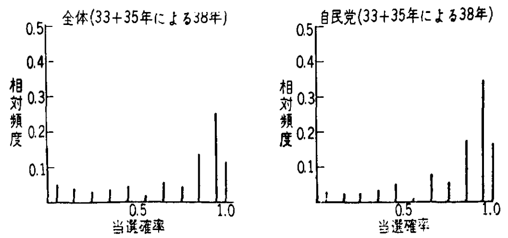
衆—26—2 図



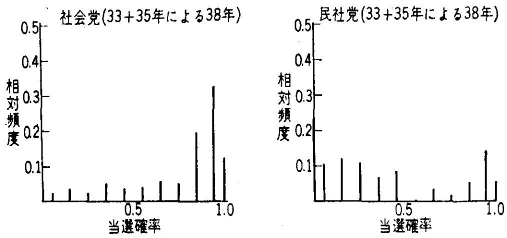
衆—26—3 図



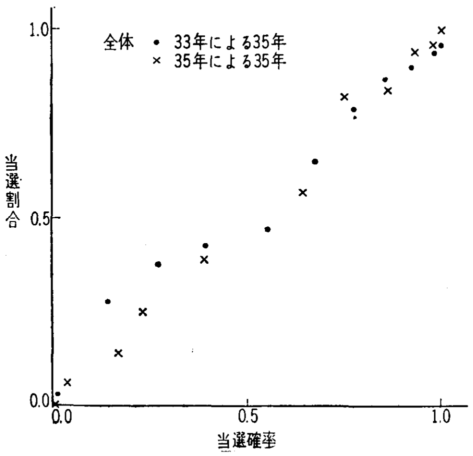
衆—26—4 図



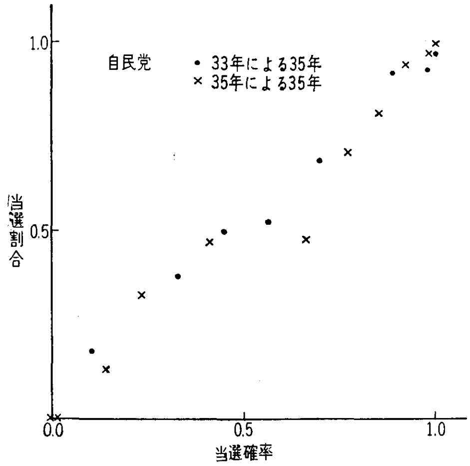
衆—26—5 図



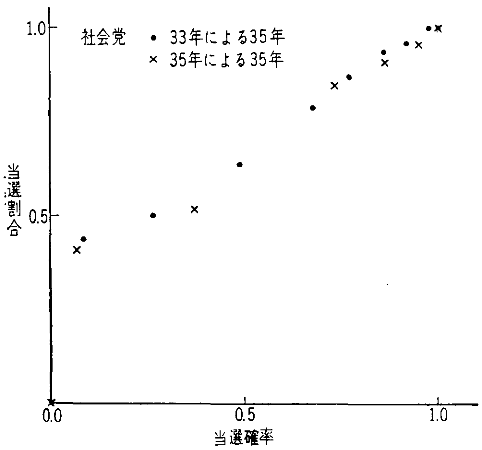
衆—26—6 図



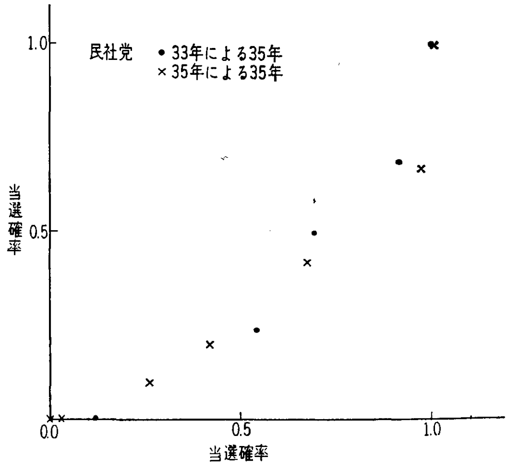
衆—27—1 図



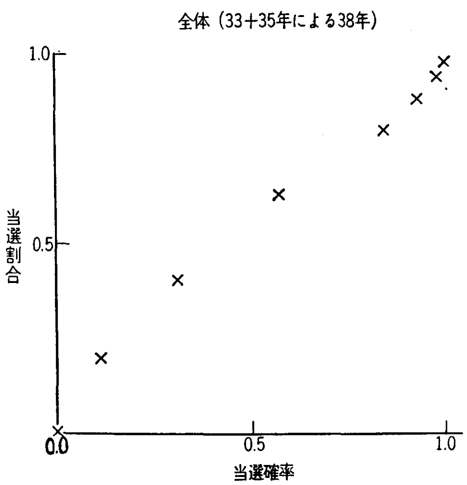
衆—27—2 図



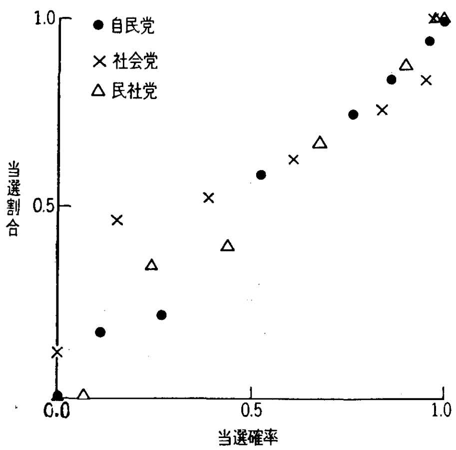
衆—27—3 図



衆—27—4 図



衆—27—5 図



衆—27—6 図

が普通である．――これは当選確率を各選挙区で調整して出した場合でも同様である．こうして出した結果を示そう．

このときも，例示として35年による35年の平面的分析，33年による35年の予測，(33+35)年による38年の予測をみよう（このときの予測は(33+35)年による $f(u)$, $g(u)$ に38年の N_1, N_2 を用いた)．

推定をみると，35年の社会党の当選確率は低目の推定で民社党のは高目の推定となっているが，これは，民社出現の特異的事情のためである．(33+35) 年による38年ではよい一致が示されている．当選確率の分布をみると民社党をのぞき，かなりよいU字型分布を示しており，大局としては満足すべきものである．

さて，推定につける幅の問題である．選挙区で定員調整しないとき l 党の候補者当選確率 $p_i(l)$ の分散の和，すなわち，$\sum_i p_i(l)(1-p_i(l))=\sigma_l^2$ によって l 党の推定当選者数 $\sum_i p_i(l)$ の分散が出る(確率の出方は一応独立と見做せる)．これは一応ガウス分布に従うと考えてよいので，幅は $\pm 2\sigma_l$ によって与えられるのである．これらの当選者数及び幅を整数化するとき，総計が定員数となるべきことを考え合わせ丸めの誤差が最小になるようにする必要がある．

一方，各選挙区において定員数に調整する場合には，このままではまずく，同一選挙区の中に属する同じ党派のものについては関連状況を勘案し，当選確率算出の基礎となるいろいろな場合の様相に応じて計算する必要がある．ただ一般に上述のものとの差はそれほど大きいものではないので一般的に分散は上述のもので代用しても大過はない．

§5.　当選確率と当落割合――結果の検討

§4では，党派別当選者数の予測と実際との比較をしたが，方法論的には十分ではない．当選確率が当選割合をよく表現していなければならないし，当選確率が政党別に同一の意味をもつのでなければならない．つまり l 党の p と l' 党の p とは p が同一なら同一の意味をもたなくてはならないのである．そこで当選確率と当選割合との関係をくらべてみよう．当選確率 p は，当選者割合と対応づけるので，p をいくつかにまとめあげておかなくては，対応がつけられない．そこで p はいくつかの段階に区切り，横軸にその区切りの p の平均値，縦軸にその区切りに属する候補者中の当選者割合を出し，これを目盛ってみるとき45°近辺に小さくバラツキ，確率0は0，1は1と対応してくれることが望ましいのである．これが検討となるのである．

データは前のように35年データによる35年，33年による35年，33年+35年による38年のものである．この結果によると，35年の社会，民社については，民社の突如の出現のため当選確率の低いところで不満足な結果を得たが，(33+35) 年による38年では十分満足の行けるものとなったが，検討の余地もいくらかある．これについては，エキスパートの判定を入れることも考慮し，調査と主観の融合をはかることが必要になろう．

第2篇　参議院議員選挙における予測方法

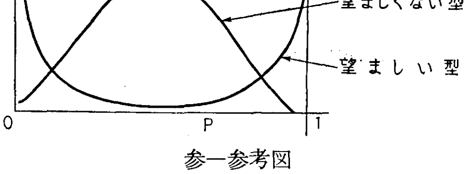
参議院の場合も，選挙調査の方法は同様であるが，予測の方法においては，選挙の性格が異なるため，異なったものを考えることが望ましいことになる．以下全般の見通しをよくするため目次をかかげておこう．

第1部　序　　論

分析は，全国区，地方区にわけて行ない，地方区は1人区，2人区，3人区以上の三つにわけて行なうことにした．ここでは，37年度の参議院選挙の予測を行なうという立場から，34年度におけるデータ（調査ならびに選挙結果，特に終盤調査*）の分析を中心とし，31年データをそのチェックとして用いた．すなわち，34年データの分析から，37年の参議院選挙予測がどの位の精度で可能であるか，あったかを示してみようと思う．さらに，34年データと37年データをつきあわせ，その異同を検討し，さらに妥当性ある予測方法を練磨し，その結果を示しておくことにする．

なお，34年，37年ともに終盤調査（投票日6～7日前）は，原則として，各県800の標本で行なわれている．各県内を地区——市郡別，産業構造，政治率（過去の参議院選挙における保守・革新の比率）によって区市町村層別，このとき，地盤の考えを入れるため，これも原則として郡の区分をこわさないようにする——さらにこの層内に含まれる投票区を，投票区特性によって，投票区を層別する．標本の割当は，有権者に応じた比例割当，各層から，一つの投票区を確率比例抽出，この抽出された投票区から指定標本の大きさだけ，選挙人名簿から個人を等確率で抽出するのである——投票区における標本の大きさは12前後であり，7人～14人にわたっているものが多い．一言にしてつくせば，投票区を第1次抽出単位，有権者個人を第2次抽出単位とした層別副次抽出，比例割当法によるランダムサンプリングによる調査といえよう．なお全国への積みあげにおいては，県別のウエイトを乗じ，推定を行なっている．なお，序盤調査（19～20日前）は，標本数をおとして，大勢をみるために行なうもので——方法は同一——この予測方法のためには，参考程度に用いることにする．

なお，前にも述べたように，われわれの予測方法の趣旨は，各候補者に対する調査情報から，それに解析を加え，各候補者の当選確率を算出することにある．この当選確率は，衆議院の場合と同様に，党派別にみても，全体的にみても同一の意味をもち（当選確率 p とは，その確率を示すものの人数が n_p 人と



参一参考図

* 34年の序盤調査は全地方にわたっては行われなかった．

するとき，平均当選者が $n_p p$ 人であること，また当選者の分布は一応二項分布となる)，この当選確率ができるだけ，U字型分布（0と1とが大きくなる）になるようにし，しかも，その方法に時間的安定性のあることが望まれるのである．

すなわち，なるべくはっきりとものがいえるようにすることで，理想的には $p=0$, 1 の二つにしたいのであるが，もちろんこれは不可能である．

§1.　調査項目

調査では，各人の性，年令，職業，支持政党のほか，投票する人をきめているか，きめていれば誰に投票するか（候補者の名前を書いたもの等を一切みせずに，このようにたづねる）をたづねる．こうして「名前の出たもの」を「(イ)によるもの」と名づけておく．ここで名の出なかったものについて，「いま投票に行くとすれば，誰に入れるか」，これで名が出なければ，さらに「一番よい候補者と思うのは誰か」「ぜひ当選させたいのは誰か」とたづねかけて行く．

これらのものを総合して「名をあげたもの」ということにする．「名をあげたもの」の総数を以下では t であらわすことにする．

集計においては，名をあげたものを100として，各候補者別の支持率を算出することになる．

なお，以下の分析をするに当ってまづ標本誤差の問題を考えておかなくてはならない．この考察を最初にあげておく．

§2.　標本誤差の問題

(イ)　全国区

i なる候補者の調査支持率 x_i（母集団の値は $\bar{x}_i$ とする）の分散 $\sigma_{x_i}{}^2$ を考えるのである．調査は，県別に層別し，これに対して標本800（1,000の所もある．北海道，東京，愛知，大阪，兵庫，神奈川，福岡）調査地点は大体のところ60としておく．煩をさけるため添字の i をおとしておく．

しかも県別にはさらに層別をほどこし，投票区を抽出し，それより個人を抽出してある．これを一々勘定するのは複雑すぎるので層別なしに全国から約2,900地点，40,000人を抽出するものとして，分散を勘定すれば，大き目の分散を示すものと考えられるのでこれを用いれば，なお安全である．標本にしても40,000人であるが，名をあげたものは25％と考えられるので（前回昭和31年の調査のときの数字）結局，地点（抽出調査投票区）10,000人の標本を抽出したものとして調査区間分散，調査区内分散を勘定してみよう．これから，

$$\sigma_x{}^2 \doteqdot \frac{\sigma_b{}^2}{m} + \frac{\sigma_w{}^2}{mn}$$ によって $\sigma_x{}^2$ を計算するのである．

$\sigma_b{}^2$　は調査区間分散

$\sigma_w{}^2$　は平均調査区内分散

m　は抽出調査投票区数

n　は平均一調査投票区当り標本数

$mn \doteqdot 10{,}000$

$m \doteqdot 2{,}900$

x　はある候補者の調査支持率

$\sigma_w{}^2$ の推定は簡単で $\bar{x}(1-\bar{x})$ とみればよい．$\sigma_b{}^2$ の推定には次のような仮説を置くことにする．

まず見透しをよくするため，調査区の大きさは一定としておく．次に調査区における候補者

の支持者数はポアソン分布をするとみるのである．つまりある候補者の支持者はランダムに調査区の中に配分されていると考えるのである．

なお一調査投票区の大きさは平均的に31年度のデータによれば 1,300 (57,010,000 (有権者数)/44,000(投票区数))であるから0.25を乗じ320と考えてよい．したがって平均は $320\bar{x}$ となる．この分布の分散は $320\bar{x}$ となる．したがって320で割ったところの平均値は $\bar{x}$（得票率）となり調査投票区間の外分散は，$\bar{x}/320$ となる．これが σ_b^2 である．

$$\sigma_b^2=\bar{x}/320$$

これは非常に小さいものとなる．

こんどはポアソン分布をしないで，調査区におけるある候補者に対する支持率 v の分布を，

ce^{-kv^2}，としてみる．

これについてその平均 $\bar{x}=0.005,\ 0.012,\ 0.02$ として（この値の程度が調査でよくでてくるものである）分散を勘定してみると，

参一1表

$\bar{x}$	分　散
	ce^{-kv^2}
0.005	0.0000140
0.012	0.0000767
0.02	0.000617

となりやはり小さい．そこで地盤のあること等を考慮に入れ，次のような仮説をとってみた．分布としては，ce^{-kv^2} をとり $v=0.8$（地盤のところで支持を得る比率が0.8であるとする）とし，それ以上は存在しないとする．

さて，上述の場合，すなわち地盤を考慮に入れた場合，$\bar{x}$ が，前の $\bar{x}=0.02,\ 0.012,\ 0.005$ にそれぞれ対応して 0.025, 0.016, 0.008 になるように考えて $v=0.8$ のところの割合をさだめれば，参一2表のようになり，一般におこりそうな型になる．これに対応して分散はそれぞれ，0.0040, 0.0031, 0.0024 となる．これから σ_b^2/m と σ_w^2/mn とを比較してみよう．

両項が等しいと見れば，安全といえることが解る．そこで，この規則を応用して，

$$\sigma_x^2\fallingdotseq 2\frac{\bar{x}(1-\bar{x})}{mn}\fallingdotseq 2\frac{x(1-x)}{mn}$$

を用いることにする．これが調査支持率 x の標本分散であると見做しておいてもよいことが了解せられる．

なお，調査不能による偏りや調査におけるその他の誤差はとり立てて問題にするほどのものではないことは，第1篇に触れておいた通りである．

(ロ) 地方区

この場合も，第1次抽出単位たる投票区の間の分散を加味する必要がある．この種の調査では，第1次抽出単位の間の分散は，かなり小さいが，第1次抽出単位の数を顧慮し，第2次抽

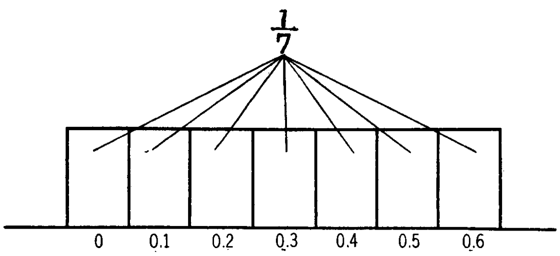
参一2表

$\bar{x}$	$v=0.8$ の比率
0.02 →0.025	0.63 %
0.012→0.016	0.48 %
0.005→0.008	0.375%

参一3表

調査支持率	σ_w^2/mn	σ_b^2/m
0.025	2.44×10^{-6}	1.38×10^{-6}
0.016	1.57×10^{-6}	1.07×10^{-6}
0.008	0.79×10^{-6}	0.82×10^{-6}

出単位の間の分散 $\bar{x}(1-\bar{x})/n$ ($\bar{x}$ はある候補者の得票率，n は標本数）の2倍をみておけば十分である．この検討は，日本人の国民性（統計数理研究所国民性調査委員会，1961）における標本抽出の項によって，その妥当性が示されており，以下の例示によってもこれを確めることができる．なお，ここに標本数の n としては，名をあげたものの数をとるものとする．これは，終盤調査で40％程度なので，調査可能性を考えて，一応250程度としておこう（この分析については以下に示す）．例示をあげておこう．ある候補者の支持率の母集団の値として $\bar{x}=0.30$ を仮定してみよう．第1次抽出単位の間の分散の，いわば，平均的な分散を予想してみるのである．第1次抽出単位の各層内での層平均との差のバラツキが，下図のようであるとしてみよう．分布は0.3を中心とし，±0.3 等分布で拡がっているとしよう．各候補者についてみると，この位の拡がりはあるかも知れないが，一般にいってこれは十分大きい見積りであろう．



参一参考図

$$\begin{aligned} mn&=250\\ m&=60\\ \bar{x}&=0.3\\ \sigma^2&=\frac{\bar{x}(1-\bar{x})}{mn}+\frac{\sigma_b^2}{m}\\ &=\frac{0.3\cdot0.7}{250}+\frac{\sigma_b^2}{60}\\ &=0.00084+0.00067\\ &=0.00151 \end{aligned}$$

となり，これでも第1項にくらべ 第2項は小さい．そこで，分散として，第1項の2倍をとっておけばまづ安心である．十分大きな標本誤差の見積りとなることが知られよう．

§3. 名をあげたものの比率

なお，以下，分析するにさいして，名をあげたものの人数を基礎として用いることになるので，この分析を示しておこう．

名をあげたものの比率は，終盤調査でみると参一4表のようになる．

この37年の数字は調査をしてみてわかったことであるが，きわめて安定したものであることが了解される．このことは非常に興味のあることで，調査の安定性をみる上にも大切なことである．

序盤調査の場合をみると，34年は (13～14) 日前，37年は (19～20) 日前の調査なので，この間の差があらわれている(参一5表)．

なお参考のため，終盤における34年と37年の名をあげた率の関係，34年，37年における序

参一4表

	全国区	地方区
34年(6～7日前)	24.7%	41.7%
37年(6～7日前)	23.9%	38.7%

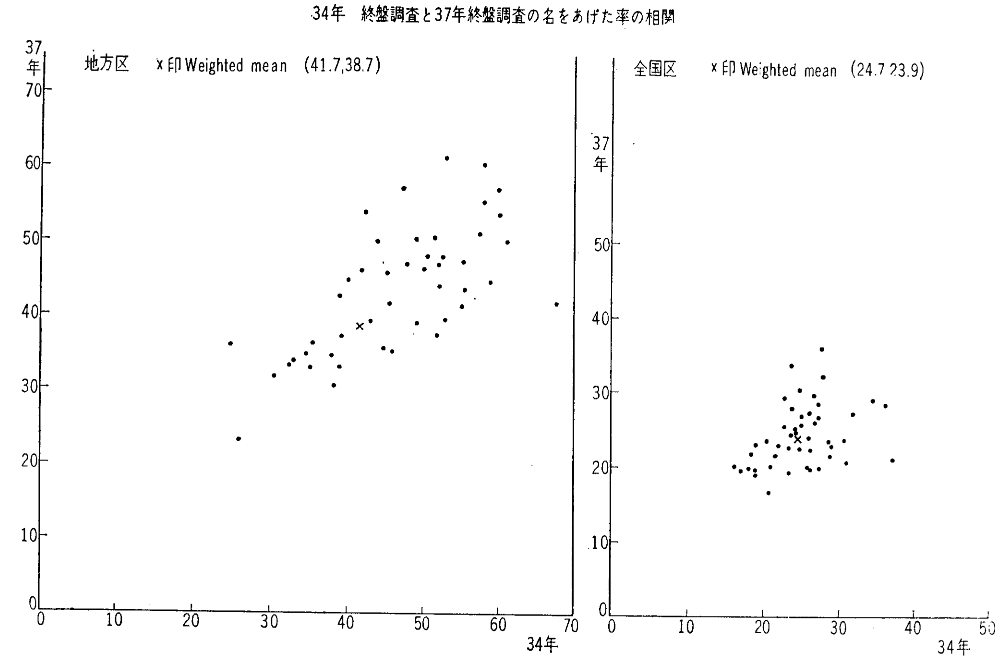
(注) 地方区単位に考えた単純算術平均は全国区 25.1% (34年)，24.1% (37年)，地方区 46.6% (34年)，43.3% (37年) となる．

参一5表

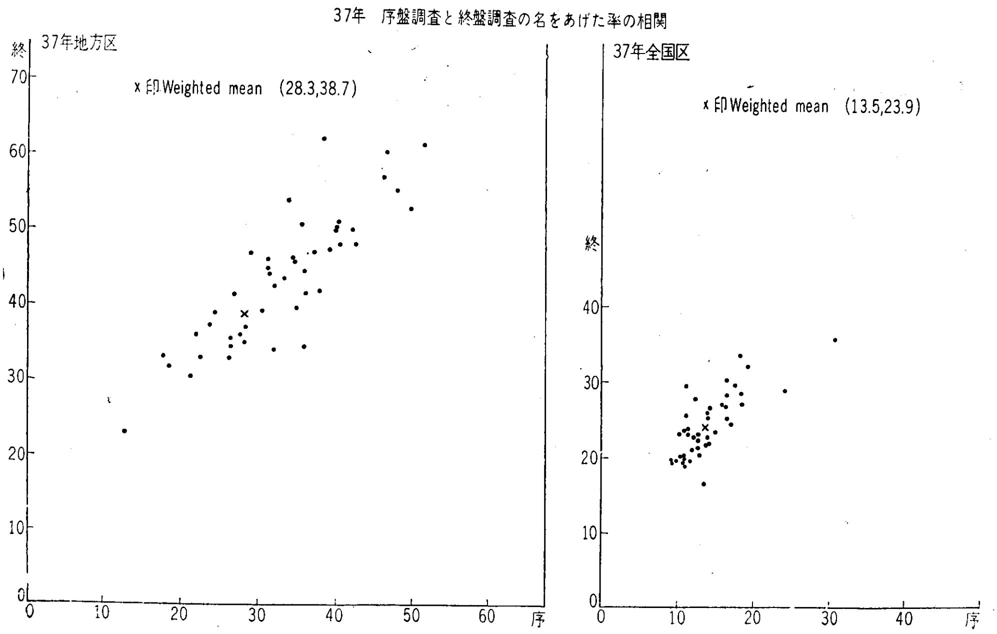
	全国区	地方区
34年(13～14日前)	19.8%	41.0%
37年(19～20日前)	13.9%*	33.2%**

(注) なお34年は序盤では全地方区の調査を行なっていないので，調査できたところのみの単純算術平均の比較を行なってみた．
*，** の全選挙区の比率は，それぞれ13.5%，28.3%である．

盤，終盤の名をあげた率の相関表を目盛ってみた.(参—1図～3図) この後者の間には高い相関が見出されたのは注目すべきであるし，前者においても地方区ではかなりの相関が見られるのも注目してよい——名をあげる率の高い県（地方区）は34年，37年ともに高く，低い県ではともに低いという関係が見られるのである.

