

## ⑤ 質問調査法のもつ誤謬について

青山 博次郎

終戦以来我が国では急速に新しい統計的研究法が、種々の調査、研究に用いられるようになった。先ず医学、工学、農学を始めとして、心理教育、社会、経済の各方面に亘り新しい統計解析法が応用せられ、所謂推計学的分析の施されていないもの、ランダム・サンプリング法に基かないものは、眞の調査でも、研究でもないかのような観を呈するに至った。しかし乍ら推計学が凡ての調査対象の本質を解明するものであり、ランダム・サンプリング法の使用が調査を完全無欠にするものであるかの如き誤った考えが一部に蔓延していることは誠に残念なことといわれはならない。

我々統計数理を研究している者としては、既にこの推計学万能の行過ぎを警戒し、現実に立脚した研究と応用に苦心しつつあつたのである。このことについては既にこの記念講演会でも論じられたことがあり、詳しい説明を避け、後に一寸ふれるに止めようと思う。

然るに最近に至って社会統計学派といわれる人々の側からも、相当辛辣な批判が行われて来ており、その悉くが正鵠を得てはおらず、また特別なイデオロギーをもつて批判されているとしても、ここで取上げてみることも無意味ではないと思う。それで私は次の三つの面、即ち調査に対する批判、質問調査形式の利害得失、調査のもつ一般的な *bias*、について述べてみようと思う。

統計的方法はそれが自然現象を対象とする限りに於てあまり大した失敗もなく利用することが出来たのであるが、一たび之を社会現象に利用するとき單なる数理統計学或いは推計学を以てしては抜き得ない何かがあるような感がある。社会統計学派の人々のいう如く統計の対象を

誤差に解することは質成出来ないにしても、社会現象の測定という仕事は自然現象と全く本質的に異なる困難に直面するのである。(調査全体については勿論両者の取扱いに差がある訳ではない。)

我々は社会調査としての在論調査にこの例をとって考えてみよう。既に古くなってしまった感があるが、米国の大統領選挙の予想がその忘れ得ぬ一例であらう。これは1948年の米国大統領選挙に於て Gallup, Roper, Crossley 等の各選挙予想が悉く外れたことで、米国は勿論、サンプリング調査による在論調査の將に行われつつあった我国の識者をも驚かせたものであった。嘗て1936年リテラリ・ダイジェスト誌が大統領選挙の予想に失敗して、それに代る Gallup の American Institute of Public Opinion が一躍名を成してより、12年を経て、皮肉にも今度は Gallup 自身が劇的な失敗をしたのであった。これに対して在論調査、特に選挙予想は到底不可能なものであるという考えぬ、サンプリング調査は役に立たないというような議論まで出たことは未だ記憶に新しいことと思う。このため米国では Social Science Research Council (社会科学調査委員会, S.S. Wilks を委員長とする) が結成され、予想失敗についての科学的究明が行われたのである。その結果は次のような勧告に示されている。

1. トルーマン、デューイの両者は接戦であったので在論調査ではどのような事前の情報をもとどちらが勝者が予想することは不可能であった。

サンプル数を2000としてランダム・サンプルと考えると  $\sigma_p = 1.1\%$  位だから  $3\sigma_p$  の中では有意差はないといえない。しかし Gallup

	Dewey	Truman	Thurmond	Wallace	その他
実際の得票%	45.1 (45.4)	49.5 (49.8)	2.4 (2.4)	2.4 (2.4)	0.6
Gallupの予想	49.5	44.5	2.0	4.0	—

( )内はその他を除く場合の%

ほどあらかが当選するといわざるを得ない立場にあつたのである。

2. サンプルリングが所謂 *random sampling* ではなく *quota* 法 (州, 性別, 都市農村, 年齢, 収入, 政党関係などを *Control* とする) であつて, 教育程度の高いものをサンプルにとり易く, 客観的な誤差が推定できない。
3. 面接上の誤りがあつたと考えられる。即ち従来から *republican* に対する支持が3%内外多きい目にてる偏りがあつた。
4. 浮動票の行方と選挙間際の流票の測定を怠つた。実際 *Gallup* は選挙当日の10~12日前位で調査を中止しており, その後になつて *Treman* に流票が集つたことが分つた。当時未決定者の割り振り (投票しないと考えていた), 最後の流票, 投票するか, しないかの決定にきめ手になつたのである。
5. 勧告として *random sampling (Area sampling)* を用い, 面接技術を高めること, その他心理学的行爲, 政治的行爲に対する基本的な研究を必要とする。また選挙前の世論調査の分析呈示, 発表の方法が調査結果の誤解を招き, 予想失敗が大きい反響を与えて了。それ故結果の解釈, 信頼度の限取を示さねばならない。