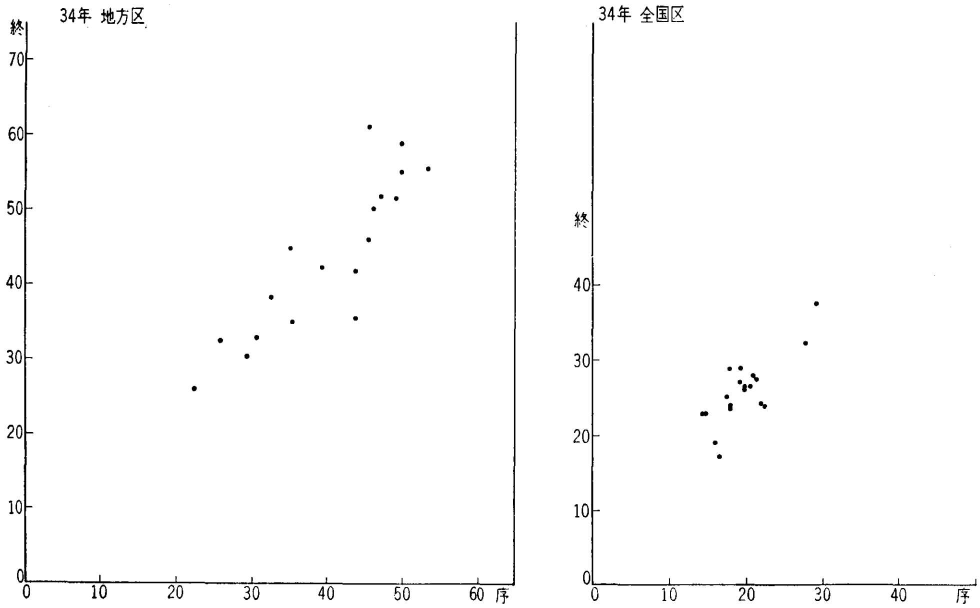


参—1図



参—2図

34年 序盤調査と終盤調査の名をあげた率の相関



参—3図

第2部 地方区の分析

考え方としては，34年度の終盤調査結果と選挙結果をつきあわせ，調査結果の候補者別比率から，選挙の得票率を予測し，これにもとづいて，候補者の当選確率を計算することにある．これにさいして，人区別に特性が異なるので分割して，1人区，2人区，3人区以上とした．

§1. 調査の支持率と選挙得票率との関係

まづ，調査で名をあげたもののみをとりあげ，この中に占める各候補者の比率を計算し，これと選挙結果の得票率との関係をみる．調査支持率を x（%で表示），選挙得票率を y（%）とするとき，$y=f(x)$，$f(x)$ は3次の多項式で精度よくあらわせることが了解される——2次式より3次式の方が望ましいことは直ちに了解される，4次式以上は繁雑になるだけで操作的に見てさして意味がない——．この $f(x)$ は当然人区によって異なるのである．

これについてみると調査支持率が低いところは，選挙得票率が高く，調査支持率が高いところでは，選挙得票率が低目に出るということを物語っており，調査方法と支持率の出方の関係からみて興味深い．45° の直線にのらないところに妙味がある．この3次曲線についての考察は衆議院の場合と全く同様なので省略する．なお，知事選挙においても，3次曲線は全く同様な傾向を辿り，参議院1人区の場合に近いのである．さて，この3次曲線については，

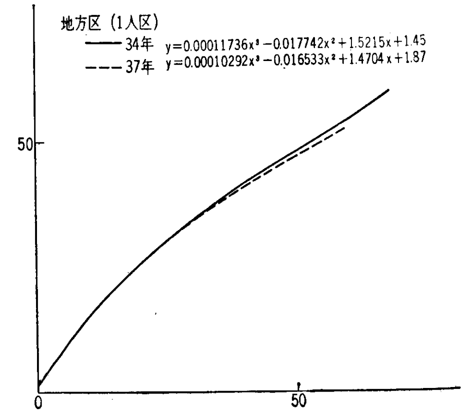
(イ) 34，37年の安定性

(ロ) 序盤，終盤の関係

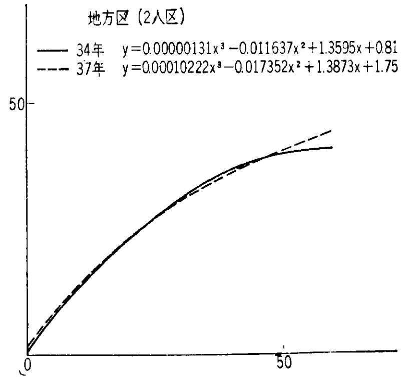
を検討することにする．

(イ) 34，37年データの終盤について，人区別に3次曲線を求めてみた．

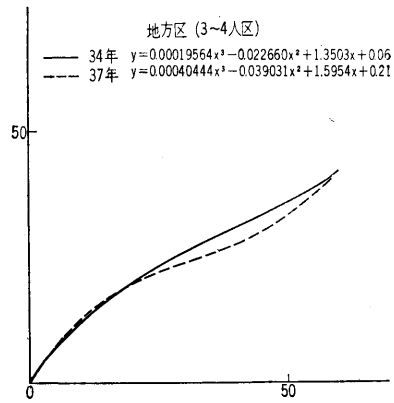
1人区は参—4図に示す通り全くよい一致を示している．2人区については参—5図をみら



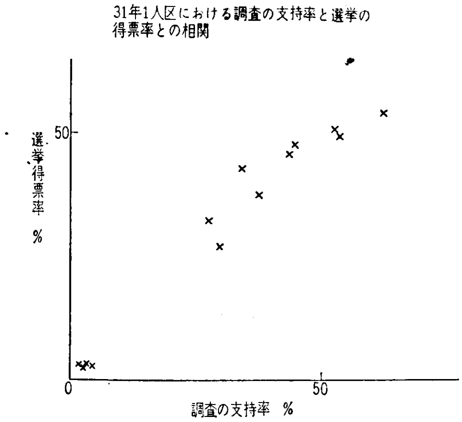
参—4図



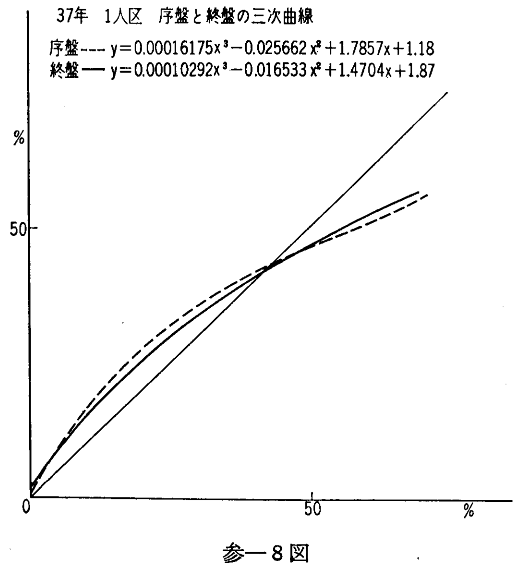
参—5図



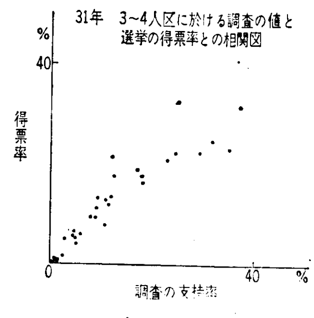
参—6図



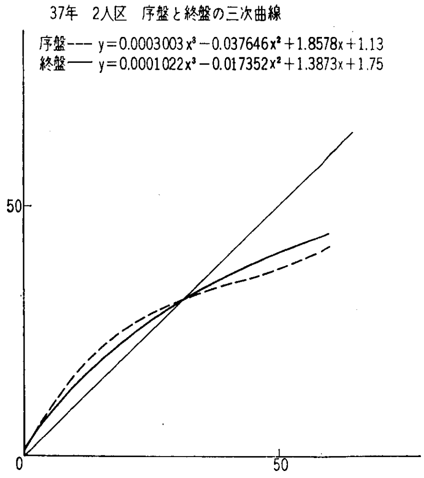
参—7—1図



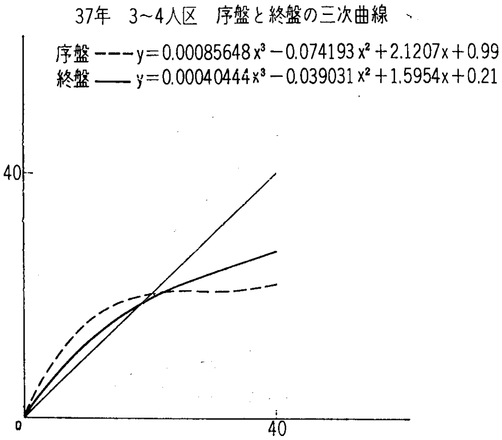
参—8図



参—7—2図



参一9図



参一10図

れたい．50%程度（ほとんどの候補者がこれまでに入る）までは実によい一致で，安定性のあることを示している．また，3～4人区では，前二者とくらべ少し多い食いちがいがみられる．参一6図にみられるように，25%位までは，かなりよく一致しており，それ以上でも高々2.5%の食いちがいである．全般的にいって，かなりよい安定性を示しているとみてよかろう．

なお，31年データをこのグラフの上にのせてみると，よく乗っており，3次式の安定性を物語っているということができる．なお，31年は全県調査を行なっていないので，グラフをのせてみる程度の参考資料である（2人区についてはデータ過少であるので，割愛した）．

(ロ) 序盤，終盤について

37年データについて人区別に序盤，終盤の3次曲線を求めてみた．すべての人区に予想される結果があらわれている．特に3，4人区で甚だしい．なお序盤は20日前，終盤は7日前の調査である（参一8，9，10図参照）．

参一参考図

なぜ予想される結果がそうなるかといえば，選挙前の日数が遠いときには，選挙運動が滲透しないため，調査では通常時の有名人が強く出すぎ，あまり有名でない人が少なく出すぎることが示されることになる．つまり有名人は調査の支持率ほどは選挙の得票率が出ない．一方，有名でない人は，調査の支持率にくらべて，実際選挙ではよけいに得票の比率を示すということになる．このことは，調査方法——候補者名を示さずに，名をあげたもののうちで占める比率を考えるから——から首肯されるところであろう．ところが選挙日が近くなると，運動が滲透してくると，実際に近く，すなわち調査の支持率と選挙での実際の得票率とが45°の直線に近く立ってくることが予想される．この傾向が序・終盤調査に出てくることになるわけである．

このようなことが如実に示されるのは，興味あることである．

(ハ) なお，以上の3次曲線を求めるとき，すべてのデータを入れて計算したのである．しかし，x が小さいところを入れるため，曲線がかなり異なるおそれがあると考えられたので，$x>5\%$ で曲線を引く，$x>2\%$ で曲線を引く，全体で曲線を引くということを34年度のデータ

で2〜4人区を一緒にしたもので（データを増加さすため）試みてみた．結果は，x が65%以下では殆んど差を見出し得なかったので「候補者全体での曲線」をひくことにした．

§2. 調査支持率 x から予測される選挙得票率 y の偏りと分布

x は標本調査から得られる値であるから確率変数である．この分散は $2\frac{\overline{x}(1-\overline{x})}{n}$ によって与えられるし（$\overline{x}$ は x の母集団の値），その分布は n が大，調査地点数が大きいことからガウス分布であると見做しうる．しかし，推定される y は x の3次曲線なので，分布が異なり，また偏りができると考えられるので，これがどの程度のものかを見積っておかなくてはならない．

$$y=ax^3+bx^2+cx+d$$

としておく．

$$\text{確率変数 } y \text{ の平均値}\quad E(y)=aE(x^3)+bE(x^2)+cE(x)+d$$

が求めるものであるが，われわれは x の値の平均値から y の推定を考えていることになるから，それは，

$$[\overline{y}]=aE(x)^3+bE(x)^2+cE(x)+d$$

によって $E(y)$ を推定することになる．この偏りを見積るのである．

$$E(x^2)=\sigma_x{}^2+\overline{x}^2$$
$$E(x^3)=3\,\overline{x}\sigma_x{}^2+\overline{x}^3$$

したがって，偏りは，

$$E(y)-[\overline{y}]=3\,a\overline{x}\sigma_x{}^2+b\sigma_x{}^2$$

予測式が，

$$y=ax^2+bx+c$$

のときの偏りは $a\sigma_x{}^2$ となる．

なお，x の分散から y の分散を求めるためには，$y=f(x)$ とすれば，x が，n のかなり大きいサンプリングによって得られている場合は x の分散はかなり小となるから，

$$\sigma_y{}^2\fallingdotseq f'(E(x))^2\sigma_x{}^2$$

とみてよい．従って，

$$\sigma_y{}^2=\{3\,aE(\overline{x})^2+2\,bE(\overline{x})+c\}^2\sigma_x{}^2$$

として y の分散は計算される．2次式のときは，

$$\sigma_y{}^2=\{2\,aE(\overline{x})+b\}^2\sigma_x{}^2$$

として計算される．

実際にこれを計算してみよう．一人区の場合をとりあげてみる．

標本数は，県単位にみたときの平均名をあげた率が46.6%のところから，800×0.466=372.8とした．

偏りは，

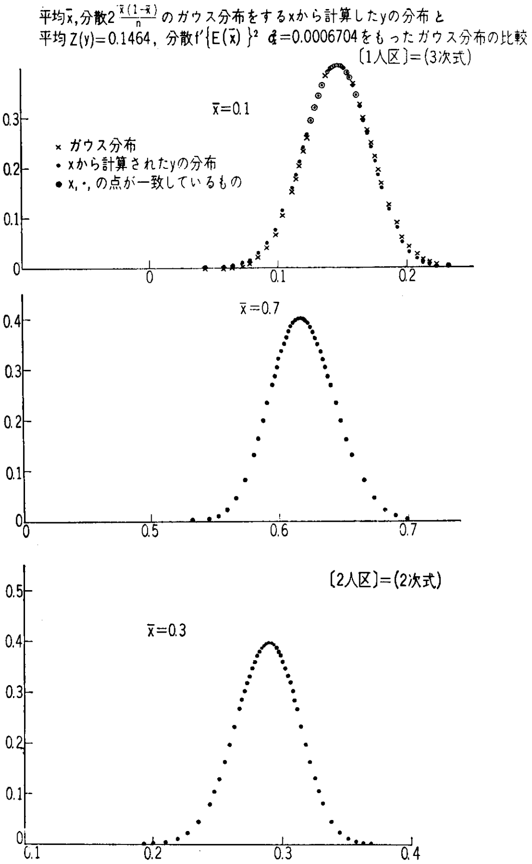
$\overline{x}=0.9$	0.05423×10^{-2}
0.8	0.07111
0.7	0.06014
0.6	0.03082
0.5	−0.00738
0.4	−0.04500
0.3	−0.07254
0.2	−0.08054
0.1	−0.05952

となり，十分小さく問題にならない．つまり実質的には，偏りはないものとみてよい．２人区以上でも全く同様に実質的な偏りはない．

次に x の分布をガウス分布としたときの y の分布をみよう．やはり一人区のところでためしてみる．３次曲線の直線に近い部分を x が動くとき当然 y の方もガウス分布となるが，線が彎曲しているところで試してみるのが一番適切である．このため $\bar{x}=0.1$, $\bar{x}=0.3$, $\bar{x}=0.7$ で試みた．

$\bar{x}=0.1$ のとき推定値 $E(y)=0.1464$ となり（偏りは問題にならない），分散は 0.0006704 となる．

このときの平均 $\bar{x}$, 分散 $2\dfrac{\bar{x}(1-\bar{x})}{n}$ をもつガウス分布をする x から実際に計算された y の分布と，$E(y)=0.1464$ 及び x の分散 $f'\{E(\bar{x})\}^2\sigma_x{}^2=0.0006704$ をもったガウス分布との比較を行なってみたところ，きわめてよい一致を得た．$\bar{x}=0.1$, $\bar{x}=0.3$, $\bar{x}=0.7$ のときのグラフを示しておく（参—11図）．$\bar{x}=0.1$ の場合，ガウス分布と x から計算された y の分布をあわせ目盛っておいた．



参—11図

２人区以上でも全く同様であろう．いづれにしても，平均値に殆んど偏りのないこと，分布がガウスによく近似されることは，それぞれの x が変動する範囲内（x の分散が大きくないので変動する範囲は狭い）では，y と直線的な関係とみなしてよい——局部的に直線，大局的には曲線——と言うことのためであろう．

以上によって，偏りは問題なく，また推定 y の分布はガウス分布と見做せることが判明した．

§ 3.　選挙得票率の推定

まず，調査の支持率 x から $y=f(x)$ によって得票率を予測するのであるが，まだこれでは十分ではない．そこで，候補者ごとに，選挙得票率と調査からの予測得票率の差，d, をもとめ，これを県(地方区)の特性，候補者のもつ要因から推定して行こうとした．もちろん１人区，２人区，３人区以上の３分類にわけて考えた．

(イ) まず，第一に県（地方区）全体における政党支持の状況を加味すべきか否かの問題がある．衆議院の場合などは，候補者の所属政党と政党支持の関係が緊密なのであるが，参議院のときは，衆議院の場合にくらべて，より多く人物本位の要素が加味されること，無所属のものの力が強いこと，などが考えられるので，一応検討を加えてみることにした(34年データ)．

このため，政党の得票率の推定方式を考えることにした．

候補者を i，政党を t とする．

t 政党支持者の中の i なる個人の支持者を $N_i(t)$，政党支持率を s_t（全サンプルで支持者を割ったもの，前に衆議院の時の政党支持率の定義と異る），t 党支持者を $N(t)$ とする．

$$s_t+\left(\frac{\sum_{i\in t}N_i(DK,\ なし)}{N(DK,\ なし)}\right)\cdot S_{(DK,\ なし)}$$
$$+\left(\frac{\sum_{i\in t}N_i(なし)}{N(その他)}\right)\cdot S_{(その他)}$$

$S_{(DK,\ なし)}$ は好きな政党なし，と答えぬものの比率．$i\in t$ は t に含まれるの意味によって t 党の得票率とする．

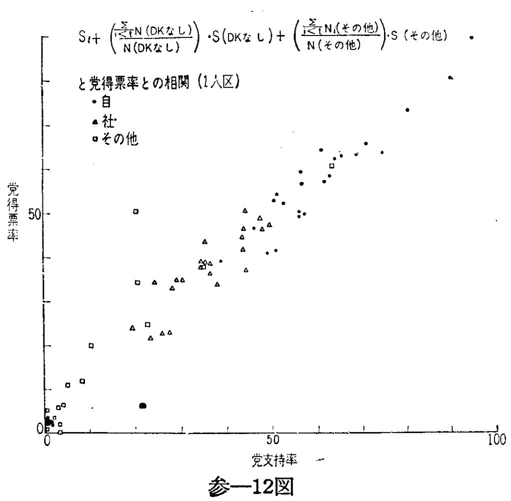
ただし，緑風会（同志会）のときは，

$$s_{緑}+\left(\frac{\sum_{i\in 緑}N_i(DK,\ なし)}{N(DK,\ なし)}\right)\cdot S_{(DK,\ なし)}$$