以上のような批判と勧告が行われ, 1950年における中間選挙では *Gallup* は *Area sampling* を用い選挙の3日前まで調査し, 電報による報告, 集計を行った。また未決定者のわりふり方を政党に対する傾向と投票しような行動とを集約的に調査して推定の精度を増したのである。

このようにして米国では *random sampling* の世論調査への採用, 面接技術の改良, 研究などが行われるよつになつたが, 我が国の状態はどうであらうか。

最近雑誌「自然」に世論調査の魔術, 失業統計の手品, C P I は正しいか, 危機における世論, など一連の調査批判が載せられている。これ

らを材料として調査のもつ誤謬について融れてみよう。先ず昨年度に行われた2, 3の新聞の傾向を概観してみよう。(次表参照)

批判されている論点は次のようである。

1. 調査の主体の問題である。現在在論調査の専門機関中30前後が実質的に活動しており、大新聞社が多くの在論調査を行っている現状である。

	回数	地域		項目数					回収率 記入	面接 記入	原簿の 判別法	誤差 記入
		全国	地方	政治	社会	経済 (生活)	娯楽	計				
朝日	5	4	1	5	3	1		9	5	2	1	
毎日	6	5	1	3		2	2	7	4			
読売	9	3	6	4	2	2	2	10	8	2	1	
東京	6		6	3	3	2		8	2	6	1	1
計	26	12	14	15	8	7	4	34	19	10	3	1

その思想的立場、調査の目的、組織、資金関係などが問題になっている。「思想的に無」という立場で調査することは科学的調査たる限り誠に大切なことであるが、その色合いは正確なる調査情報の提供以外批判の仕様のない水掛論になってしまう。我々は調査の失敗の記録を書くだけの勇気も欲しいもので、最初の仮説と異なる結果を得ても厳に我田引水は慎しまねばならないと思う。

2. 面接法における問題が取上げられており一度に多くの項目を調査したり、面接時に同席者いたり(1950.11. 国立在論調査所、「郵便事業に関する在論調査」では同席者なし、55%、一時的同席14.7%、終始同席29.4%)、調査票を置いてきて誰が書いたか分らぬような調査票を回収したり、また調査員の社会的構成が問題として取上げられている。

これについては各調査機関が専属の面接員をもたず、訓練すること  
ができぬ所に向題があるが、調査員の社会的構成のみをもつて結果を云に  
することは常識的抽象的すぎるように思える。実験してみての上の話な  
らよいがたと影響があるたろうとだけでは科学的な批判とはいえない。  
我が国ではこの種の研究として朝日新聞社の堀川直義氏の「インタビュ  
ーの研究」がある。これは主として新聞記者について心理学的実験を行  
い、面接に際し経験年数、発問形式等かどのように影響するかが研究さ  
れている。米国の例では面接員のもつ意見によつて回答者の意見が影響  
をうけることがあり、面接員の性、年齢、収入の差による影響はみられ  
ず、*socioeconomic status* により差のみられる場合も、差のみ  
られない場合もあつたといわれている。また面接の経験による差もあま  
りないといわれている。このような面接員の構成は訓練により、或いは  
対象を *split* することによつて或程度修正することができる。

また調査の *questionnaire* の問題が重要である。面接員の判断  
を要する問題では若干の差を生ずるか、具体的な自動車の有無、電話の  
有無などでは同一回答者に対する面接員の二度の調査、異なる面接員の  
同一回答者との面接に於て殆んど差はでないといわれている（これらの  
数字例もあるが *random sample* かどうか不明のため例示し  
ない。）これらの結果は米國に於ける例であつて人種、環境の異なる我  
國でも同様かどうかは飽々まで実験的研究に俟たねはならないであらう。