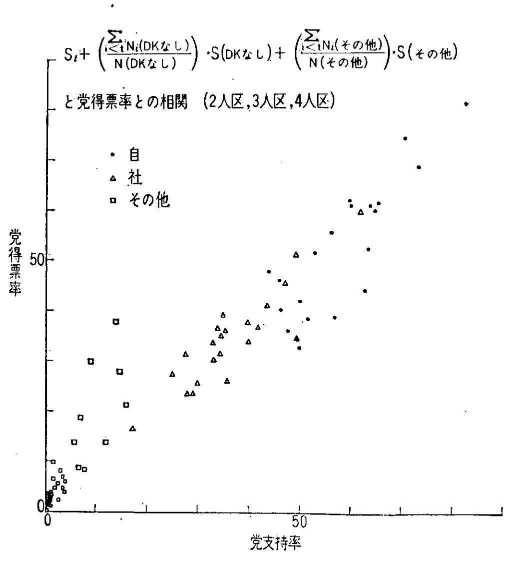
とする．なお，その地方区に緑風会の候補者があるときは，$s_{他}$ はすべて緑風会に，緑風会がなくても“諸派，無所属”があるときに $s_{他}$ はすべて“諸派，無所属”にする．このとき，自民，社会，共産においては $S_{(その他)}$ に入れるものとする．ただし，滋賀では，自民の候補者がないため，$s_{自}$ は緑風会へ，島根，佐賀では社会の候補者がないため，$s_{社}$ はふり分けることにした．

こうして，前述のような政党の得票率と関係が深かろうと一応推定される量をつくった．これと党の得票率との関係を示すと参—12図，参—13図のようになった．

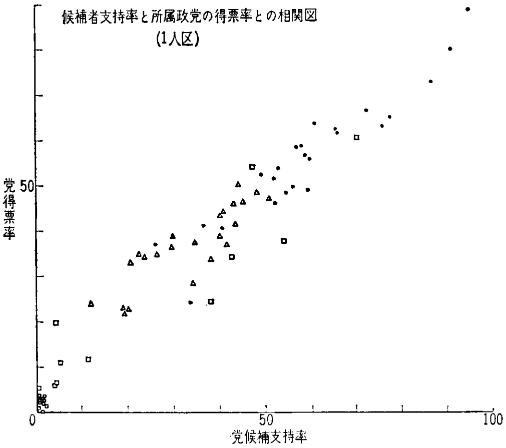
次に，党に属する候補者の支持率と，党の得票率との関係をみるため，それらを目盛ってみると参—14図，参—15図のようになり，党得票率の推定に関しては，政党支持率から推定するよりも，むしろ直截的に政党に属する候補者の支持率のみによる方がすぐれていることが了解される．これは，強力な無所属（とくに創価学会）のあるところや緑風会（同志会），諸派のあるためであることが知られる．



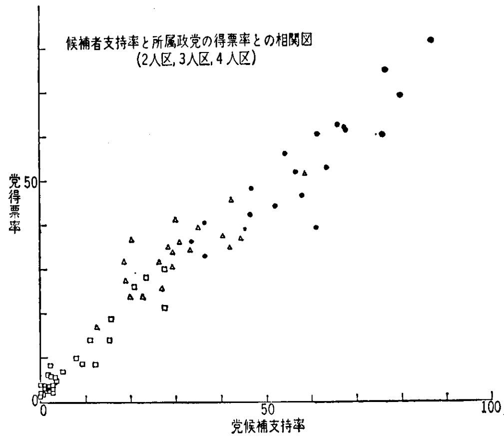
参—12図



参—13図



参—14図



参—15図

なお，さらに政党支持率の小計比（政党名をあげたものを100とした支持率）と選挙結果の党別得票率との関係を目盛って，その関係式を求めてみた．（自，社，共）について，それをみると，$y=0.8479x+1.83$（y 及び x はそれぞれ選挙結果及び政党支持の比率），相関係数は0.910であった．なおこの場合，共産が含まれると，見掛け上，精度がよくなるので，自と社だけで考えると $y=0.9426x-3.18$ となり，相関係数は0.714と落ちる．また $y=ax$ と，原点を通るものとして無理においてあてはめると $y=0.8824x$ となった．これらを通してみて，政党支持と党得票率との関係は，さして強力なものとは思えないことが示された．

以上のような考察によって，政党支持率は，候補者（ある政党に属する）の得票率の推定に関して敢てとりあげる程，有力な要因となっていないと判断された．すなわち，県（地方区）の特性として，政党支持の要因を加味することは，参議院議員選挙のときにはそれほど有効な方法でないと認められるのであって，これからの分析では，政党支持率の知識は用いないことにした．衆議院の場合と異なるところである．

(ロ)　候補者の特性

特性としては，所属政党及び現，新，元，序盤調査からの得票率の延び（終盤におけるものとの比較において）をとりあげることがよいと考えられるので，これらを検討した．まず，延びの問題から考えを進める．

延びといった場合，二つ考えられる．第1は，調査されたものを100として，それぞれの候補者の名をあげたものの比率を出し，これが序盤と終盤でどう変化しているかである．すなわち，

$$\frac{\text{名をあげたものの数(終)}}{T}-\frac{\text{名をあげたものの数(序)}}{T}\qquad (T\text{ は調査可能サンプルの大きさ})$$

を求め，これと d（選挙結果における得票率－3次曲線の読み）とを関係づけることである．

第2は，名をあげたものを100として，候補者別の比率を出し，これが終盤と序盤でどう変化しているか，すなわち，候補者別相対的強弱関係のバランスの変化を算出し，これと d との関係を求めようとするものである．

すなわち，$\dfrac{\text{名をあげたものの数(終)}}{t}-\dfrac{\text{名をあげたものの数(序)}}{t}$　（t は名をあげたものの総数）を求め，これと d とを関係づけることである．

以上二つの方法を検討するためにグラフ化してみたところ，d と関係ずけてみて，いずれも

有力な要因とみなしがたいと判断された．つまり，序盤調査からの知識は，終盤調査の知識以上に加えるものをもっていないと判定されたのである．

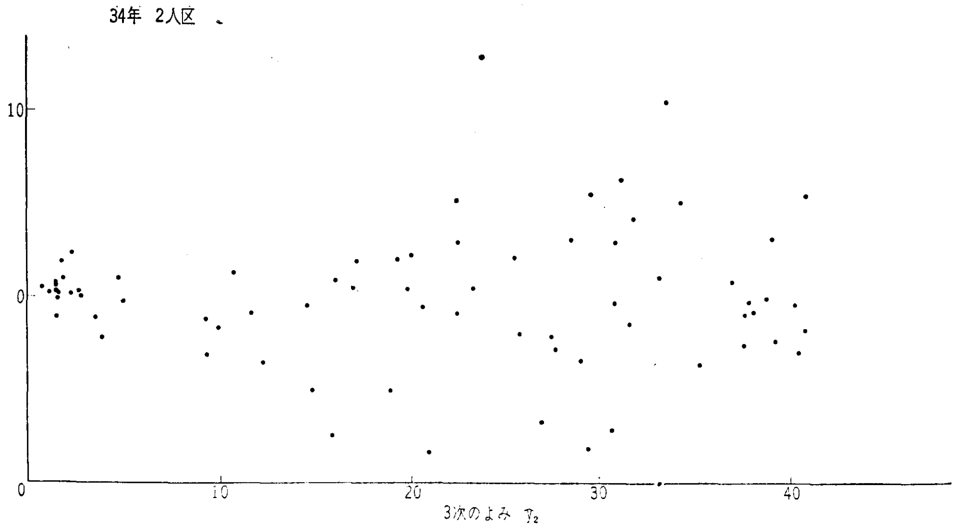
ただし序盤調査の予測における効用は，調査の安定性をみるという質的意味で役に立つものと思われる――甚だしい変化のあるときには，調査法の信憑性を検討することによって，データからの結論を控えめにするという点において．

そこで残る要因をみつけるのであるが，これには，政党，新，元，現の別，さらにこれに調査での支持率を加味することの必要が分析された．これを1人区，2人区，3人区以上について行なってみた．34年データをもとにしたその結果を示そう．

なお，以下の分析では，当落に全く関係のないもの，調査の支持率 x が2％以下の諸派，無所属のものは除いてある．1人区ではとりあげた要因は，政党{(自)，(社)，(共)，(諸派，無所属同志会)}及び経歴{(新元)，(現)}である．これらを要因とし（実際の得票率－3次の読み）を外的基準として要因カテゴリーの数量化を行った．この結果，相関係数は $\rho=0.52$ となった．なお，37年に適用するとき民社のカテゴリーは（諸派，無所属，同志会）として扱かうことにした．2人区では，自民，社会については，政党{(自)，(社会)}経歴{(新元)，(現)}を要因とし，(実際の得票率－3次の読み)$=d$ を外的基準として要因カテゴリーの数量化を行なった．この結果，相関係数は $\rho=0.44$ となった．諸・無所属，同志会，共産については，3次の読み (y) で区分し，この区分を (0≦，<10)，(10≦，<20)，(20≦) とあらわし，このカテゴリーを数量化した（このときは，各区分に属する候補者の（実際の得票率－3次曲線の読み）の平均値をもって各カテゴリーの数量とすることが最適となる)．このときの相関係数は $\rho=0.56$ となった．これを37年に適用するとき，民社は（諸・無所属，同志会，共産）と同じあつかいにした．

さらに3人区以上では，政党{(自)，(社)，(共)，(諸・無所属，同志会)}と経歴{(元新)，(現)}を要因として前述のように数量化を行なった．この結果 $\rho=0.27$ と極めて低く出た．37年に適用するときの民社のあつかいは，1人区と同様（諸・無所属，同志会）と同じにした．

以上のように，各要因カテゴリーに与えるべき数量すなわち，d の推定値を「ゲタ」と名づけ，3次曲線の読みに要因カテゴリーの数量を加えることを「ゲタをはかせる」ということに



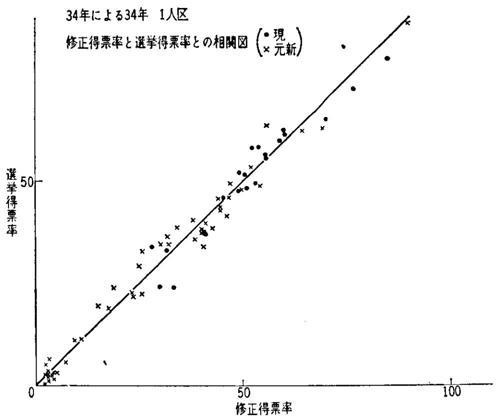
参―16図

する．

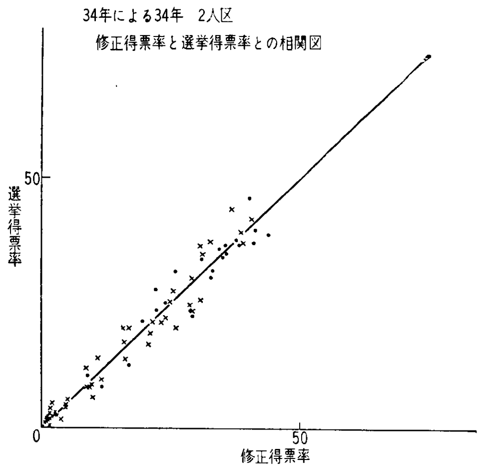
なお，$y=f(x)$ の読みと d' との関係をみるに，参—16図の様に，2人区ではかなりよく関係があり，1人区，3人区以上ではあまりないので，2人区のみ特異の情況となった．

(ハ) 総　　合

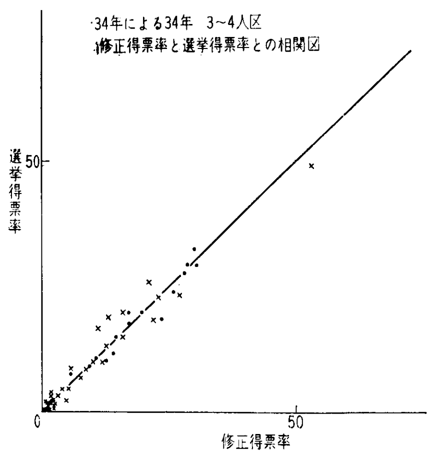
以上のように $y=f(x)$ と各候補者の特性によって与えらるべき数値を総合して，ゲタをはかせ各候補者の推定得票率を算出することができる．なお，人区別に推定得票率 z を出したとき，この人区で100％とならねばまづいので，100％調整を行なった．100％調整のとき，前にゲタ算出のときぬかした2％以下のものについては，3次の読みをもって推定値としたのである（3次の読みで十分よく推定されている）．なお，100％調整以前の総和を求めてみると殆んどが90％～110％の間に入ってきて，（そのうちの大部分が97.5％～107.5％の間にある）それ以外に出るものは，特別の県の事情（ある候補者の得票がアンバランスに大きい場合，立候補者数の多いとき等）によるものである．これと実際の得票率を34年データについて目盛ってみると，参—17～19図のグラフのようになり，一応満足すべき関係がみられた（x が2％以下諸派・無所属をのぞく）．



参—17図



参—18図

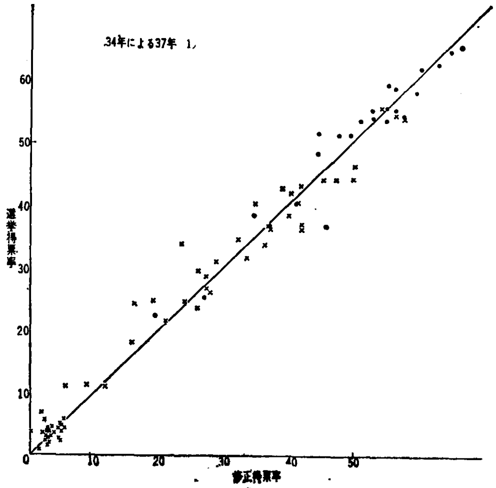


参—19図

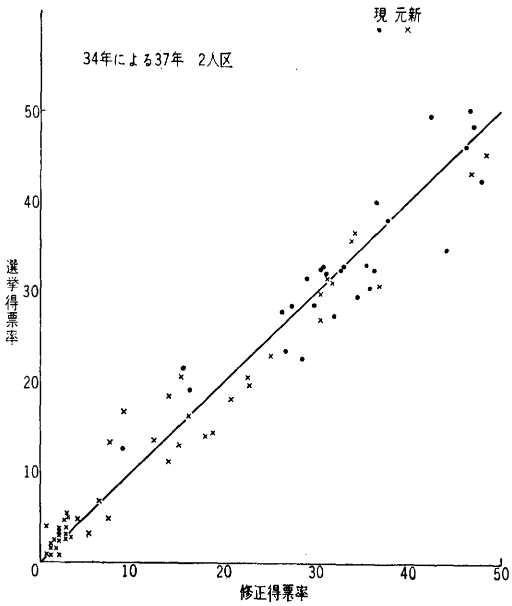
なお，この方式を用い全く同様に37年度のデータに適用してみたところ，参—20, 21, 22図のように，かなりよい推定を得ることができた（x の２％以下の諸派・無所属を除く）.

㈡　候補者の特性の数量の安定性

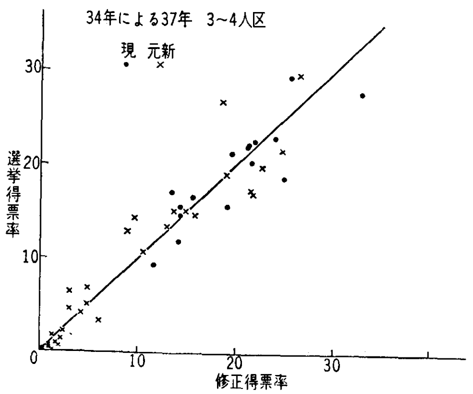
さきに，３次曲線の安定性について論を進めてきた．ここでは，さきに行ったカテゴリーの数量化によるゲタの安定性をみることにしよう．34年の３次に基礎をおいて，34年データから出された数量，37年の３次（34年と殆んど一致）を基礎において37年データから出された数量との比較をグラフ化したものを示しておくと，直線上にならび，要因カテゴリーに与えらるべき数量たるゲタの順序，方向が大局的に34年，37年でくるいがなく，かつ，34年，37年同志の数量が共に一次的な関係にあることは甚だ注目すべきところである．グラフにある印は各カテゴリーに与えられた数量である．45°の直線上にあれば，絶対的に一致していることになるが，45°より傾く直線上にあれば，相対的には一致しているということになる．われわれの場合は，相対的に一致していることがわかった．37年の方が，ゲタのひろがりが少ないのである．



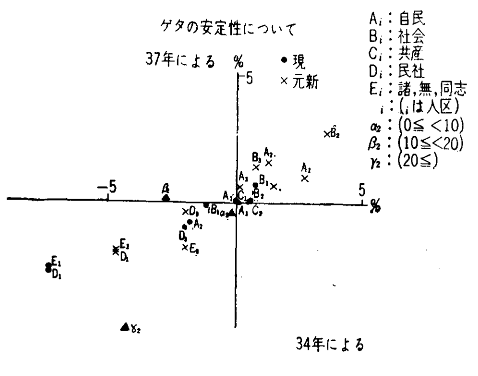
参—20図



参—21図



参—22図



参—23図

こうして，一応，相対的意味での安定性が示されたことは，注目される（3次曲線は絶対的意味で安定している）．

§4.　推定の精度

$z=g(x)$，($g(x)$ は $y=f(x)$ による推定に各候補者のゲタをはかせたもの）によって推定するわけであるが，x のサンプリング分散があるため z も一つの分布をもつべきである．x が前から述べているように分散 $\dfrac{2\,\overline{x}(1-\overline{x})}{n}$ のガウス分布になると考えられるので，z の分布も近似的にガウス分布になると考えられよう（$y=f(x)$ の y が近似的にガウス分布になることは前に示した．また，各特性値における d の平均値をとっているので，加えるべきゲタの値も一応ガウス分布に従うと見られるので，上述の考え方は大過はなかろう）．また，$g(\overline{x})\fallingdotseq\bar{z}$（$\overline{x}$ は x の平均，$\bar{z}$ は z の平均）であることも，$f(x)$ が偏りをもたぬと見做されることから予想されるので，このように考えてくると，調査によって得られる x を知り，これから y を推定するとき y の精度をいかに考えたらよいかは，論理的にはフィデューシァル推論の思想から y のフィデューシァル分布を考えることによって決定するのが妥当となる．すなわち，上述の考えに従うならば x を知ったとき推定すべき z は $\bar{z}$ を平均とするガウス分布として推定されると考えるのが妥当となる*.

以上の考察は，推定理論にもとづく理論的考察である．これを，どう実証的立場から取扱うべきかを考えてこなくてはならない．

§4.　推定得票率の精度

そこで，まづ z の分散を計算することになる．この z の分散の第1次推定として取敢えず，3次の読みによるものだけを考えることにする．もちろん，これだけでは不十分であり，ゲタによるものを加えておかなくてはならない．

一まづ z の第1次の推定分散として $\sigma_y{}^2\fallingdotseq f'(\overline{x})^2\sigma_x{}^2$ を考える．もちろん，$\sigma_x{}^2=2\,\overline{x}(1-\overline{x})/n$ である（n は標本の大きさ，$\overline{x}$ は調査支持率の母集団の値）．以下の計算では，34年においては，$n=250$，37年においては300で計算してある．この値を名をあげたものの平均値に近い丸められた数である．

そこで，$x\xrightarrow{\text{推定}}z$ による誤差，各候補者別にみた(実際の得票率$-z$)$=D_z$ を求め，D_z/σ_y 詳しくは $D_{z_i}/\sigma_{y_i}:\sigma_{y_i}{}^2=f'(\overline{x}_i)^2\sigma_{x_i}{}^2$：を計算し，この分布がどうなっているかをみよう．これが，ガウス分布をするならば，フィデューシァル的思想が，この意味で実際的な意義をもつことになる——フィデューシァルは実証さるべきものではない．われわれはこれを，実際的方法を導くための導火線として用いているに過ぎない——．

ここまでくれば，フィデューシァルの思想はどうでもよいのである．z をもって推定したとき，本当の（選挙での）得票率推定の誤差をどうみれば妥当かを実証的に出してくればよいのである．誤差が平均0，分散云々のガウス分布として表現される，したがって，推定はこれにもとづいて確率的にされる，ということになればよいのである（z であると推定すれば誤差は

---

* この考え方は §4 の項で実証的立場から修正，検証される．フィデューシアルの考え方は通常の統計的方法では出てこない．y は常数であるからである．しかし x を知り y の推定がどうなるかを知りたいのが我々の場合である．このためには y の事前確率を想定する行き方（ベイズ推定）があるが我々の場合，事前確率の推定のしようがないのである．（昔はよく一様分布を想定していたが我々の場合これを裏づける根拠がない）．そこでフィデューシァル的考え方をとったが，この検討は上述のように実証的立場からすべきものであり，純論理上からのみ具体的な妥当づけをしようとしても，することはできないものである．

確率で表現できる）．この誤差を算出するのに，誤差は z について独立で同一の分布をもつ，とする．そこでこれをデータによって出してくることになるわけである．なお，$\sigma_{y_i}{}^2$ を出すときの $\bar{x}_i$，$\sigma_{x_i}{}^2$ はすべて標本の値で代用することにした．また，これからの分析でも，当落に関係のない x 2％以下のもの（諸派，無所属に限らず，すべて）は除いてある．

なお，これも人区別に出す必要があるのであるが，その前に D_{z_i}/σ_{y_i} が z_i と関係があるかどうかをたしかめておく必要がある．このため，z と D_z/σ_y のグラフを目盛ったところ z の値との間に顕著な関係はない．念のため差の大きいと思われる1人区，2人区について z で区切り $|D_z/\sigma_y|$ の平均を示してみたが，次のように，まづ z との関係を云々するほどのものでないことが解った．34年度のデータ，37年度のデータ（37年を34年のデータと全く同じ方法で処理したもの）について行なってみた．

参—7表 $|D_z/\sigma_z|$ の値と z との関係

1人区

年	34 年		37 年	
z	0≦ <20	20≦	0≦ <20	20≦
平均($|D/\sigma|$)	0.93	1.102	0.87	0.87

3〜4人区

年	34 年		37 年	
z	0≦ <10	10≦	0≦ <10	10≦
平均($|D/\sigma|$)	0.70	0.86	0.61	0.93

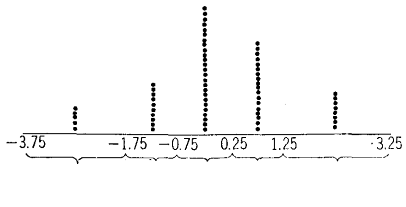
そこで D_z/σ_y を人区ごとに別にして，その分布をとってみた．因みに1人区についてプロットしたものを参—24図にあげておこう．34年，37年（37年による37年）ともに，その分布をあげておく．さらに34年データにもとづく37年の推定の場合の検討をあげておく．

人区別に，平均と分散とを計算し，ガウス分布と見做したときの χ^2 検定を行なってみたところ，全体的にみて満足すべき結果をみたといえよう．

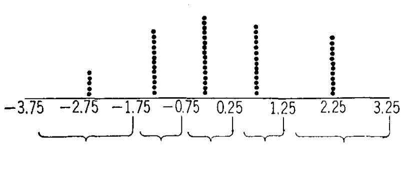
参—6表

		平 均	標準偏差	χ^2_s の値	自由度	確率 $=a<P_r\{\chi^2\geqq\chi_s{}^2\}<b$
1人区	34 年	0.003	1.282	0.950	2	0.70< <0.50
	37 年	−0.052	1.152	1.337	2	0.70< <0.50
	34年による37年推定	−0.026	1.201	3.105	2	0.30< <0.20
2人区	34 年	−0.085	1.158	0.783	2	0.70< <0.50
	37 年	−0.009	1.252	1.824	1	0.20< <0.10
	34年による37年推定	−0.077	1.281	0.646	2	0.80< <0.70
3人区	34 年	0.056	1.073	1.520	1	0.30< <0.20
	37 年	−0.043	1.058	0.245	1	0.70< <0.50
	34年による37年推定	0.028	1.305	1.47	2	0.30< <0.20

37年 D_z/σ_yの分布



34年 D_z/σ_yの分布



参—24図

なお，平均の差が0と有意かどうかをみてもどれもまったく有意ではなかった．

	k
1 人 区	1.2
2 人 区	1.2
3人区以上	1.1

なお，ここで D_z/σ_y の分散が1よりも大きいことがわかる．つまり，これを $k>1$ とすれば，z の分散 σ_z^2 は (D_z/σ_y) の分散の k^2 倍になっていると考えれば自然である．この増加分がゲタによる分散と見做せるわけである．したがって，推定 z の精度の表現としては，$\sigma_z^2=k^2\sigma_y^2$ を分散とし，平均0のガウス分布と考えて行なえばよいことが予解されよう．k は人区別にきめればよかろう．k を大き目に見積り過ぎれば精度をわるく見るので安全目になるが，推定の損失が大きくなり過ぎるわけである．k の値としては，34年の分析，37年による37年の分析でみて，著しい差はなく(かなり安定)，一応，上のようにしておけばまづ大過はなかろう．人区により著変のないのは注目してよかろう．なお，34年による37年予測の場合でも k はこの値に近いことが知られる．

以上要するに，地方区各候補者の得票率の推定は，次のようにまとめられる．しかも，この推定には34年，37年データの分析から安定性がみられるのである．

(i) 調査支持率──→$y=f(x)$ による3次曲線の読み

(ii) ゲタをはかせ推定得票率 z を算出する．

(iii) z による得票率の推定の精度は，平均0，分散

$$\sigma_z^2=k^2\sigma_y^2=k^2f'(\bar{x})^2\sigma_x^2=2\,k^2f'(\bar{x})^2\bar{x}(1-\bar{x})/n$$

のガウス分布によって与えられる．k は人区によって定まるある常数; x は調査支持率，$\bar{x}$ はその母集団値，n は標本数．なお，実際の計算では $\bar{x}$ を x で置き替える．

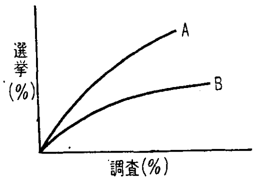
第3部 全国区の分析

全国区における分析は次のような方法によった．まず i なる候補者の l なる県での調査支持率から得票率を推定し，これに対し県のウエイトをつけてつみあげる方式を根拠においた．全国を一本にして，調査での県別支持率（名をあげたものを100とした比率）にウエイトをかけて積みあげたものを3次式で変換して推定得票率を出す方法も考えられる．しかし，積みあげられた調査での支持率が同一であっても，県別に支持傾向が異なるとき，結果は異なってくることが当然考えられる．たとえば，ある県で多く票をあつめ(支持を得ている)，他の県ではほとんど票をあつめていない候補者をAとする．各県でまんべんなく少しずつ票をあつめている候補者をBとする．この二人が全国的につみあげられた結果として，調査で同一の支持率を得たとしよう．県別にみたとき，調査支持率と選挙得票率の関係3次曲線に従うことが予想される．Aはある県で支持を多くあつめているので，調査では非常に高い支持率を示すであろうから，3次曲線でひとまず読むとき，調査より相当引き下げられた比率が推定値とされるであろう（実際の選挙結果もこれに近いであろう）．従って，積みあげられた数値は，かなり低めに出る．一方Bは，各県で少しずつ票をあつめているから，調査における支持の比率は低いので3次曲線の読み（曲線の性質上調査支持率よりも高くなっている！）を読みあげて出した推定は全国一本で出したよりもかなり高く出ることになろう．従って，積みあげられた全国における推定は見掛けより高くなり，Bの推定得票率はAのそれよりも高くなる．このことは，実情によくあうことが実証されている．従ってわれわれとしては，県別に3次曲線を求める方針をとって，分析を進めることにした．

これからの34年データの分析では，100位までの候補者をとり，これに分析を加え，以下のものは別扱いにした．それ以下のものは，0.05%未満であり，全く当落予測には関係しないと考えられたからである．また全国区の定員は50人なので，この2倍をとることにしたからでもあった．(34年では当選者は補欠があったため52人となったが，今後のことを考えれば，一般的に50人としておくことが妥当である)．

§1. 県別得票率の推定

まず参考のため，全国的に積みあげられた調査支持率と選挙結果による得票率の関係を概観してみよう．これによると，やはり2次あるいは3次式の関係（3次式の方が望ましいが，それほどの差はない）が見受けられるのは，いつものことながら興味深い．



参—参考図

さて県別に調査の支持率と選挙結果の得票率とを目盛ったところ，直観的に大別してA，Bの二群にわけられるように思われ，Aは高目に出る曲線，Bは下を通る曲線である．Aには18都道府県，Bは28県ある．(参—参考図参照)．

A，B二群のうち，主観的にA群のうちでBに近いもの(AB)と然らざるもの(AA)，B群のうちでAに近いもの(BA)と然らざるもの(BB)にわかれるので，この4群の代表として，

AA　大阪
AB　岩手
BA　群馬
BB　鹿児島

の4府県をとり，3次曲線をあてはめてみた．その結果，3次曲線の常数項がほとんど変らないのは面白い．なお，4群の曲線を重ねてみると，主観に相違して，曲線はかなりよく一致している．こまかくみると，それぞれの相違が目につくが，さらに検討を重ねてみると，曲線が重なっているA群においては，調査よりとくに高い得票率（選挙結果において）を示す候補者が少しあるために曲線が上方へ移行する結果になっているのであって，この少数の候補者を除けば全くB群に近くなるのである．従って，これらの特例を強調して考えることは，分析の安定性からいって望ましくないので，A，B両群を一本にして引いてみることにした．ただ，このとき，創価学会は特異の傾向を示すので，これを除外することにした(この検討については後述する)．したがって県別にみた3次曲線の算出は34年について，95人，37年については81人となった．これを示すと，参—25図，37年については参—26図のようになる．創価学会候補者については別個に3次曲線をつくることにした．この曲線は一般のものにくらべ同じ調査支持率でも高目に出ているのは注目すべきである．この曲線は参—27図のようになる．

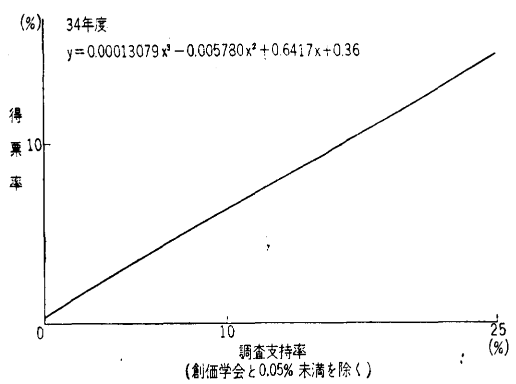
なお，参—27図は，34年データ＋37年データで求めたものであるが，34年の点，37年の点の傾向が全く同様で，34，37年データでともに安定した曲線を示しているのであわせたものを示しておいた(後述)．

こうして，調査の x から y をよみ，これに県別のウエイトを乗じて全国の推定得票率を第一次的に計算することにする．

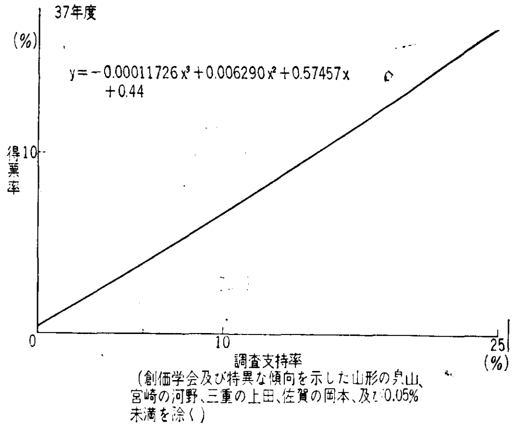
全国につみあげる場合のウエイトとしては，有権者比率を用いることが当然考えられるが，もしも投票率（というよりは，むしろ有効投票率）の相対的傾向が常に安定しているならば，これを用いることが望ましい．これを検討した結果，28年，31年，34年でそれぞれ投票率そのものは変化しているが，県別の投票者（有効投票者）の比率はあまり変化していないことが

確かめられたのでこれを用いることにした．例えば34年の分析を37年に適用するときは，34年と31年との投票率の平均を用い，これに37年の有権者数を乗じ，推定有効投票者数（もっと正確にいえば，これにある常数を乗じたもの）を出し，これの比率をもってウエイトとしたのである．34年のデータ分析のときもウェイトは全く同じ考えによった．

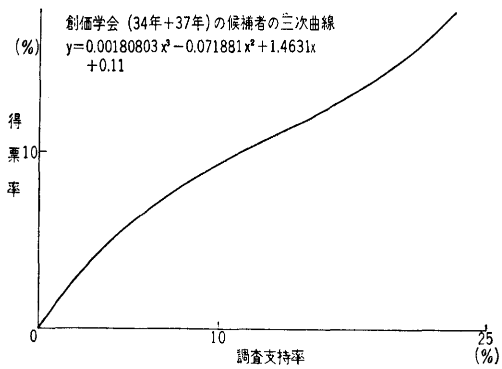
以上の方法によって，i なる候補者の k なる県での得票率 x_{ik} から $y_{ik}=f(x_{ik})$——f は前述の3次曲線——によって県別得票率を推定し，これを積みあげ $y_i=\sum_k w_k y_{ik}$,（w_k は前述の



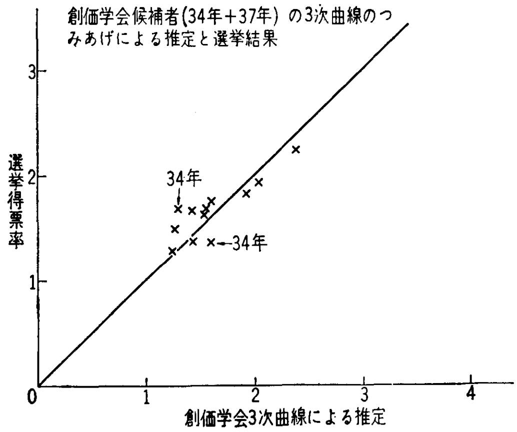
参—25図



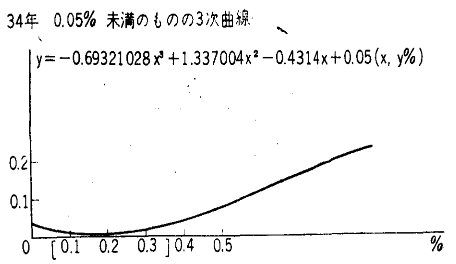
参—26図



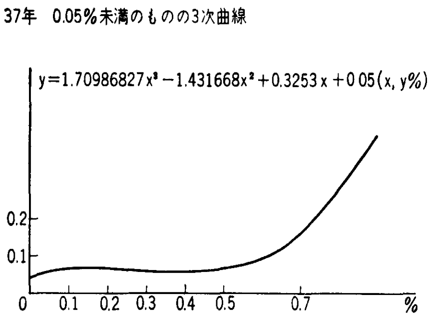
参—27図



参—28図



参—29図



参—30図

ウエイト）によって y_i を推定するのである．しかし，これだけでは実際の選挙結果を十分あらわしているものではない．その差 d を他の要因からつめてゆく必要がある．

なお，ここで除外した0.05％未満のものについても考えておく必要がある．これについては，調査の支持率が0.05％以下のもの（0％も含む）について別に3次曲線をつくってみた．この結果，参―29図をみられたい．34年度については22人のデータを用いた．なお，3次曲線で(0.1～0.3)％のところがへこんでいるが，このところにあてはまるデータがない（$n≒250$，支持者が1人あればすでに0.4％となってしまうのである．0.2，0.3はおこり得ない）．ので問題はない．37年のもの参―30図については後述する．

§3.　候補者の特性の分析

d をつめるに当り，34年の調査データを用いるにさいして検討を加えるべきところがあった．これは，非常な最高得票率をとった第一位の米田議員である．調査では，彼の支持はそれほど強くあらわれなかった．何かの偶発的事情とみなされ，これを含めるよりは，むしろ除いた方が安定した結果が得られると考えられるので，分析には除外することにした．しかし3次曲線を引くときは，さして強い影響がないと思われたので，含めることにした――つまり米田議員を含めても含めなくても，曲線の形はほとんど変化はないのであるが，しかし（実際選挙での得票率－3次の読み）$=d$ をつめるときは，米田議員の d が極めて大きいので，問題がおこってくるからである．そこで分析では米田議員の除外し方が問題になる．

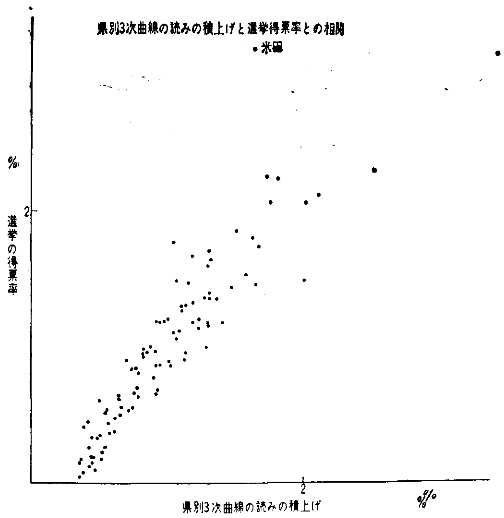
除外し方にはいろいろあるが，(i)各県別に選挙での得票率から3次曲線で逆に調査支持率を推定し，この推定されたものを米田議員の調査支持率とみなし，のこりを各候補者の調査支持率（この算出には，米田議員の分は含まれている）で按分するという行き方，(ii)県別3次曲線を全国につみあげて，各候補者の調査支持率を出し，米田議員は選挙結果と同じ得票率をもつとして，のこりを，得られた推定得票率によって按分するという行き方，がある．

前者の方が望ましいのであるが，実際に，予測方式をつくるとき時間的制約があったため，(ii)の方法をとった．(i)，(ii)の差の全体に与える影響は，さして大きいものではない．

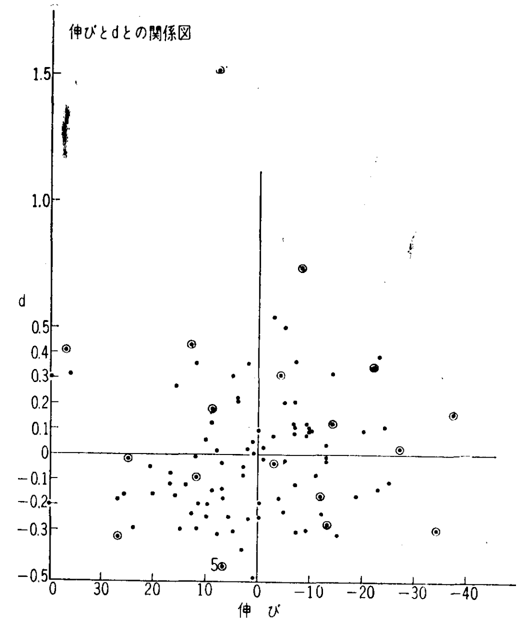
このようにして得られた各候補者別の d を各候補者の特性からつめることにした．まず県別3次曲線からの積みあげと選挙得票率との関係をみよう．両者の関係を参―31図に示しておいたが，45度の直線上にのるべきものと予測されたが，実際はそうなっていない．選挙結果で弱いものが強く出すぎる傾向がある．これは，調査で小さい支持率を示すものに2通りあり，本当に弱いもの（すべての県で全く弱いもの，従って全国での支持率は低い）と，かなり強いが，その強さが調査に現われないもの（他の県ではかなりの支持率があるが，この県では出ていない，従って，全国での支持率はかなりある），との2通りがあり，この二つの平均が3次曲線として出ているので，弱いものが強めに出ていることがわかる．この修正をすべきことが第一に注意される．つまり，積みあげられた推定得票率を要因にとるべきことが示唆される．次には政党，現元新の別である．さらに序盤調査からの延びの問題である．

延びの検討から行なってみよう．34年調査では，序盤調査は全部の県（地方区）には行なっておらず，19県しか行なっていないので，十分な検討は行なえなかった．

延びの第1として，名をあげたものを100％としての序盤と終盤の差を問題にしてみる．これは，調査得票（支持率）の相対的バランスの崩れをみるためのものである．34年の序盤は19県の結果にすぎないので問題はのこるが，一応これを延びの一表現として検討を加えてみる．参―32図に示すように全く関係はみられない．それでは19県が，全県といかなる関係にあるかをみるために，参―32図の各象限にある○印の18人の候補者にたいして終盤で調査された19県の終盤での調査結果から推定した（県別のウエイトつき）全国の結果と終盤での調査で得た



参—31図

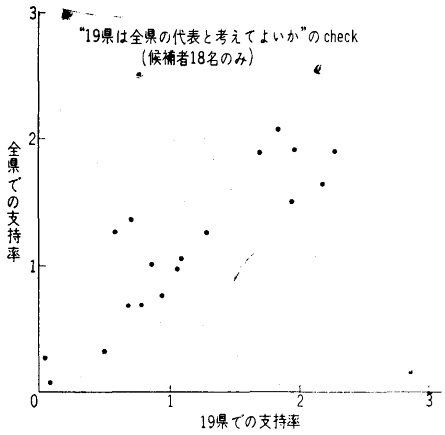


参—32図

全県の結果（これは調査してある）とを比較してみると参—33図のようにかなりよい一致が見られるので，19県を用いての延びの検討もそう不当なものではないと考えられる．

次に，県別にみて序盤・終盤の相対的バランスで＋になった県の数，－になった県の数（このときは両方のデータのある19県のみをとりあげる）と d との関係をみたところ（候補者別）全く関係を見出せなかった．

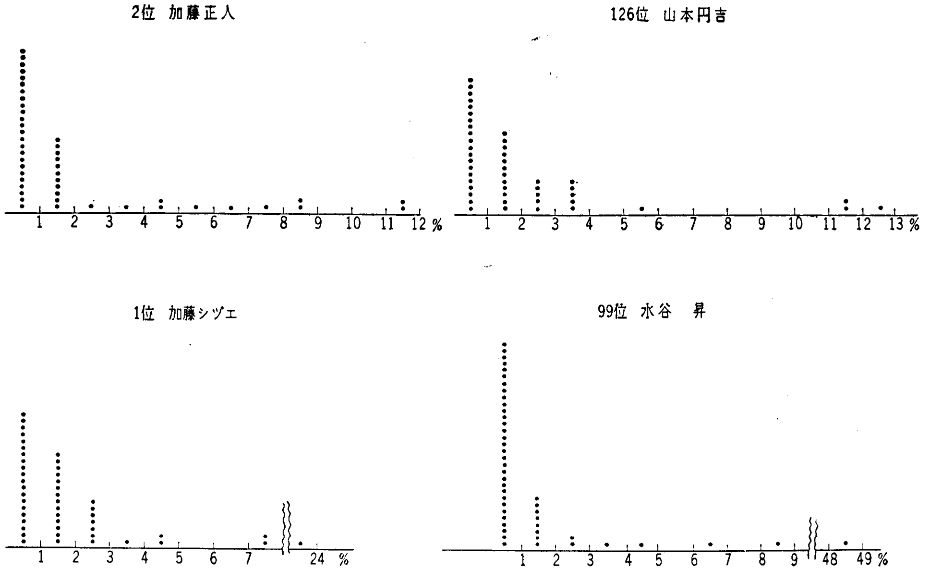
また19県における本当の延びとの関係をみるため，名をあげたものの比率の県平均，序盤19.0%，終盤25.1%を用い，これを前に述べた名をあげたものを100とした支持比率に乗じて，名を得た率（のび）を算出し d との関係を目盛ってみたが，関係を直ちに見出すことができなかった．