次に面接員の虚偽 *cheating* の問題である。実際に面接を行わな  
いで回答を自分で作り上げたり、サンプルを指定のものから勝手に変更  
したりするものがそれである。我國のように時間的にも、経済的にも無理  
な調査をやらせるときにこのようなことが多い。我々の経験では学生を  
用いたときノ割位のもものが虚偽報告をしていることが分つた。

批判には融れられていないもので回答者自身のごまかし、心理的変動  
などが結果に *bias* を与える原因となることに注意しておこう。

3. 次に *random sampling* に対する批判がでている。本

論から若干は離れる問題かも知れぬが重要なことであるから取上げてみよう。

推計学が万能的一般方法学であるといわれ、社会現象の分野に应用する場合にも盲目的にこれが使用されているというのである。確かに「標本と母集団ということを区別し得た」ことは推計学の功績であつたが、その母集団を明確に把握し得なかつたこと、種々の仮定を忘却した行過ぎは推計学の失敗であつたともいえる。我々の問題とするのは母集団の前提たる調査対象 (*universe*) であつて、そのパラメーターを平均という形で把握するため、確率の場を導入した母集団 (*population*) を考えるのである。結果は母集団についてではなく調査対象について論じたいのであつて、正規母集団を仮定する検定を用いたり推定をしたりするだけでは母集団の範囲を出ない議論にすぎない。信頼性も妥当性をしらべ、調査対象に対して結論が正鵠を遠していないかどうかを予測という絆を通して追求して行くものでなくてはならないのであつて、これを統計数理は重大な項目として取上げているのである。

また *random* に抽出することは概念的な操作上は可能であつても、面接調査という手段を用いて測定するとき、サンプル相互は調査員を通じて相関々係をもつて成らうし、現実には独立性の条件を完全に満足することはまず望み得ないのである。これがまた一つの *bias* の原因となるのである。

次に層別ということについて「種々の調査では調査内容の如何に拘せず形式的に人口、工業率、商業率、ラジオ普及率、政党支持率を規準として採用していること、層別が任意抽出の科学的準備であつて結果的にみて全数調査の結論である標識一般と近似的な値をとることを目的とする限り、それぞれ異質的なものを厳密に区別しなければならぬのであるから無作為性をいれる余地は殆んどなくなるといった方が正確であらう」という批判がのべられている。これ

は層別の意味の誤解であつて、我々が調査項目（標識）に關係の深い規準、これは過去の調査から分つておればそれをとるにこしたことはないが、これを用いるとき母集団パラメータの推定の精度が上がるというのであつて該批判は層別的手段と本質を混同したものである。（*Quota* と *stratification* の差がはっきりしてないためである）。

層別の失敗があつても精度は層別しないときより悪くなることは殆んどないのである。また層別して了えはランダム性をいれる余地がないというのは調査項目たる標識自身を以て層別して了えはそうかも知れぬが、それなら調査は不要であり、現実にはそのようなことは不可能なことなのである。ただ多くの調査が層別による効果について論じていないため、層別が本質であると誤解して了つたといえよう。

4. 次に質問法と分析法について批判されている。これは主体が意識的、無意識的に回答者を誘導し、質問の過程で回答者を教育しているというのである。これは確かにそうである。例えば、『日本も講和条約ができて独立国になつたのだから自分の力で自分の国を守るために軍隊を作らねばならぬ』という意見があります。あなたはこの意見に賛成しますか、反対しますか。（1951/9/20朝日）—— 賛成 71%、反対 16%、DK 13%。「講和後日本は自衛のための軍隊をもつべきだといわれていますが、あなたはこれに賛成ですか、反対ですか」（1951/10/4読売）—— 賛成 56.0%、反対 24.9%、DK 17.1%。これらは何れも *bias question* といわれるもので賛成側にひき入れようとするには一寸した言い廻して%の開き（有意差あり）が出てくるのである。これについては昨年度輿論科学協会による実験的研究が行われている。

しかしこの反面に於て政党の支持率などは同時期において種々の新聞の%は殆んど同一で有意差を示さない。即ち質問の項目と形式

によっては歪んだ結果を生ずることがあるのに注意せねばならないのである。

またある事柄を知っているかどうかを聞く様な場合は必ず程度をしらべなければならぬ。単なる言葉だけの知る知らぬということは実態と全く関係のない場合さえあり得るのである。このことは二項選択法のもつ欠点であり、後にまた述べるであろう。