参—33図

以上のようなわけで延びの要因はとりあげることはやめた．

また，各県別にみた得票傾向が要因になるのではないかと考えられ，いろいろ分析を加えてみた．各候補者の得票総数を100として，これが各県にどう分布しているか，をみたのである．つまり，各県同じように票をこまめにとっているのが強い要因か，あるいは弱い要因か，地盤があって得票が偏る方が強い要因か，弱い要因かを見出せないかを検討してみることにした．もし，強い要因となっているならば，調査のときの候補者特性としてとりあげようとするのである．調査のデータを用いず，昭和28年参議院議員選挙結果を用い，各候補者の総得票に対する県別得票率の分布をつくってみた．しかし，ここには何の特性も見出し得なかった（参—34図参照）．



参—34図

また，31年データを用い各候補者の県別得票率の分布（県別有効投票総数を100とした得票比率の分布）を，いろいろ特性化してみた．例えば，

- 1％以上が15県以上（A）
- 1％以上が14県以下
  - 5％以上有
    - 1％～5％4県以上
      - 1％以上が8～14県（B_1）
      - 1％以上が7県以下（B_2）
    - 1％～5％3県以上
      - 15%以上が有，しかし1%～15%も1県以上あり（C_1）
      - 15%以上あり，しかも1％～15%はなし（C_2）
      - 又は15%以上なし
  - 5％以上無（D）

のようなことを試みたが，この場合はもちろん利く（総得票率が多いものはよい方の傾向を示す）のであるが，それ以上に当落を左右する要因となっていないことが明らかになった．その他，職業などの要因を参考にしてみたが，調査支持率を一定にしたところでは，当確を左右する要因となっていないことがわかった．

以上のようなことから，全国区でも各候補者の要因はしぼられてきた．政党，経歴{(新元)(現)}その他 y の値の位置である．政党では創価学会を別あつかいしたので，これを除外する．創価学会候補者については3次曲線の読みだけから推定を行なった．

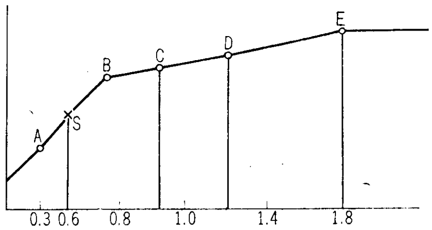
政党では，34年の分析に当って，(自)，(社)，(共)，(同志，諸派，無所属）にわけた，経歴では（新元)，(現）の2分類とした．さらに3次曲線の読み y について，(0.6%未満)，(0.6%≦　<1%)，(1%≦　<1.4%)，(1.4%以上）の4分類を要因とした．

なお，このとき（34年のとき）諸派・無所属・同志会の県別3次の読みの積み上げが1％以上のものは，特異傾向を示すので除外し，87人のデータを用いた（0.05%以上のもの100人−{創価学会5人＋米田＋諸派・無所属・同志会7人}＝87人)．

この3つのアイテムを用い数量化の方法によって{(実際の選挙の得票率−県別3次の読みの積みあげ)＝d}を推定することにした．すなわち，各候補者がそれらの要因でどんなパタンを示しているかを調べ，これと候補者 i の選挙結果と3次の積みあげによる推定得票率との差 d_i とを関係づけるにさいし，各要因の分類（前述のカテゴリー）に数量 x を与えることとし，

各候補者がどんな要因パタンを示しているかを調べ，各要因で反応しているところの数量を加え合わせて i 候補者の推定さるべき値とし，これと d_i との相関係数が最大になるように各 x の値を計算するのである．この x の解は外的基準が数量である場合の数量化の方程式を解くことによって求められる．このとき得られた相関係数は0.66となり，かなり高い値を得た．

さて，d の推定値すなわちゲタから推定値を出すのであるが，一寸問題のあるのは，連続的な y の読みにも拘らず，これをいくつかに区切ってしまったところである．切れ目附近のところ，すなわち，例えば1%近い0.98%と1.02%では与えらるべき数値が，大きくはなれてしまうことになり思わしくない．そこで，段落がないように平滑化して数量を与えることにした．すなわち，下図のように区分されたとき与えられるそれぞれの数量がA，B，C，D，Eとするとき，これを直線でむすび，この直線上の数値をもって，各候補者に与えるべき数値とした．なお，直線で結ぶといったが，A，B，C，Dまではそれぞれの中央値に与えられた数量を対応させて内挿の直線をつくる．Aの中央値0.3以下はそのまま外挿する．Eではそれに属する y の平均値(1.8%) をもってDの中央値と結ぶことにする．1.8%以上は水平に外挿(以上はEに与えられた数量そのものを用いる)することにする．



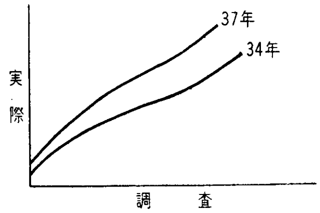
参—参考図

§3. 3次曲線および特性値の安定性の問題

地方区のときも触れたのであるが，3次曲線及び特性値がどのくらい安定しているかを検討しておくことは必要である．

(イ) 3次曲線の安定性

まず，34年と37年の3次曲線の比較から始めよう．ただし，調査支持率で0.05%未満をのぞいたもので，34年では，100人，37年では，88人であった．3次曲線をみると，37年の曲線が34年にくらべて高く出ることがわかった．この原因をしらべたところ，前述の様に創価学会の候補者が常に上方に偏っていることがわかったので，創価学会の候補者を別にしてみることにした．



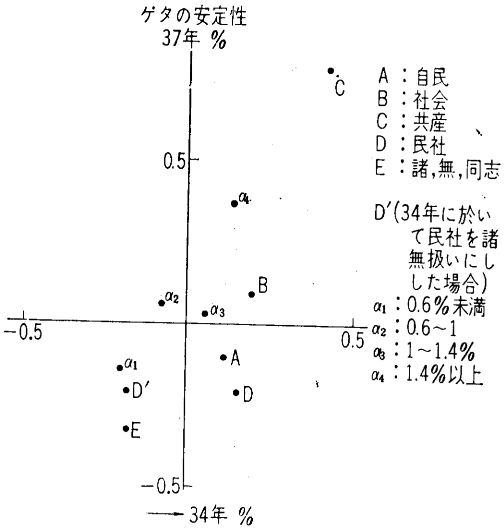
参—参考図

こうして，県別に，各候補者の調査支持率と選挙得票率の関係を34年(95人)，37年(81人)について求めてみた．この結果は参—25図，参—26図の通りである．なお，このとき，37年ではあまりにも特異の傾向を示すもの（泉山——山形，河野——宮崎，上田——三重，岡本——佐賀）を除外した．これらを比較してみると，殆んどのものが現われる10%以下のところではよく一致していることがわかる．これは極めて大切なことである．それ以上のところでは1%前後狂うところがあるが，ある候補者が10%以上をとる県は，一つか二つしかないのが実状であるから（ウエイトをかけ加え合わせたものが全体の調査支持率となることを考えるとき），全体に与える影響は殆んどないといってよかろう．

また，0.05%未満のものについて，37年データによるものと34年によるデータを比較してみよう．37年データは19人である．34年データについては参—29図，37年データについては参—30図をみられたい．これを比較してみると，常数項は34年，37年ともに0.047%程度でよく一致している．0.1～0.3%は問題はないので，ぬかすと0.4～0.7%ではきわめてよく一致している．0.8%以後は外れるが，ここに出てくるものは殆んどないので，大局的にみて，安定

性が保証されているといってよかろう.

これで，一応安定性をたしかめることができた．さて，創価学会はどうなるか，をみるために，創価学会の候補者についてのみ，3次曲線を求めてみた．これは参—27図となる．因みに，参—25，26図と比較してみると，きわめて上方に偏位していることが予解されよう．なお，創価学会については，データが少なく，十分安定性の比較はできないが，点うちの状況からみて——34，37年のデータ入りまざる——34，37年で差異はないものと考えてよかろう．なお，この県別3次曲線を用いて，全国推定を行なうと，極めてよい推定を得ることができた．これは参—28図に示しておく．



参—35図

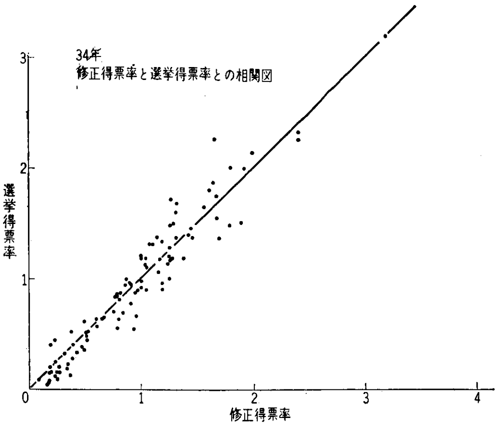
(ロ) ゲタの安定性

さきに，34年データによって特性値の数量化を行なったのであるが，こんどは37年データにもとづいて数量化によってゲタを出してみた（37年による37年の分析）——なお，そのとき政党としては（民社）の項を新につけ加えた——．その結果 0.66 となり，34年による34年の分析と全く同様になった．各カテゴリーに与えられた数量化による数量をグラフに目盛ってみると，これまた大局的には地方区のときと同様に直線的関係が見受けられた．ただし(元新)，(現)の傾向は，34年と37年とでは食い違っていることがわかる．そこで，この区分をつぶしてみると，相当安定した傾向があらわれてきた．なお，ここで34年における民社については，同志・諸派・無所属の数を与えてみた．さて，これら両表の傾向は45°の直線の上にはなく，やや傾いた直線上にのっている．したがって，ここでも3次曲線の絶対的安定性ではなく，順序は相対的位置が安定している意味の相対的安定性がたしかめられたことになる．この安定性を知っておくことは非常に大切なことである．

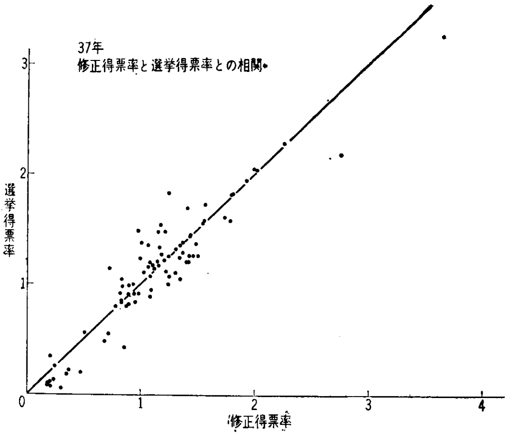
このような安定性が見出されたので，34年と37年のデータをまとめて3次曲線をつくり，かつ，両者のデータをあわせゲタを求めておくことは将来のため役に立とう．

§4. 各候補者の得票率の推定

以上のように，3次曲線及びゲタの値がきまれば，各候補者の推定得票率を求めることができる．しかし，これができるのは，調査支持率が0.05%以上のもの（ただし，創価学会，1%以上の諸派・無所属・同志会をのぞく）である．創価学会については，創価学会県別3次曲線の読みの積みあげによって推定，1%以上の諸派・無所属・同志会は，一般3次曲線の積みあげのみによって推定，0.05%未満のものについては，0.05%未満のものの3次曲線の常数項 0.047% を各人の推定値（常数であるから積みあげても同じものになる）とした．こうして推定したものを100%調整したのが最終の推定得票率となる．100%調整のし方は，0.05%未満のものの票のとり方が非常に安定して固いために，100%から0.05%未満のものの推定得票率をのぞいておき，残りを推定得票率に応じて按分比例によってわけることにした．総和が100%に近い数字になるので，そう問題になるべきものではない．なお，34年度の米田は，ずうっとぬかしてきたのであるが，一応，調査支持率と選挙得票率が同じものとして見做しておく



参—36図



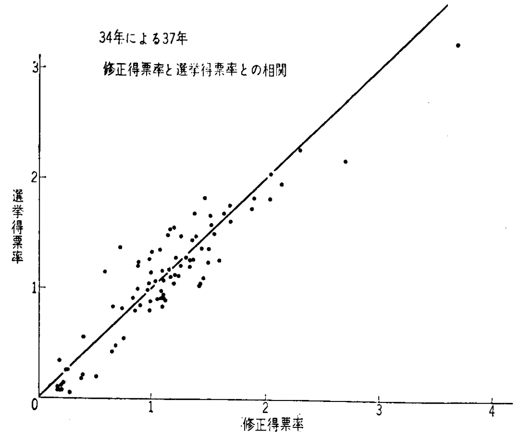
参—37図

——これは，今後予測のための分析においては，こうしておく方が望ましい，すなわち，特例と見做せるので，入れておくとかえって安定性がみだれると思われるからである．調整前のデータ，34年度の推定得票率の総和は，99.42％，37年度のでは101.36％であり，100％にごく近いのは注目してよい．

さて，こうして得られた推定値と実際の選挙の結果を目盛ってみると，参—36，37図のようになり，相当よい一致をみる．34年データによると，相関係数としては0.96となっている．やや離れるのは3人の候補者であった．

また，34年による37年の推定で問題となるのは民社であるが，民社については自＋社のゲタの1/2を用いた．こうした推定の100％調整値と37年の選挙結果を目盛ったのが参—38図で相当よいものがみられた．これは，さきに示した分析の安定性からみれば当然首肯されるところである．1人の候補者のみは離れている．(37年による37年のときも同様)．

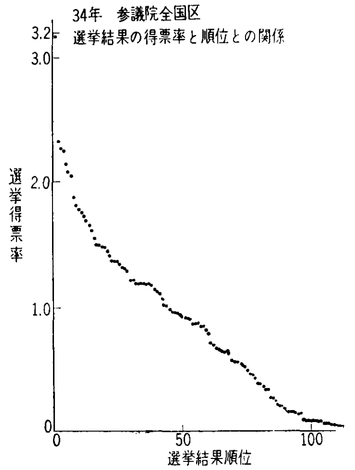
次に，参考のため34年データについて推定得票率による順位と選挙結果の順位との関係を目盛ってみたが，両者の関係はかなり高く，外れは8人であった．すなわち，推定得票率で52位より以上で落選したものは4人であった．52位以下で当選したものも56位までに入っているのである．37年データについて（37年による37年）行なってみたところ，外れは12人であった．すなわち，51位以下で当選するもの6人，51位以内で落選するもの6人あった．34年よりも多少わるくなっている．高順位のもので落選するもの，低順位で当選するものが出たためで，調査の欠陥によるのではないかと判定される．なお，34年データによるものを用いた37年の予測では，これが悪くなり，51位以内で落選するもの9人，51位以上で当選するもの9人あった．したがって，順位による傾向は一応よみとれるが，当選確率の付与の上で推定の分散が増大するものとして処置すべきことを示唆している．



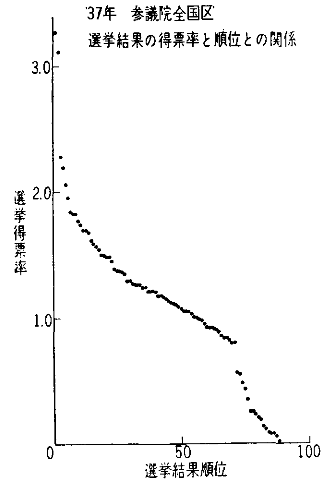
参—38図

なお，参考のため推定順位と推定得票

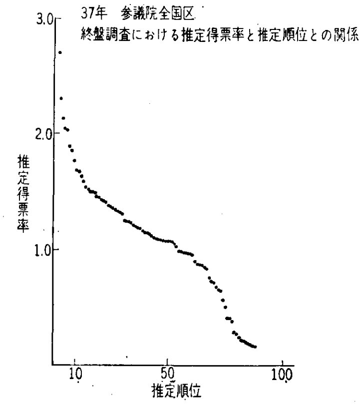
率，選挙順位と選挙得票率とを見よう．34年，37年の後者，37年（34年による37年の予測）の前者をグラフ化してみたところ，参—39, 40, 41図のようになる．37年の実際と34年による37年予測とはほぼ70位まではよく一致しているのは興味深い，つまり人は入れかわっているが，それぞれの順位では得票率と推定得票率との関係には恒常性があるということである．



参—39図



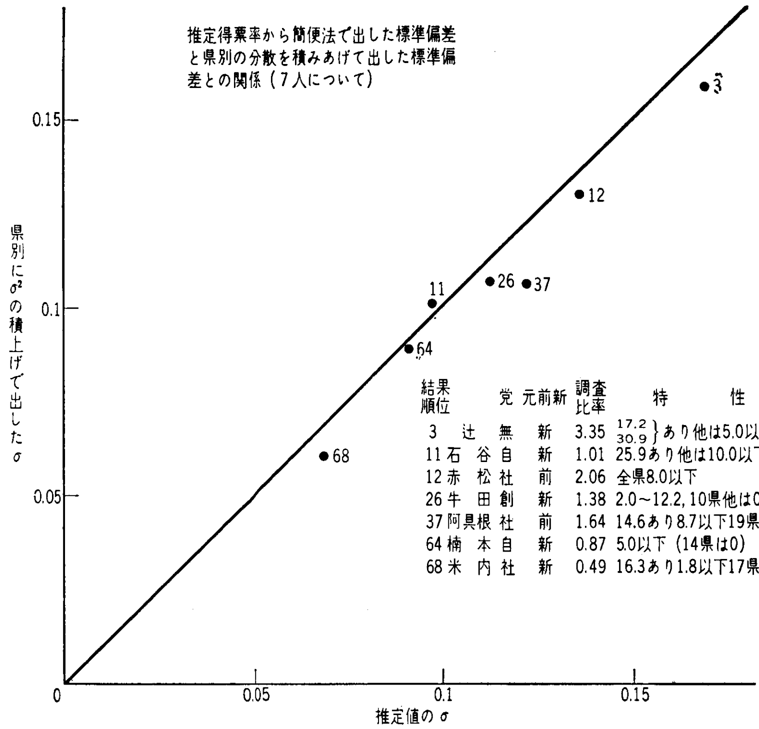
参—40図



参—41図

§ 5. 推定得票率の分散と分布

推定得票率が偏りが少なく，ガウス分布をすることは，地方区の場合と同様に考えられる．3次曲線で県別に支持率を出し，これにウエイトを乗じ全国に積みあげるのであるから，ここまでは偏りはなく，ガウス分布に近くなることは容易にわかる．これから d をつめるところに問題があるが，大きな修正ではないし，地方区で述べたように，偏り，分布の型は，まず変りはないものと考えてよかろう．次に推定得票率の分散であるが，地方区の場合と同様，3次曲線までのところの分散をもって代用することを考えた．$y_i=\sum w_k f(x_{ik})$ であるから，$\sigma_{y_i}^2 \fallingdotseq \sum_k w_k^2 f'(\bar{x}_{ik})^2 \sigma_{x_{ik}}^2$ となる．$\sigma_{x_{ik}}^2$ としては，§1. で述べた分散を用いることにする．しかし，これをいちいち計算するのは，急を要する予測作業においては，大変なので，$x_i=\sum_k w_k x_{ik}$ を用い，$\sigma_{y_i}^2$ として $f'(\bar{x}_i)^2\sigma_{x_i}^2$ をもって代用しようとした．両者の差がどのくらいあるかを



結果順位		党	元前新	調査比率	特性	結果比率	結果順位
3	辻	無	新	3.35	17.2, 30.9 }あり他は5.0以下	2.32	3
11	石谷	自	新	1.01	25.9あり他は10.0以下	1.78	11
12	赤松	社	前	2.06	全県8.0以下	1.75	12
26	牛田	創	新	1.38	2.0～12.2, 10県他は0	1.37	26
37	阿具根	社	前	1.64	14.6あり8.7以下19県他は0	1.19	37
64	楠本	自	新	0.87	5.0以下（14県は0）	0.66	64
68	米内	社	新	0.49	16.3あり1.8以下17県他は0	0.63	68

参—42図

確かめるために，数人の特色ある候補者をとりあげ，両者の関係を検討した．ある県で多くとっているもの，少しずつ票をあつめているもの，得票率の多いもの，比較的少ないもの（ただし，当選の可能性ある範囲内）等7人をとりあげた．両方の標準偏差を参—42図に目盛ってみた．大局的には，簡便法がに従っても，まづ大過はないものと思われる（著しく外れるものはなく，7つのうち6つは簡便法の方が多目に出ている）．

こうして代用された標準偏差をもとにして考えを進めることにしよう．つまり，推定値 z_i は $k\sigma_{v_i}$ をもつガウス分布に従うであろうと予想するのである．さて，ここでまた，地方区の時と同じようなフィデューシャル的な考えが分析を進めるときの出発点となるのである．推定値 z_i を得たとき，それによる選挙得票率の推定をフィデューシャルの考えによって行なおうとするのである．z_i の分布がガウスならフィデューシャル分布もまたガウス分布と考えられるのである．

しかし，前にも述べたように，これは実証することができない——できないのが特性で，できるくらいならフィデューシャル分布を用いなくて推論が可能となる——．したがって，考えの動機としてのみ用いるのである．そこで，z_i の分布，すなわち，z_i で推定したとき，選挙結果がどのくらいに出るかをしらべる——$D_{z_i}=\{(i$ 候補者選挙結果の得票率$)-z_i\}$ を考える——この食いちがいがどんな分布をしているかをみるのである．z による推定の誤差はある確率分布に従い，独立におこるものとし，この誤差の分布は D_z の分布として表現されると考えるのである．これは，極めて実践的な立場となる．

さて，この D_z の分布を考えるのである．前と同様に D_z/σ_y を考え，これに分析を加えてみるのである．その前に $|D_z/\sigma_y|$ と z との関係をみたところ，強い関係は見当らないが大局的にみると z の小さいところで大きく，z の大きいところで小さいという傾向を読みとるこ

とができるのである．そこで0.8％未満，0.8％以上にわけて——2分するならば，ここで分点を入れるのが有効であることがグラフからわかる，2分以上にすることは困難である——$|D_z/\sigma_y|$ の平均をとってみた．この結果，次の表のようになる．

参—8表　最終推定値と $\alpha=(|D_z|/\sigma_y$ の平均）について

z	34 年	37 年	34年による37年の場合
0≦ ＜0.8	4.169	2.449	4.166
0.8≦	1.827	1.728	1.620

0.8％未満のときは，データも大小さまざまありバラツキも多く，年度によってかなり不安定であるが，0.8％以上との食いちがいの傾向は十分みることができる．そこで，大局的に考える場合は，つかえるものと思われる．そこで，以上のようにして0.8％未満では $\alpha\sigma_y$ の α の値をさだめ，これをもって z の分散 $\sigma_z'^2$ とするのである．

z が0.8％未満では z の分散 $\sigma_z'^2$ は　$\alpha_1^2\sigma_y^2$

0.8％以上では　〃　$\alpha_2^2\sigma_y^2$

となる．α_1，α_2 は表によってきめられるのであるが，大きくつかめば，

$\alpha_1=3.0$

$\alpha_2=1.8$

と考えてよいものであろう．

こうして σ_z' をつくり D_z/σ_z' の分布をつくって検定を行なってみた．なお，このとき34年データは米田をのぞいた99人，37年データは86人（88人中特異の傾向を示した2人をのぞく）である．

参—9表

	平 均	標準偏差	χ_s^2 の値	自由度	確率 $a<P_r\{\chi^2\geqq\chi_s^2\}<b$
34 年	−0.091	1.278	0.758	2	0.70＜ ＜0.50
37 年	−0.058	1.250	2.896	2	0.50＜ ＜0.30
34年による37年	−0.119	1.427	4.314	3	0.30＜ ＜0.20

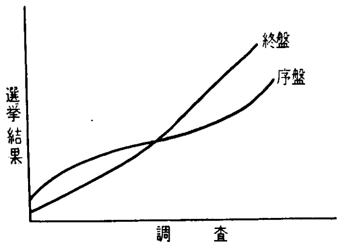
分散としては一応1.6程度と考えてよいようである．そこでこれが $(\alpha\sigma_y)$ に対する分散の増大分で k^2 とあらわすことにする．

これをみるとガウス分布によく一致し，また平均値0と有意差はないと見做せるので，z による推定の誤差は——この誤差分散は z の値によって異なる——ガウス分布，平均0，分散 σ_z^2 は $k^2\alpha^2\sigma_y^2$，$k^2=1.6$ に従うものとして，以下の解析を加えて行こうと思う．このように z の推定誤差の分散 σ_z^2 が増大しているのはサンプリングによる誤差以外のものが多くあるためであると思われる．

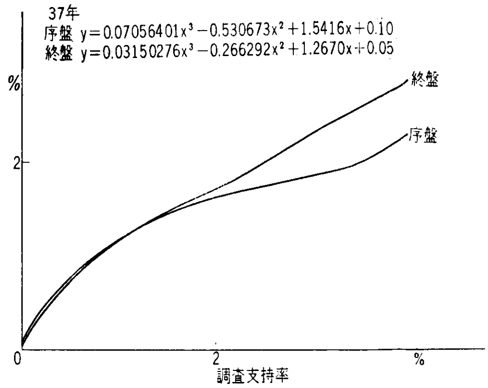
§6. 3次曲線細論

以上によって予測の方法の筋を述べてきたのである．3次曲線決定に到るまでいくつかの分析をほどこしてみた．これについて述べておくことは，選挙調査の実際としてなかなか興味あることが多いので記述してみよう．

(イ) 序盤調査と終盤調査の3次曲線



参一参考図



参一43図

(ロ) 県別にみた名をあげた率と投票率との関係を利用した県の分散について

(ハ) 個人の得票率別にみた3次曲線について

の3つについての考察を述べてみよう.

(イ) 序盤，終盤の差異について

理論的に考えるとき，序盤，終盤で，3次曲線がかわると思われ，拡大して考えれば，上参考図の様相を呈し，投票日から遠ければ遠いほど，曲線はねてくるものと判断され，逆に近ければ近いほど45％のように近くなるものと考えられよう（遠ければ，有名人が調査支持率で多くなるが選挙では，それほどとらない，一方普通の人はまだ滲透が十分でなく調査には出にくいが選挙ではそれよりずっと多く得票する）．まづ37年について，この比較をしてみよう．参一43図をみよう．なお，これは全国一本（つみあげられた調査支持率と選挙結果の関係）で候補者107人別に求めたものである．終盤と序盤との3次曲線をみると，はたして，両者の間に予想された結果が出ている．しかし，その差異はそれほど大なるものではないことが見られた．なお，序盤は投票日19～20日前，終盤は投票日6～7日前のデータである．

34年については，序盤は全県行なっていないので（9県のみで行なった）．次のような推定方法で全体への推定値とした．

$$x'(i)=\sum_{j=1}^{19} w_j x_j(i)$$

$x_j(i)$: i 候補者の第 j 県での調査支持率

w_j : 第 j 県のウエイト(有権者数による比率)

$$x(i)=x'(i)/\sum^{19} w_j$$

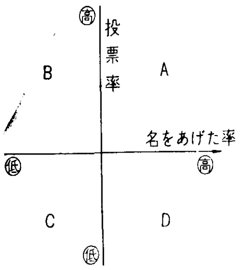
の $x(i)$ によって全国での調査支持率として画いてみた．これは，必ずしもよい方法ではないが，一応の比較のため，このようにしてみた．終盤での3次曲線（122人を用いた）

$$y=0.01636\,x^3-0.20542\,x^2+1.235\,x+0.0335(x,\ y\ \%)$$

と比較してみると，同様の様相が見られた．

(ロ) 県別にみた名をあげた率と投票率の関係を利用した県の分類について

次に考えられることは，県別にみたとき，名をあげ率の高いもの，低いものがあり，また投票率の高いもの，低いものがある．この関係が3次曲線に影響を与えるのではないかと思われることである．図式的に書いてみると，



参一参考図

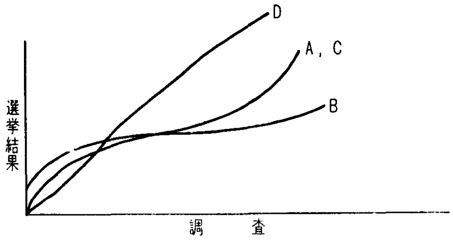
投票率高，名をあげ率高いグループA
投票率低，名をあげ率高いグループD
投票率高，名をあげ率低いグループB
投票率低，名をあげ率低いグループC

の4つにわかれる．このとき3次曲線の予想としては，右参考図のように考えられる．Dが立っているのは，投票率低く，名をあげ率が高いので，調査はよく実際をあらわす可能性が高く，Bがねているのは，投票率高く，名をあげ率が低いので，序盤的様相を帯びると考えられるからである．

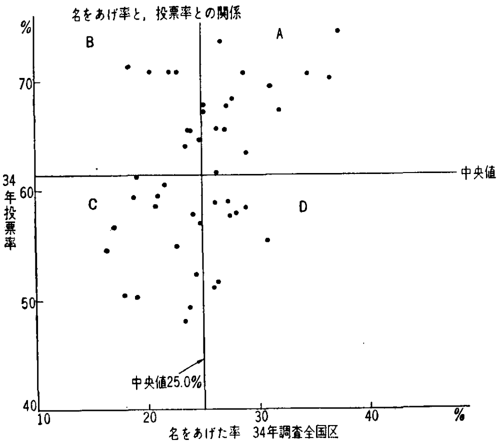
まず34年のデータでみよう．投票率は34年を用いてある．参—44図に示されている．A，B，C，D の4グループに分けることを考えた．分割点はそれぞれの中央値によった．すなわち，名をあげ率では25％以上と未満，投票率で，61.4％以上と未満とでわけた．A，B，C，D 別の3次曲線を求めてみた．これは，その i 県の調査支持率とその県での選挙得票率の関係を求めることを意味する．

3次曲線は参—45図をみる．これによってみると，予想が傾向的にみて全く裏書きされていることがわかった．このことは，甚だ興味があることである．

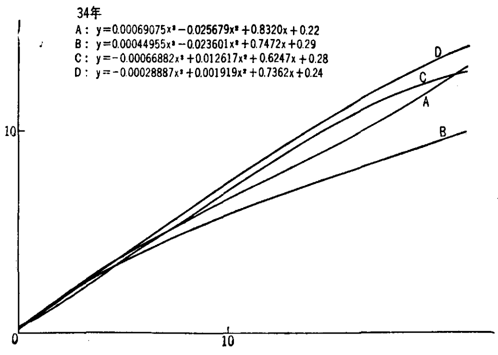
さて37年をみよう．まづ，県の分類から



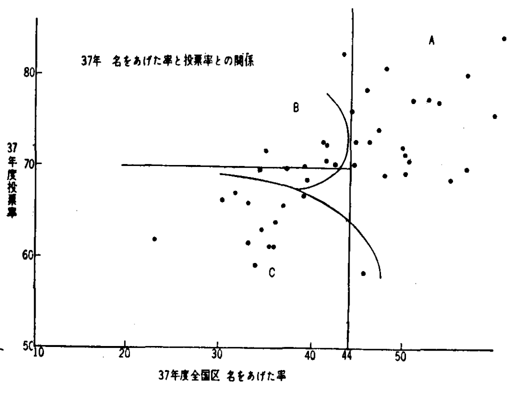
参一参考図



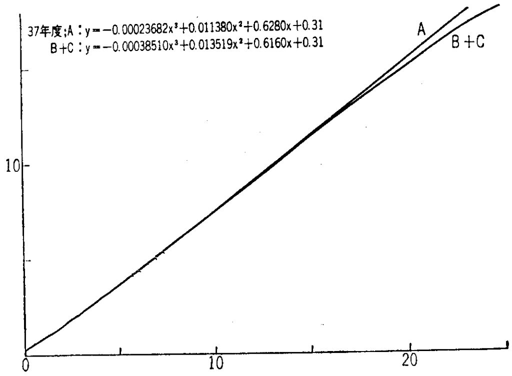
参—44図



参—45図



参—46図



参—47図

みよう．投票率，名をあげ率が異なるので分類は当然異なってくる．参—46図が分類のためのグラフである．同様に中央値でA，B，C，Dグループをわけてみた．ところが，このときは，県の分布がかたより，A，B，C，Dを形式的に保持すると，県の数が少なくなりすぎ，3次曲線が引けなくなるので，A，B，C，D 分類の意図を傾向的に掬んで，曲線でかこったものをグループとしてみた．このときはA，B，C 3グループとなった．BとCとが県別分類のとき近いので一緒にし，3次曲線を引き，またAの曲線を目盛ったのが参—47図である．これらを比較してみるとAとB＋Cの曲線はよく一致していることがわかる．なお，37年ではBグループの位置がCグループに近いので，Bグループの特色的曲線は得られなかったのである．

なお，34年と37年のAグループでの3次曲線の比較をするとき，37年の方が上方に偏倚していることがわかる．これは，さきに述べたように創価学会の候補者のためにおこるものである．

(ハ)　個人の得票率別にみた3次曲線について

個人の得票率によって3次曲線が異なるのではないかと考えられる．すなわち，強い候補者（選挙結果で，高い得票率を示すもの，事前にはこれがわからないので，全国につみあげられた調査支持率によって表現することになる）は，調査支持率の低い県でも選挙得票率はかなりあると考えられるし，弱い候補者の県別にみたときの調査支持率が低い場合は，本当に選挙得票率も調査のときと同様に低くなっていると考えられるからである．このようなことが，得票率のいろいろのところでおこると考えられるからである．すなわち，

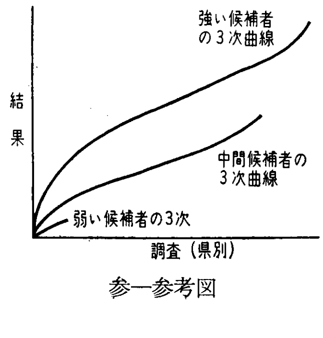
$y=f(x;y)+\varepsilon$，x は県別にみた調査支持率である．

しかし，y はわからないので，分析に際しては，x で代用し，

$$y=f_x(x)+\varepsilon$$

と考えるのである．

予想は右参考図のようになると考えられる．このように，3次曲線はその人の得票率に依存している．すなわち，各3次曲線の係数はそれとパラメーターにしているということができる．一人一人が異なった3次曲線をもつが，選挙得票率とその間には一定の関係が存在するということができる．これは3次曲線が $y(x)$ を介して連続的に変化すると考えることになる——これは3次曲線を考えるのではなく，さらに高次の非線型のものを想定することになる——．

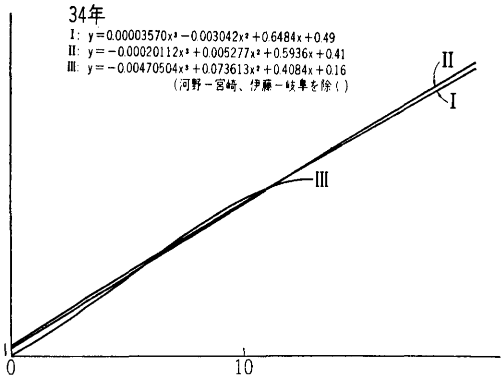


参—参考図

この傾向が実際にあるかどうかをみるために，つみあげられた調査支持率で候補者を分類し，その各グループごとに3次曲線を求めて傾向を検討した．実際のデータとして1人1人をあつかうことができないので，グループをつくり，大づかみにその傾向を検討することにした．これには34年データを用いた．グループの構成は次の通りである．

参—10表　調査支持率（全国つみあげ）

グループ	範　囲	候補者数
I	1.2%以上	36
II	0.4%—1.199%	37
III	0.045%—0.399%	27
IV	0.045%未満	22
計		122

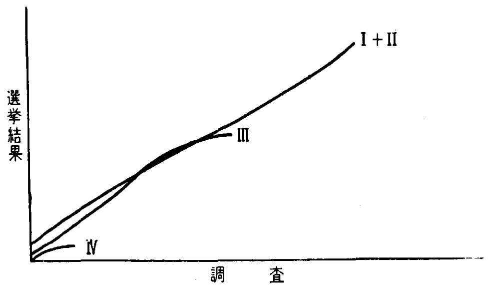


参—48図

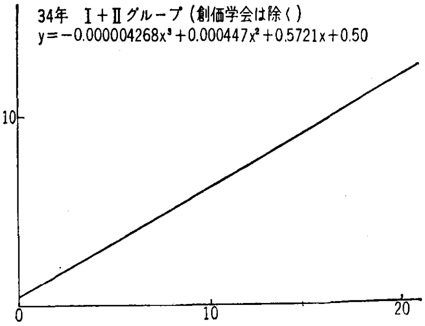
この4つのグループ別の県別調査支持率と選挙結果との関係をながめ，3次曲線を引いてみよう．なお，参—48図のⅢグループでは，全く特異の傾向を示す河野の宮崎のデータ，また，離れた点の伊藤の岐阜のデータは除外した．前掲の参—29図のⅣグループで，横山，伊藤(義)が除外されたが，分類基準の変更のためであった（あらためて計算するほどの値打ちは認められないので取止めた)．Ⅰ，Ⅱ グループの3次曲線はほとんどかわらない．Ⅲ グループでは，この候補者たちの最もよくあらわれる5％以下で下まわる曲線を示している．Ⅳグループでは Ⅲグループにくらべ，さらに下まわることがわかった．図式的に表現すると，下の参考図のようになる．なお参—48図曲線がとめてあるのは，そのグループに属するものでは，それ以上の調査支持率を示すものがないことをあらわしている．当選に関係するⅠ＋Ⅱグループで3次曲線が殆んどかわらないのは興味がある．参考のためⅠ，Ⅱグループをあわせた3次曲線を参—49図に示しておこう．なお，このとき創価学会をのぞいたものとする．これとさきのⅠ，Ⅱグループとを比較してみると除いたものの方の曲線が系統的に下に出ていることがはっきりわかる．このグループとⅢ，Ⅳグループとの曲線の差異ははじめに述べたように，弱い候補者のしかも調査支持率の低いところにおこっているのが，特色である．

次に37年でみよう．

このとき34年と同一グループ構成でみる．34年でⅠ，Ⅱ グループが同じとみなせたので



参—参考図



参—49図

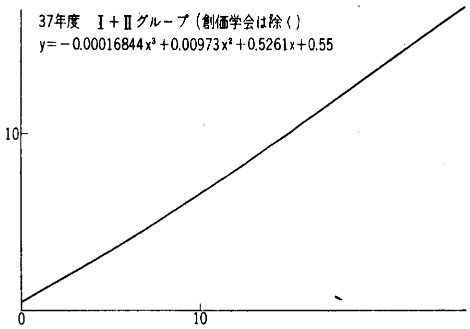
Ⅰ＋Ⅱグループで出してみた．なお，このときⅠ＋Ⅱ，Ⅲ，Ⅳグループ（すべて創価学会を除く）で全く34年と同様の傾向がみうけられた．Ⅰ，Ⅱ，Ⅲ，Ⅳ別の34年との比較は，創価学会の候補者が入っているために問題である．そこでⅠ＋Ⅱグループから創価学会の7人を除いた67人のもので3次曲線を引いたのが参—50図である．

参—11表

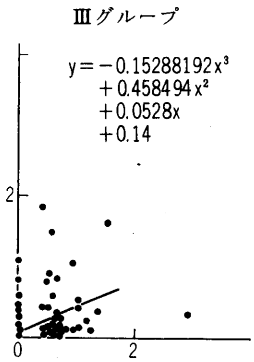
Ⅰ＋Ⅱグループ	74
Ⅲ　グループ	14
Ⅳ　グループ	19
計	107

これと前の参—49図とを比較してみればよいわけである．この結果は，(イ)で述べたように大ていの人が示す10％以下のところでよい一致を示している．まれにしか現われない10％以上でも実は1％前後なので，全体に与える影響は全く小さいものである．安定性に関しては(イ)と同様の結末を得たことになる．

次に，Ⅲグループの曲線を参—51図，Ⅳグループの曲線を参—30図に示す．通常多く現われる得票率が小さいところでは，Ⅲ，Ⅳグループでは，殆んど差が出ない．ただし，Ⅲグループについては，34年と37年とを比較すると傾向が異なる．34年ではかなりの調査支持率（12％程度のものあり）を示したものがいたが，37年では特殊のものを除いて3％以下であった．そこで全体的な比較はできないが，3％以下のところではかなり一致していることがわかる．



参—50図



参—51図

以上の分析を通して，得票率による候補者の分類による3次曲線を求めることは，さして重要ではないと考えられた．当選に関係あるⅠ，Ⅱグループでは差はないし，Ⅲ，Ⅳグループでは，得票率の少ないところの問題であるので，3次曲線の読み以後の数量化による補正にまかせて，全候補者をあつめて，近似的な推定力をもつものとして3次曲線をつくることにした．

以上，(イ)(ロ)(ハ)と検討を重ねたところ3次曲線の取扱いについて次のような結論を得た．

県別，候補者の得票別に3次曲線を引くことは望ましいが，データの数が減少し，安定性を欠くおそれがあるので，それらの功罪を考えてみなくてはならない．県別は37年のデータをみると，一本にしても差支えないし，候補者別得票率による分類は，0点附近が問題になること，弱い候補者のみが特に問題となることを考えると，爾後の数量化による補正の方法で覆うことができることを考えて，県別，候補者別の分類は3次曲線のときには用いないことにした．傾向をたしかめることで満足することにし，さらに，分析をこまかくするときの手掛りとしての知識とするに止めた．ただし，調査支持率が0.05％未満のものは別に扱い，別の3次曲線を考えることにしたのは前述の通りである．また，前述のように創価学会候補者は傾向が異なる——3次曲線が上方に偏位する——ので別に扱かうことにしたこと，3次曲線に一応安定性のあること，をここにあらためてつけ加えておこう．

第4部 地方区予測

以上のようにして，推定及び精度の表現ができたならば，これを用いて各候補者の当選確率を算出することになる．この方法について，まづ地方区から述べてみよう．

§1. 地方区，1人区

x から得票率を推定するのに z をつくる．前述のような立場から推定誤差の確率分布を考えることにする．すなわち調査で得られる x から，選挙結果をある分布をもって推定しようとすることになる．x を知るとき，前述したような方法で推定された平均 z，分散 σ_z をもつガウス分布をもって選挙結果が示されるものと考えることにするのである．