例えば、1950.3の国立社会論調査所「青少年不良化防止に関する社会論調査」に「児童委員、児童福祉司または少年保護委員というものを知っているか」というのに次のような結果がでてゐる。

	テスト質問をしたとき	テスト質問をしないうとき
知っている	17.5%	63.1%
聞いたことはあるが詳しいことは知らぬ	58.3%	10.7%
分らない	24.2%	26.2%

これらの質問法の問題は推計学の全く関知せぬ領域に属することは明らかである。

また「全有権者の意見の縮図がでてくることが保証されている」と称して対象中得られたサンプルの性別、年令別、職業別、教育程度などの *control* だけについてのべることは母集団の構造を説明する根拠にならないと述べられている。これは母集団とサンプルの相違(ランダム性が確保されたかどうか)を調べているのであつて、構造そのものの説明ではない。通常述べられていない所の無回答者についてもその構成をしらべ、その一部を更に、*random sampling* して *bias* の程度を調べなければならぬ。そして始めて歪みのない構造が推定できる。唯新聞などではそのような、*space* もないといった方が当る



ているかも知れないが、調査井関の発表に際してはこのことに触れられぬは価値はない。而して調査項目と諸種の *Control* との関係は分析の後始めて構造化されるのである。

次に D 区グループの取扱いが不十分だということである。我が國の各種の調査では女子、それも小学校卒業程度の高年令の者は大抵 D 区グループに属し、この取扱い方によって種々の都合のよい解釈をつける余地が生ずる。それ故 *intensive survey* を行って十分その動向を把握することが必要である。

尚多項分布の推定に関し、信濃市のつけ方についての批判は理論的となるので避けるが、その考え方は誤解である。我々は調査して了うと得られた数字に心を惹かれ、兎角溺れる着け藁をも掴むの譬の如く正当なる評価を誤ることが多い。サンプル数の小なることを忘れて有意差のある採に考えたり、相関のあることを忘れて *data* の比較をしたりする採なことがそれである。失敗は失敗として認めてこそ進歩があり、問題を後に残してこそ前進があるのである。

以上は面接調査に関する批判についてであつた。この他労働力調査、消費者価格調査についても論じられているがこゝでは取上げる余裕がないので省略する。概括的にいうならば一般的に行われているこの種の調査に於ては定義の曖昧なこと、調査対象の枠の設定の落失、層別確率比例抽出の第一次抽出単位の数不足（これは層別の上すべりなことにも原因する）、ランダム性の誤解などがあげられる。

以上の如き面接調査に対応するものとして郵便調査による場合も多く、誤謬の入りこむ余地がある。これと原理は同一であるから客観的テストと呼ばれるものを例にとつてみよう。戦後の一傾向として教育方面でも、生徒を各方面から評価するのに種々の客観的テスト、所謂 O X 式テストが行われている。東京都に於ても高校進学のためにアチーブが行われているが、採算が客観的に行えるようにというので再認型や再生型のテスト形式が多分に用いられている。これによって我々は一体何を評価

しているのであろうか。真に生徒の学力が分るのであろうか。これらについて研究は米国における結果の追従に甘んじていて未だ十分行われていないとは云い難い状態である。多項選択法を用いるのがよいか、二項選択法でもよいか、昔のような論文型即ち自由回答式がよいか、どの方法が学力（但しそれが具体的に定義し得るとして）と一番高い相関をもつ方法かなど枚挙にいとまのない程の問題を含んでいる。これらは一学校、一学校の如く実験室的の研究も必要であるが、更に一般的な結論を出すために広範囲の対象についての研究を必要とする。これに關係して我々の調査して得た結果の一部を紹介することにしよう。

取上げた問題は質問の形式のもつ利害得失である。このことは先にも述べた輿論科学協会でも先年東京都で共産党についての問題で研究されたことがあるが、我々は一足とびに面談調査に行かずに、集合調査という形式で具体的には東京都の中学2年生全体を対象として基礎的な研究調査を行ったのである。

一般に質問調査法における技術はその媒体（郵便、面接、電話等の media）、形式（二項選択、多項選択、順序、言葉づかい等）の二つの面から考えねばならない、ここでは形式についてののみ考えてみる。形式としてはよく使われるものとして自由回答法、多項選択法、高等法、尺度法などがある。これらの利害得失は次のようであると称されている。