1人区では2人以上の候補者が必らず存在するので，そのいずれが勝つかを計算しなくてはならない．

いま，候補者が2人としよう．調査支持率は x_1，x_2 とする．これより第2部で述べた推定方式により推定得票率 z_1，z_2 を計算する．この x_1，x_2 よりする選挙得票率に対する推定は，前述の意味で確率変数となるので ⓩ$_1$，ⓩ$_2$ と表現する．これらはそれぞれ平均 z_1，z_2 をもち分散 σz_1，σz_2 をもつガウス分布とする．そこで ⓩ$_1>$ⓩ$_2$ となる確率を計算すれば，1なる候補者が勝つ確率を知ることができる．

ⓩ$_1-$ⓩ$_2=$ⓩ$_{12}$ とおくとき，z_1，z_2 による選挙得票率の推定としての誤差分布としてさきに示したガウス分布にしたがうので，この ⓩ$_{12}$ はガウス分布であり，平均は z_1-z_2 であり，その分散は $\sigma z_{12}^2=\sigma z_1^2+\sigma z_2^2-2\rho_{12}\sigma z_1\sigma z_2$ となる．この σz_{12}^2 を計算してみよう．

$$dz_{12}=dz_1-dz_2=g'(\bar{x}_1)dx_1-g'(\bar{x}_2)dx_2$$

$$\begin{aligned}\sigma z_{12}^2&=E\{(dz_{12})^2\}=g'(\bar{x}_1)^2\sigma_{x_1}^2+g'(\bar{x}_2)^2\sigma_{x_2}^2-2g'(\bar{x}_1)g(\bar{x}_2)E(dx_1dx_2)\\&=g'(\bar{x}_1)^2\sigma_{x_1}^2+g'(\bar{x}_2)^2\sigma_{x_2}^2+2g'(\bar{x}_1)g'(\bar{x}_2)\sqrt{\frac{\bar{x}_1}{1-\bar{x}_1}\frac{\bar{x}_2}{1-\bar{x}_2}}\sigma_{x_1}\sigma_{x_2}\end{aligned}$$

（ここに，$\sigma_{x_i}^2=2\dfrac{\bar{x}_i(1-\bar{x}_i)}{n}$，$n$ は標本数（われわれの場合，名をあげたものの数）前述のように，$g'(x)=f'(x)$ とみなすことにする．）

となり，σz_{12} は計算できる．なお，2人のときは $\rho_{12}=-1$ となる．

さて，ⓩ$_{12}$ の分布は $\dfrac{1}{\sqrt{2\pi}\,\sigma_{z_{12}}}\exp\left[-\dfrac{\{z_{12}-(z_1-z_2)\}^2}{2\sigma_{z_{12}}}\right]$ であるから，この ⓩ$_{12}$ の平均，$=z_1-z_2$，σz_{12} がわかると，ⓩ$_{12}>0$ の確率（すなわち，推定 z_1 による選挙得票率が，推定 z_2 によるものよりも大となる確率，つまり候補者1が当選する確率）は，

$$\frac{1}{\sqrt{2\pi}\sigma z_{12}}\int_0^\infty\exp\left[-\frac{\{z_{12}-(z_1-z_2)\}^2}{2\sigma z_{12}}\right]dz_{12}$$

$$=\frac{1}{\sqrt{2\pi}}\int_{-\frac{z_1-z_2}{\sigma z_{12}}}^\infty\exp(-t^2/2)\,dt$$

となる．ガウス分布の表から1が当選する確率を容易に計算できるのである．

2人の場合は，きわめて簡単に上述の計算をそのまま $\rho_{12}=-1$ とおいて実行すればよいのである．たとえば，$z_1=0.6$，$z_2=0.4, \sigma_{z_{12}}$ を 0.2 とすれば，$-\dfrac{z_1-z_2}{\sigma_{z_{12}}}=-\dfrac{0.6-0.4}{0.2}=1$ となり，

P_r(ⓩ$_1>$ⓩ$_2$)$=\dfrac{1}{\sqrt{2\pi}}\displaystyle\int_{-1}^\infty\exp(-t^2/2)dt$ として計算できる．

候補者が3人以上あるとする．そのとき推定値を z_1, z_2, z_3 とすれば，$\textcircled{Z}_1$, $\textcircled{Z}_2$, $\textcircled{Z}_3$ が得られる．この（$\textcircled{Z}_1$, $\textcircled{Z}_2$, $\textcircled{Z}_3$）はそれぞれ平均（z_1, z_2, z_3），分散・共分散マトリックス

$$\begin{pmatrix} \sigma {z_1}^2 & \sigma z_1 z_2 & \sigma z_1 z_3 \\ \sigma z_2 z_1 & \sigma {z_2}^2 & \sigma z_2 z_3 \\ \sigma z_3 z_1 & \sigma z_3 z_2 & \sigma {z_3}^2 \end{pmatrix}$$ をもつガウス分布をするものと見做すことが出来る．

ここに $\sigma z_i z_j$ は $-g'(x_i)g'(x_j)\sqrt{\frac{\bar{x}_i}{1-\bar{x}_i}\frac{\bar{x}_j}{1-\bar{x}_j}}\,\sigma_{x_i}\sigma_{x_j}$ とかける量であり，計算することができる．

ここで $\textcircled{Z}_1$ の当選する確率は，

$\textcircled{Z}_1 > \textcircled{Z}_2 > \textcircled{Z}_3$

か $\textcircled{Z}_1 > \textcircled{Z}_2 > \textcircled{Z}_3$

のいずれかがおこればよい．

このために，同時分布 $G(\textcircled{z}'_1, \textcircled{z}'_2, \textcircled{z}'_3)$ において，

$\textcircled{z}'_1 = u_1$

$\textcircled{z}'_1 - \textcircled{z}'_2 = u_2$

$\textcircled{z}'_1 - \textcircled{z}'_3 = u_3$

と変数変換し，

$$\iiint_R G(\textcircled{z}'_1, \textcircled{z}'_2, \textcircled{z}'_3)\, d\textcircled{z}'_1 d\textcircled{z}'_2 d\textcircled{z}'_3$$

R: $\{1 \geqq \textcircled{z}'_1 \geqq 0,\ \textcircled{z}'_1 - \textcircled{z}'_2 > 0$ で，かつ $\textcircled{z}'_1 - \textcircled{z}'_3 > 0\}$ なる領域

を計算すればよい．ヤコビアンは簡単で1となるが，計算は面倒であるので，簡便法として次のような立場をとる．

一般に候補者のうちの一人は弱いので（これを3としておく）$\textcircled{Z}_1 > \textcircled{Z}_2$ か $\textcircled{Z}_1 < \textcircled{Z}_2$ に焦点をしぼって議論を進めればよいと考えられる．

このときは2人の候補者のみをとりあげ，$\textcircled{Z}_1 > \textcircled{Z}_2$ か，あるいは $\textcircled{Z}_1 < \textcircled{Z}_2$ の確率を計算するのである．つまり3次元分布において $\textcircled{Z}_3$ を一応不問に付して，$\textcircled{Z}_1$, $\textcircled{Z}_2$ の2次元分布において $\textcircled{Z}_1$, $\textcircled{Z}_2$ いずれが勝つかの確率を計算することになるのである．この妥当性は，

$\textcircled{Z}_3 > \textcircled{Z}_1$ あるいは $\textcircled{Z}_3 > \textcircled{Z}_2$ の確率がどのくらいあるかによって定まるのである．一般にこの確率は十分小さく，1％以下のことがほとんどであるので，まず簡便法で十分用が足せると思われる．もし，この確率が，問題にするほどあるとすれば，$\textcircled{Z}_1$, $\textcircled{Z}_2$ の計算のみでなく，

$\textcircled{Z}_1 > \textcircled{Z}_2 > \textcircled{Z}_3$

$\textcircled{Z}_1 > \textcircled{Z}_3 > \textcircled{Z}_2$

の立場から，補正を加えることを考えればよいのである．つまり $\textcircled{Z}_1$ が $\textcircled{Z}_3$ に負ける——$\textcircled{Z}_1 > \textcircled{Z}_3$ であっても $\textcircled{Z}_3$ が当選する場合の——確率（当選確率）を考えに入れて，差し引けばよいのである．4人候補者がある場合でも同様である．さて，簡便法による場合，この計算を実行しやすくするために，z_1, z_2, $1/\sqrt{n}\,\sigma z_{12}$ の表を作っておけば便利である．n は標本数で，県ごとに変わるので，パラメーターにしておいた．求めた値に，その標本数の $\sqrt{n}$ を乗し，求めた $\frac{z_1 - z_2}{\sigma z_{12}}$ からガウス分布の表で $\frac{1}{\sqrt{2\pi}}\int_{-\frac{z_1-z_2}{\sigma z_{12}}}^{\infty} \exp(-t^2/2)\,dt$ の値を求めて，当選確率とするのである．

§2. 地方区2人区

このときの方法も，1人区と同様であるが，確率の計算の仕方が異なるのである．

候補者は3人としよう．$\textcircled{z}_1$, $\textcircled{z}_2$, $\textcircled{z}_3$ が得票率を示す確率変数である．

$\textcircled{z}_1>\textcircled{z}_2>\textcircled{z}_3\cdots\cdots a$

$\textcircled{z}_1>\textcircled{z}_3>\textcircled{z}_2\cdots\cdots b$

$\textcircled{z}_2>\textcircled{z}_1>\textcircled{z}_3\cdots\cdots c$

$\textcircled{z}_2>\textcircled{z}_3>\textcircled{z}_1\cdots\cdots d$

$\textcircled{z}_3>\textcircled{z}_1>\textcircled{z}_2\cdots\cdots e$

$\textcircled{z}_3>\textcircled{z}_2>\textcircled{z}_1\cdots\cdots f$

の $3!=6$ 通りのいずれかおこっているので，このうち，

1の当選の確率は		$a+b+c+e$
2	〃	$a+c+d+f$
3	〃	$b+d+e+f$

によって与えられるのである．

これを計算すればよい．4人以上あっても残りの一人は一般に弱ければ，この三者の争いに帰着される．

これを計算するのにも二つずつの組の積みあげで，計算しても大過はない（人数が多いと相関の程度はかなり小さくなっている）．すなわち，1の当選する確率（1は z_1 が第一位とする）をみよう．平均値は $z_1>z_2>z_3$ の順としておこう．1, 2, 3 の番号のつけ方は，どうでもよいので，この通りにしておいても一般性を失わない．

$\textcircled{z}_1>\textcircled{z}_3$ の確率を計算する．これは a, b, c を含んでいる．これに加うるに $\textcircled{z}_3>\textcircled{z}_1>\textcircled{z}_2$ (e) をもってすれば1の当選する確率が計算できる．これを P_1 とする．

3が当選する確率は $\textcircled{z}_3>\textcircled{z}_1$（$\textcircled{z}_2$ はどこにあってもよい），これに $\textcircled{z}_1>\textcircled{z}_3>\textcircled{z}_2$ をつけ加えればよい．この候補者3の当選する確率を P_3 とすれば，候補2の当選する確率は $2-P_1-P_3$ として計算できる．なお，候補者が3人以上になるとき，それぞれの相関係数が比較的小さくなる――たとえば，ほぼ3人の平均得票率が $1/3$ とするときでも，それぞれの間の相関係数は $-\sqrt{\frac{1/3}{1-1/3}\frac{1/3}{1-1/3}}=-0.5$ となる――ので，候補者の多いときは，$(\textcircled{z}_1\cdots\textcircled{z}_s)$――$S$ は候補者の総数――はそれぞれの次元で近似的に独立のガウス分布とみなして計算しても著しい誤りはないと見られる．われわれは $S=3$ を示したが $S\geqq4$ でも全く同様に当選確率を計算することができる．2人区のときは一般に3人の候補者までが一応せりあい，それ以上候補者があっても問題とならない場合が多いのであるから，3人までの考え方が根幹となると考えてよい．ここでも，z_1, z_2, $1/\sqrt{n}\sigma z_{12}$ の表をつくっておけば便利である．なお，ここで近似を用いるので，当選確率の総和が定員の2名になるような配慮が必要になることもある．

以下3人区以上でも同様であるが，以下では記述は省略する．

§3. 地方区3人区

考え方は，前述の場合と全く同様で，x_1, x_2, x_3, x_4 から推定値 z_1, z_2, z_3, z_4 を計算し，得票率の確率変数 $\textcircled{z}_1$, $\textcircled{z}_2$, $\textcircled{z}_3$, $\textcircled{z}_4$ の分布を求め，これから各候補者の当選確率を計算するのである．このとき $\textcircled{z}_1$, $\textcircled{z}_2$, $\textcircled{z}_3$, $\textcircled{z}_4$ に関し $4!=24$ 通りの順がある．

添字のみを書くと，参―12表のようになる．

1の当選する確率は□でかこんだ範囲を積分すればよい．(1, 2, 3, 4 は $z_1>z_2>z_3>z_4$ の順としても一般性を失なわない．)

$\textcircled{z}_1>\textcircled{z}_4$ の確率に，$\textcircled{z}_1<\textcircled{z}_4$ で1の当選する確率を出し，それらを加えてゆくという仕方で1の当選確率が計算できる．以下の候補者についても同様であり，計算は可能となる．

参—12表

1	2	3	4
1	2	4	3
1	3	2	4
1	3	4	2
1	4	2	3
1	4	3	2

2	1	3	4
2	1	4	3
2	3	1	4
2	4	1	3
2	3	4	1
2	4	3	1

3	1	2	4
3	1	4	2
3	4	1	2
3	2	1	4
3	2	4	1
3	4	2	1

4	1	2	3
4	1	3	2
4	2	1	3
4	3	1	2
4	2	3	1
4	3	2	1

すなわち，

$$1\text{の当選}\begin{cases}1\text{が}4\text{に勝つ}\\ 4\text{が}1\text{に勝ち，且}\begin{cases}1\text{が}3\text{に勝つ，あるいは}\\ 3\text{が}1\text{に勝ち，かつ}1\text{が}2\text{に勝つ}\end{cases}\end{cases}$$

$$2\text{の当選}\begin{cases}2\text{が}4\text{に勝つ}\\ 4\text{が}2\text{に勝ち，且}\begin{cases}2\text{が}1\text{に勝つ，あるいは}\\ 1\text{が}2\text{に勝ち，かつ}2\text{が}3\text{に勝つ}\end{cases}\end{cases}$$

$$4\text{の当選}\begin{cases}4\text{が}3\text{に勝つ}\\ 3\text{が}4\text{に勝ち，且}\begin{cases}4\text{が}1\text{に勝つ，あるいは}\\ 1\text{が}4\text{に勝ち，かつ}4\text{が}2\text{に勝つ}\end{cases}\end{cases}$$

3の当選は（3—上述の確率）となる．候補者の数が大になると，次第に $\textcircled{Z}_1$，$\textcircled{Z}_2$，$\textcircled{Z}_3$，$\textcircled{Z}_4$ の分布は独立的になってくるので，独立とみなしても，ますます安全となる．いま4人として，それらが競り合う場合を考えても相関係数は，

$$-\sqrt{\frac{1/4}{1-1/4}\,\frac{1/4}{1-1/4}}=-1/3=-0.33$$

となる．

なお，一応強い候補者は3人区のときは4人であるので上述の論議に尽きるが，もし5人以上あるとすれば，上の方法を拡張して用いればよい．

§4.　地方区4人区以上

全く3人区のときの考え方が適用される．3，4人区のときの早見表は，数が少ないので敢えて作らなかった．

§5.　地方区における県別各候補者の当選確率と党派別当選者数の予測

以上の方法に従って，各候補者別の当選確率を算出する．34年データにおいては，序盤，終盤とでは具体的予測数は大して変化はなく，個人候補者別にみても大した変化は認められなかったのは注目される．つまり，調査時期によってさして動かないことが確かめられたのは興味があった．

当選確率の党派別の分布をとると次のようになった．l 党の当選者数は，当選確率の和としてあらわされる．

$$\sum_{i_l}^{N_l} p_{i_l}(l)=m(l)$$

i_l は l 党の候補者，N_l は l 党の候補者数 $P_{i_l}(l)$ は l 党の i_l 候補者の当選確率，$m(l)$ は l 党の当選予測数となる．

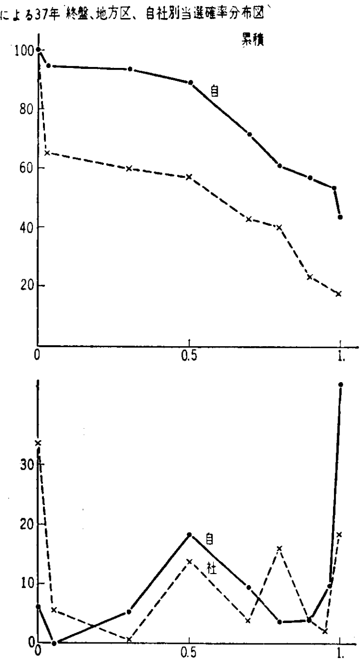
この $m(l)$ の分散は近似的に $\sum_{i_l}^{N_l} p_{i_l}(l)(1-p_{i_l}(l))=\sigma_{m(l)}^2$ で与えられるとみて大過はない．N_l が大であるから，また，県の数が大であるから $m(l)$ の分布は一般的にいって一応ガウス分布と見做せるので，信頼度95%として2×標準偏差で幅をつけておけば十分であろう．

34年データによる37年予測の当選確率の分布を示しておけば，参—52図となり，U字型とまでは行かず，中位が少しふくらむが，一応はU字型的となり，目的が達せられていることがわかる．また34年データによる37年の予測における当選確率と実際の当選者割合との関係を目盛ってみると，参—58図のようになって，両者が45°の直線の両側にかたまって，十分信頼性がみとめられる．さて，こうして党派別につみあげられた34年調査よりする37年の予測と実際の関係をみよう．参—13表のように，結果は十分満足すべきものである．なお，以上示したものは終盤のものであるが，大局的には，前述のように序盤・終盤ともに著しい変化はなかった．

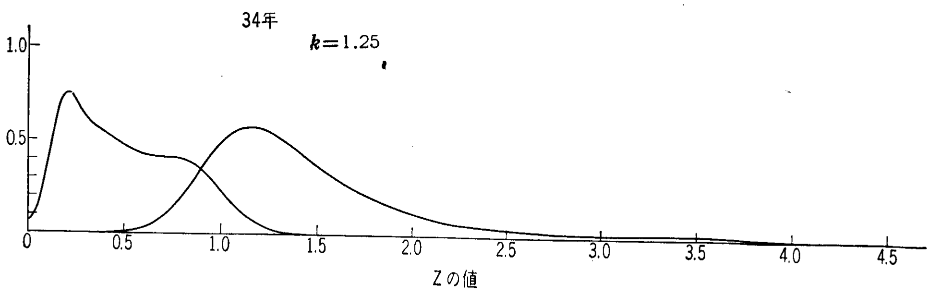
ここに，党派別当選者数は整数にしてあるが，これは，当選確率のつみあげの小数点のある数字を四捨五入を中心とした考えで整数とするのである——機械的にしたのでは，つみあげの総数が定員と一致しなくなるので考慮を要する——．この ±0.5 以内の差は，信頼幅をつけるとき，考慮に入れるのである．

参—13表

	自	社	民	同	共	諸無	創価
予測（当選確率のつみあげ）	47	23	1	0	1	2	2
95°の信頼幅	±4	±4	+1	0	−1	±1	−1
選挙結果	48	22	1	0	1	2	2



参—52図



参—53図

第5部　全国区の予測

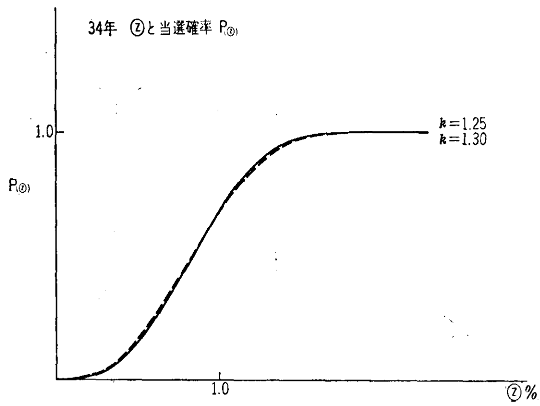
§　予測の方法

全国区においては，県別から積みあげられた調査結果からの推定値 z_i (i は候補者)を用いて

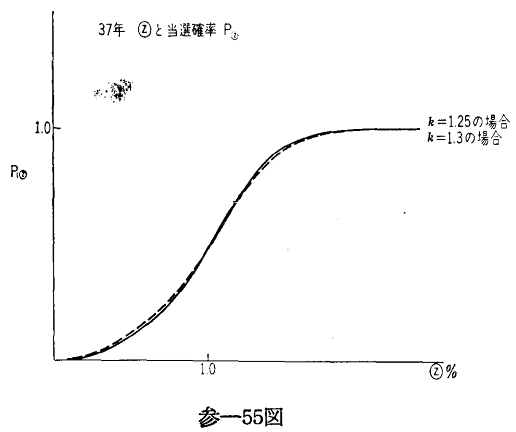
予測することになる．本来は，地方区のようにⓩの同時分布をつくり），これによって各候補の当選確率を算出すべきであるが，候補者数 n が大きいので，これを実行することはまず不可能である——ただし，n が大きく z_i が小さいので，相互の相関係数 $-\sqrt{\dfrac{\bar{x}_i}{1-\bar{x}_i}\dfrac{\bar{x}_j}{1-\bar{x}_j}}$ はきわめて小となり，候補者の選挙得票率に対する予測値 z の誤差分布は独立で $g_i(\zeta)$; 平均 z_i, 分散 $\sigma_{z_i}{}^2$ のガウス分布; と言う確率変数ⓩと近似的に見做すことが許容されることは便利なことである——．従って，この考え方に立って，当選確率を算出する簡便法を考えることが望ましいことになる．

x から推定値 z をつくり，その分散 $\sigma_z{}^2$ をつくるところは，第3部に述べた方法による．これが x よりする選挙得票率の推定分布となるのであるが，ここでまた見方をかへ，選挙得票率が z であるものが，調査による予測によって確率変数ⓩとしてあらわれると考え，このⓩの分布が上述のガウス分布をなすとするのである．従って，当選者群について，当選者群のもつ上述の確率分布曲線を積みあげて，当選者群のⓩの分布をつくり，一方，落選者群のⓩの分を積みあげ，落選者群の分布をつくるのである．しかし，誰が当選か落選かわからないので，また近似的に次のような考え方を採用するのである．この妥当性は，たびかさなる調査によって確かめられなければならないのは当然であるが，過去のデータが一応一つしかない（34年のみ）現在としては，最も妥当性あると考えられる方法によらざるを得ないのである．