1. 自由回答法は回答者が自分の意見を自由に述べる形式であつて、テストでいうと論文型というものである。回答者の個性的な意見がよく分り、事例調査に用いられる。多項選択の分岐をしらべるのに用いる。明確に意見をのべられぬ人、教養のない人、億劫がる人には不適當で *After coding* するため客観性も損われる。テストでは同じ答案に〇真から満真までつくというような欠点がある。
2. 二項選択法は賛成、反対を問いたり、〇×をつけさせたりするもので態度が未分化のときも確定的な判断を求めることができるが、態度が不明確で一方的に意見をおしつけることとなる。

3. 多項選択法はいくつかの回答を用意してこれの何れかをえらばせるのである。回答は割合に容易であるが、選択肢の作成が困難であり、幾つかのカテゴリーに自分の意見を当てはめるために無理を生じ、教養程度の低い人には「分らない」という回答が多くなる。テストではよい加減なでたらめ回答が生じる。

4. 出題法としては重要なものから順位をつけさせるものがあるが、果して順位が明確に回答者の意識にあるかどうか疑わしいものがある。

大体以上のようなことが知られてはいるが調査項目に対し、対象によりのように変わるかは数字的に明瞭に示されたものがない。

我々のとりあげた形式と材料は次の通りである。中学2年生を対象としたため、数学と社会科より題材を選んだ。問題は大きく3部に分け、第1部は数学を材料とし再認型の多項選択、二項選択、再生型の完成法を比較するとし、第2部では社会科の材料を用い順位づけ (ranking)、一番上位のもの (best one choice)、幾つでもえらぶ (multiple choice) という3方法及び選択肢の与え方の順序、問題挿入の位置による差などをみることにした。第3部では道徳的判断の問題を用いて生徒の態度を測定し、bias をかけた効果、強度の問題などをしらべることにした。

さて第1部は代数の問題を3問(各々小問5を含む)提出し、更に一問だけ自由回答による問題を対照に用いた。いま完成法をA形式、多項選択法をB形式、二項選択法をC形式(この中訂正させるものをD形式)と呼ぶことにして次表のような6種類の問題を split し6群の生徒に与えた。例えば第1問のA形式では「 $5a + 3a = \square$ 」の如き5問より成り、この  $\square$  の中に答を記入し、B形式は「 $5a + 3a = \{8 + 2a, 8a, 8a^2, 16a\}$ 」の中の正しい答を○でかこみ、C形式は「 $5a + 3a = 8 + 2a$  ( )」の( )内に、正しいときは○印、誤っているときは×印をつけて正しい答をその右に記入するようにし

た。(詳細は統計数理研究所：統計数理研究報 No. 9 参照)これらの6群の生徒については田中B式知能テストを行い、各群の知能度に差がないこと、free answer の得点にも差がないことを以て等

生徒群		問題群		
		I	II	III
第1群	1	A	B	C
第2群	2	B	A	C
第3群	3	C	A	B
第4群	4	A	C	B
第5群	5	B	C	A
第6群	6	C	B	A