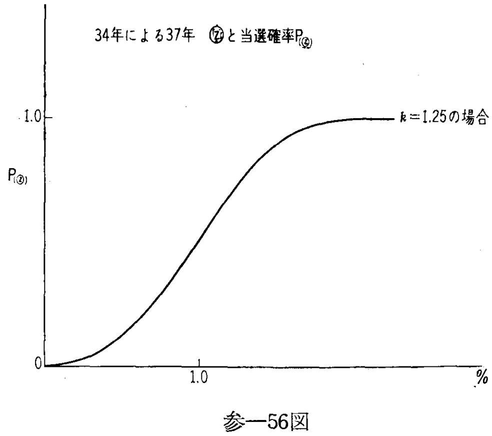
このため，調査から推定される z の値で，51番目，つまり当選定員以内のものの z を示す



参—54図



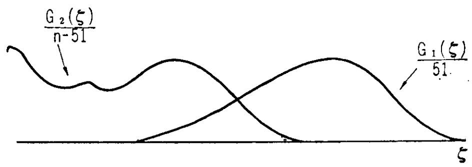
参—55図



参—56図

ものが一応当選者群をあらわすものであると見做し，――調査で推定されるその z をもつものが当選するというのではなく，その程度の z を選挙結果としてもつものが当選すると考えるのである．この考え方のズレは大きいとは思われないし，そのズレの影響は以下述べるように分布を考えるのであるから大なるものとはならない，因みに，順位と得票率の関係は推定と実際でよい一致を示していることは p. 67 §4 の終りに示してある――そのような z をもつものの分布を積みあげて，当選者群のもつⓩの分布（調査によって推定される z のもつ分布）を代表させるのである．この 51 人の z は当選者のそれではない（調査からの推定値にすぎないため）が，本当の選挙得票率はその附近にあることは確かであろう．また，たとえ，この推定された z を平均とするガウス分布をもつ確率変数ⓩをもって，個別的意味で，ある当選者の調査で示す比率の分布とみなすことは無理であったとしても，これらの51人を積みあげた場合，全体としてみるときは，これを当選者群の調査で示すⓩの分布（本当の当選者群の分布）と見做しうるということは首肯できる考え方であろう．

かくて $g_k(\zeta)$ を積みあげる．すなわち $G_1(\zeta)=\sum_{k=1}^{51} g_k(\zeta)$ をつくる．次に調査で 52 番目以下のものの $g_k(\zeta)$ を積みあげた分布 $G_2(\zeta)=\sum_{k=52}^{n} g_k(\zeta)$，$n$ は候補者数，をつくる．これはどういうことかというと $G_1/51$ は，当選者群がもつ選挙結果の示すⓩの分布をしめすことになる．$G_2/(n-51)$ は，落選者群のものがもつ選挙結果の示すⓩの分布を示すことになる．ある候補者の得票率が z と推定されたとしよう．



参一参考図

このとき，当選群にしてこの比率をもつものの，人数は $G_1(\zeta)\,(51\cdot\frac{G_1(\zeta)}{51}$ である）であり，落選群にして，この比率をもつものの人数は，$G_2(\zeta)\left((n-51)\cdot\frac{G_2(\zeta)}{n-51}\right)$ であるべきはずであるから，z のものべき当選確率は，$\frac{G_1(z)}{G_1(z)+G_2(z)}=p(z)$ であるべきはずである．こうすれば z と $p(z)$ との関係が見出されるのである．この関係の求め方は，選挙のたびに安定したものであるか――この z と $p(z)$ と関係式が常に同一であるというのではなく，こうして求めた $p(z)$ がいつも選挙結果において検証さるべきものであるという意味である――否かは，この調査と実際とをつきあわせて知られるべきものであるが，前回の調査の結果をみると，この考え方に妥当性があると予想された．

$p(z)$ と z との関係は，下に示すような関係がある．

これを実際の 34 年による 34 年のデータ，37年による 37 年のデータ，34年による 37 年の予測について求めてみた．この z, $p(z)$ の表を用い，z が調査でわかれば，当選確率 $p(z)$ を知ることができる．

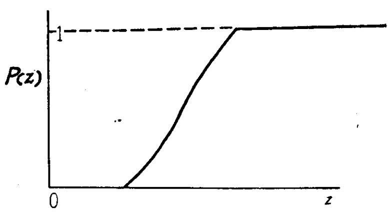
この $p(z)$ をつみあげて党派別当選者数を算出し，その分散を計算することになる．

l 党の予測当選者数を 全$m(l)$ とすると，

$$全m(l)=\sum p(z(l))$$

$$\sigma^2_{全m(l)} \doteqdot p(z(l))(1-p(z(l))$$

となる．（一応，独立と見做して大過はない）

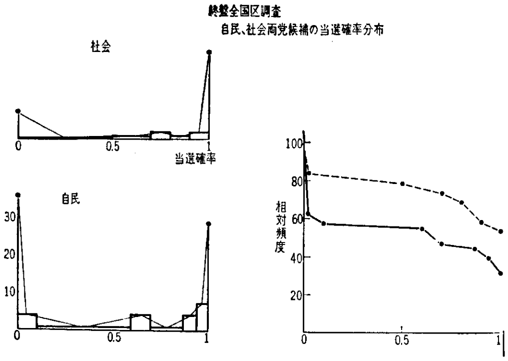


参一参考図

実際に，34 年のデータ分析にもとづく 37 年予測を行なってみよう．調査を行なって出した $p(z)$ の党派別分布，推定数，分散を終盤についてみると次のよう

になった.

まづ，当選確率の分布をみると，参—57図のようにほぼU字型に近いものが得られ，一応満足すべきものとなる当選確率の分布は，一般的にいえばガウス分布と見做してよいと考えられる．そこで当選者数及び幅（95% 信頼度）を出してみると（$k=1.25$）



参—57図

参—14表

	自	社	民	共	諸・無・同志会	創　価
推　定　数	23	12	2	2	5	7
誤　差　幅	± 4	± 3	+1	−1	±1	−1
37年選挙結果	21	15	3	2	3	7

のように，ほぼ満足すべき結果が出ている.

なお，参考のため34年データ分析による34年（3次曲線及びゲタが34年の分析によるもの）の予測，37年データ分析（3次曲線及びゲタが37年の分析によるもの）による37年の予測ということを分析のため行なってみると（$k=1.25$）

参—15表

		自	社	共	諸・無・緑	創　価	民　社
34年	予測	22	16	1	8	5	—
	結果	22	17	1	7	5	—
37年	予測	21	14	2	5	7	2
	結果	21	15	2	3	7	3

のようになり，当然のことながら34年による37年予測よりやや精度が上っている．これは当然のことであるが，方法を検討する上に必要なことである．$k=1.3$ としても変りはない.

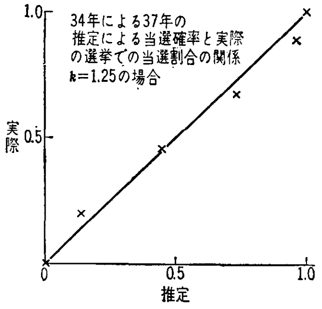
なお，党派別当選確率を出すとき，つみあげられた当選確率の和が丁度当選者数にならないのが普通である（ただし ±1 程度の差）．そこで，丁度当選者数になるように党派別当選確率の和に比例して按分するのである．また，推定数が，整数にならないのも普通である．そこで，四捨五入を中心とした考えによって整数とする——機械的に四捨五入したのでは当選者数になると限らない——のである．この四捨五入による ±0.5 までの差は，信頼幅をつくるときに考慮に入れるのである．なお，幅の $2\sigma_{全m(l)}$ もまた整数ではないので，幅をつけるのにも上述の考えにしたがい配慮する必要がある．ただし，これ以外の配慮——ガウス分布をしない場合，また候補者数に限度がある場合（候補者が，たとえば8人あり，そのうち7人が当選確率

1で残りの1人が0.5である場合などは 8(−1) とした)——によって幅を出すべきところも，少数候補者の政党（共産党，創価学会の場合）に関してはあるので注意されたい．

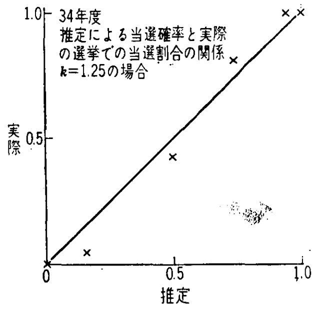
§2. 当選確率と当落割合

以上のようにして当選確率が求まったので，このようにしてつけた当選確率と当落の割合の一致度を検討してみることにしよう．これは非常に大切なことである．この当選確率の意味は，前に示した通り候補者 n 人中当選するものの確率平均は $n\times$(当選確率）であるということであるが，これが全体的に当選者割合と一致するだけでは十分ではなく，政党別にみても当選確率は全く同じ意味をもつ——当選確率は政党に無関係——ようになっていなくてはならないのである．

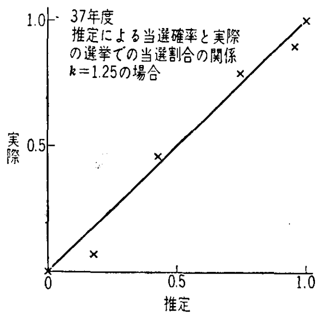
さてまづ，34年による34年データ，37年による37年データからみよう．この図は当選確率を 0, 0< ≦0.4, 0.4< ≦0.6, 0.6< ≦0.9, 0.9< <1, 1 の6区分に分類し，横軸にその区分にぞくしている当選確率の平均をとり，縦軸には当落割合を目盛るのである．すなわち，区分をそれぞれ $J_1, J_2, J_3, J_4, J_5, J_6$ とすると，横軸は $\bar{P}_{J_j}=\sum_{i\epsilon J_j}P_i/N_{J_j}$: P_i は i 候補者の当選確率，N_{J_j} は J_j 区分にぞくしている候補者総数：であり，縦軸は $Q_{J_j}=N'_{J_j}/N_{J_j}$: N'_{J_j} は J_j 区分にぞくしている当選者数：となる．これらが45°の直線上にのれば申分ないのである



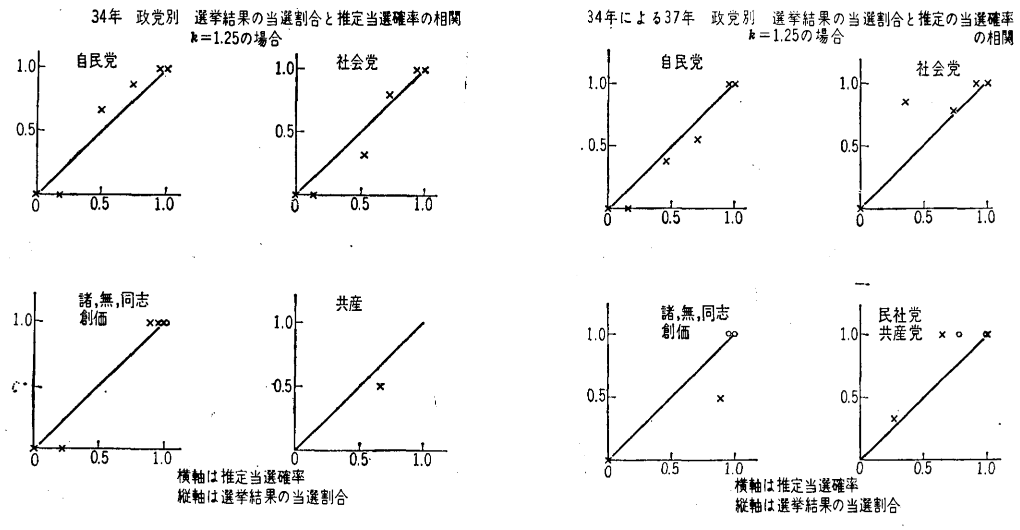
参—58図



参—59図



参—60図



参—61図

が．当然確率変動があるので，候補者数に応じて 45° の上下に変動するが，一方にかたよらない形が出ること，1及び0は 45° の直線上にのることが望ましいのである．

34年の全体は，参—59図（$k=1.25$）

37年の全体は，参—60図（$k=1.25$）

となり，満足すべき形となっている．$k=1.3$ の場合も同様である．

34年データの分析による37年の予測と実際を比較してみると，参—58図のようになり，34年による34年，37年による37年の平均的分析にくらべて悪くなっていない．十分満足すべき結果が出ている．

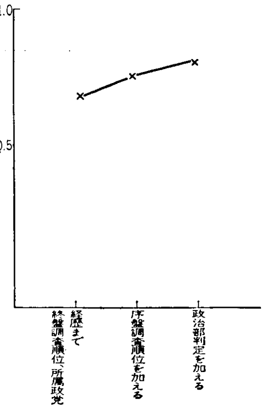
そこで政党別にこれをみよう（$k=1.25$ でみる．$k=1.3$ でも変りはない）．

煩雑にわたるので，34年による34年，34年データ分析に基く37年の予測と実際とのグラフを参—61図に示そう．両者とも自民党についてはまづ変化はなく満足すべき様相である．社会党の場合，当選確率の0.5以下でやや予測の場合には当選者が上回っていることが見られる(34年の平面的分析ではそうではない)．他の政党では数が少ないので何ともいえないが参考のためかかげておく．

§3.　当選確率の別の出し方

以上と異なった立場，衆議院選挙予測と同様の立場から予測を行なおうとするものである．データの完備している37年データ（107人）を分析の対象とする．各候補者は当落いづれかに属している．そして各候補者の要因として，終盤調査の推定順位(第3部 §4 による)，政党．経歴(新元現)，エキスパートの判定——朝日新聞社の政治部記者の各候補者による当選程度判定——及び序盤の調査支持率順位を用いるのである．これは朝日新聞社の近見晠道氏の発案になるものである．

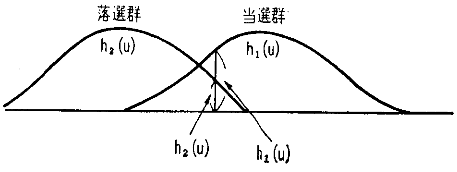
そこで外的基準が分類で与えられる場合，しかも外的基準が2分類（当落の2分類）の場合の数量化を用い，要因カテゴリーを数量化するのである．37年データについて行なってみると，弁別の効率をあらわす相関比は $\eta^2=0.8261(\eta=0.91)$ となり十分高い．各段階で η^2 がどう変るかをみると参—62図となる．



参—62図

主観的判定を加えても全体的にみた η^2 はあまり大にはならぬが，候補者個別にみるとき，要因相補い，平均化されて異様なものがなくなり，落ち着きが出てくる利点がある．

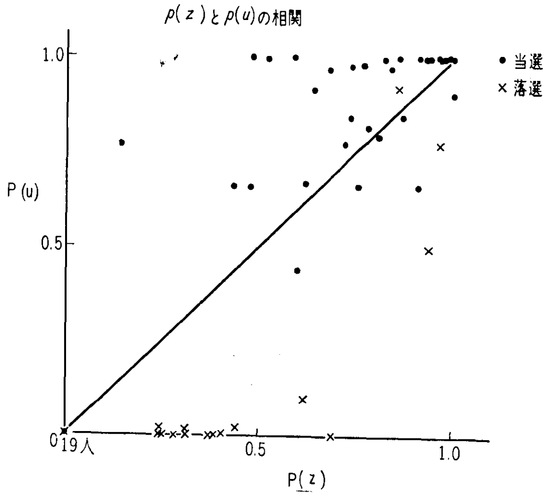
これらについて，各当落群分布図を前掲の要因中序盤をぬいたその他の要因を用いた場合について書いてみると二つにきれいにわかれ，予測の的中率は94％とみられ，十分注目に値する．こうして二つの分布図から，数量化によって求められる各候補者の数量 u における当選確率 $p(u)$ は，



参—参考図

$$\frac{M_1h_1(u)}{M_1h_1(u)+M_2h_2(u)}=p(u)$$

によって求められる．ここに M_1, M_2 はそれぞれ当選群，落選群の人数比率とする．$p(u)$ と u との関係から，各候補者の当選確率を出し，§2 の場合の当選確率 (37年による37年) と比較，また実際の当落比率と比較してみることは興味がある．これを求めてみると，



参—63図

のようになり，当落に関しては，主観的判定が加わったものが，より効率がよいことがわかる．左上半には・(当選) が多く，下半には×(落選) が多いのである．調査のみではなく，エキスパートの判定を入れることによって，推定の効率が上ることが見られる．以上は，平面的の分析であるが，今後の予測では，エキスパートの判断を加味することは，十分考えてよい問題である．

第6部　参議院議員党派別当選者数の予測

§1. 予　　測

第4部，第5部において，当選確率をつみあげて党派別当選者数を各個に出した．こんどは，これをつみあげて，参議院議員の党派別当選者数の予測を行う．ただ加え合わせばよいのであるが，夫々の予測のときは数字を整数にしているので丸めの誤差が入っている．これをそのまま加えたのでは丸めの誤差が大々的な損失をまねく．そこで，当選確率のつみあげをそのまま加えあわせ，その結果を丸めて整数にし，かつ定員数にあわせるようにすることが望ましいのである．

前に対応した記号を用いれば，地方区における l 党の当選者数は $m(l)$ である．これは l 党の候補者すべての当選確率をつみあげたものである．すなわち，

l 党の予測数は，

$$m(l)+{}^{全}m(l)={}^{T}m(l)$$

となる．この分散は

$$\sigma^2_{{}^{T}m(l)}=\sigma^2_{m(l)}+\sigma^2_{{}^{全}m(l)}$$

である．また ${}^{T}m(l)$ の分布は，$m(l)$, ${}^{全}m(l)$ の分布が概ねガウス分布と見做せるところからみて，これもガウス分布とみなせるので，信頼度95%で ${}^{T}m(l)$ に $\pm 2\sigma_{{}^{T}m(l)}$ の幅をつければよ

いと考えられる．但し，このとき，${}^{T}m(l)$ を整数にすること，全当選者数，すなわち $\sum_{l} {}^{T}m(l)$ を定数に等しくするために，${}^{T}m(l)$ の少数以下の誤差を配慮すること等と $\pm 2\sigma_{{}^{T}m(l)}$ の丸めを考えて幅をきめる必要がおこってくることをつけ加えておこう（このため，誤差が片側だけ例えば±1ではなく，－1とか＋1とかのものがおこることに注意されたい）．

§2. 検　討

予測調査を行なったとき，いちじるしく確率の小なるものが当選することが期待したより多かったり，選挙得票率が推定得票率より著しく大きくなったり，少なくなったり——当落においては問題はない．当選確率の問題においても同様問題はない——することがおこる．これを確率的変数であるといい切ることもできるが，これを検討しておけば情報が得られるかもしれない．そこで，37年度調査で主観的に「外れたと意識」されたものについて分析を加えてみた．まづ，調査について疑ってみる．

そこで，調査結果から推定されるべき県別3次曲線の値のサンプリング分散のみを示す曲線を書いてみて，3次の読みと選挙結果とを，著目した数候補者について県別に検討してみた．

サンプリングの許容範囲内にある点も少なくないのであるが，傾向的にみると選挙得票率が推定得票率より高かったものはずうっと全般的に調査よりも高く，推定得票率が高く出すぎたものは，調査よりもかなりひくいところが多い．下位にあって当選したものは得票率の小さいところでかなり傾向的に多めに出ているのが注目される．いずれにしても，特殊の県の所で，大きく外れたために，調査と選挙とのズレが出たというよりも，傾向的に動いていると見られのである．県別に調査の悪いところがあったというより，大勢として，ズレが来ていると見るべきであろう．これをどう解決すべきかは今後の問題である．概念的にいえば，調査の時期，調査の方法，かくれた要因の発掘といえよう．

第3編 知 事 選 挙

衆議院議員，参議院議員選挙に次いで問題になるのは知事選挙である．この予測はいかにすべきか．初期の都知事選挙予測の研究では「有権者が誰に入れるか，を予測する」という立場からされていたが，頭初に述べたように，これは選挙予測の研究としては本筋ではない．「誰に入れたか」は，明確にされるものではないからである——これらのことについては「数量化と予測——選挙予測を素材として」に詳述してある——．はっきりしていることは，投票したか，棄権したかである．これを個人について予測することは手がけることができるが，選挙予測としてはやはり本筋ではない．しかしながら，初期の研究ではデータが不足していたので，このような立場からせざるを得なかったのである．データが集まるにつれて予測の方法が見えてきたのである．その結果，参議院地方区1人区の方法と全く同一の方法でよいことがわかってきた．——但し一段と政党支持との結びつきが稀薄になる．ここで同じことを繰返すのは方法論的には意味がないので割愛する．手続きは，過去のデータの分析から，調査支持率と選挙得票率との関係——3次式を求め，これに d の推定値たるゲタをはかせて，推定得票率を算出する．この推定得票率のもつ予測誤差の分布型，平均，分散を算出し，これにもとづいて，候補者の当選確率を求めるのである．

統計数理研究所