質化されたものと考えたことにしたのである。

先ず7枚で、test と retest を行ったので、その結果からのへよう。これは信頼度に関係する。各問各小問各形式毎の安定度を次表の

第1表

{ test-retest の反応カテゴリー表 }

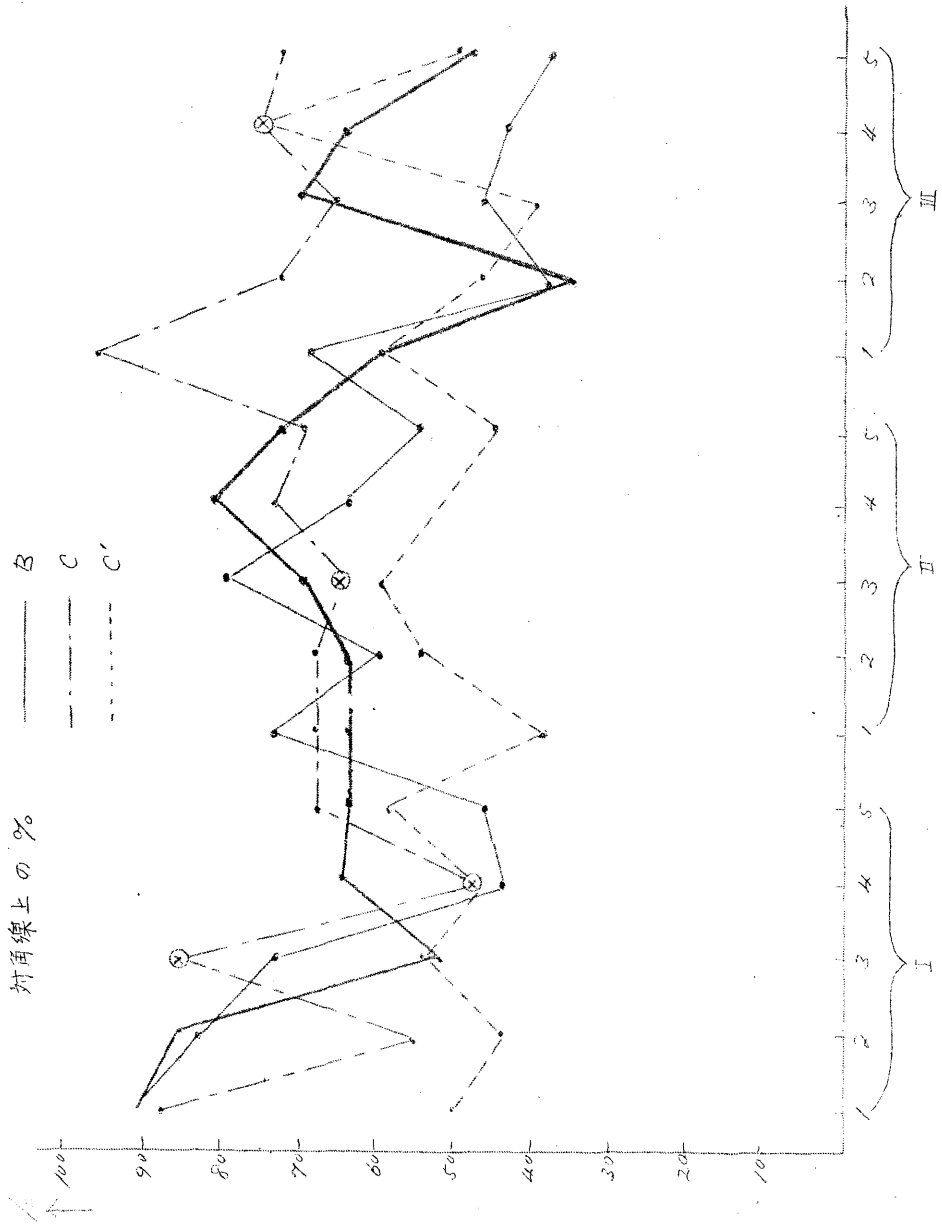
test retest	1	2	3	4	計
1	$f_{11}$	$f_{21}$	$f_{31}$	$f_{41}$	$f_{.1}$
2	$f_{12}$	$f_{22}$	$f_{32}$	$f_{42}$	$f_{.2}$
3	$f_{13}$	$f_{23}$	$f_{33}$	$f_{43}$	$f_{.3}$
4	$f_{14}$	$f_{24}$	$f_{34}$	$f_{44}$	$f_{.4}$
計	$f_{1.}$	$f_{2.}$	$f_{3.}$	$f_{4.}$	$N$

対角線上にのっている%で表わすことにすると第1図のようになる。全

第 1 図

test - retest  
 対再線上の %

— A  
 — B  
 - - - C  
 - - - C'



体を平均すると

$$C > A > B > C'$$

$$71.8(\%) \quad 66.1 \quad 60.6 \quad 51.9$$

C' は C 形式で訂正させたものについて採集したものである。この数字のまゝで見ると確かに C 形式は易観性をもつように思える。しかし答小問をノックとして採集したときの、*test* の得点と、*retest* の得点との相関係数は第 2 表の値になる。(サンプル数は 30~34) 即ち一般的には

$$r_A > r_B > r_C$$

の如き傾向をもち前とは反対の結果を示す。その理由は次のように考えればよいであろう。

でために回答す

るものとしてみると C 形式では 2 回とも同じ答をする確率は  $\frac{1}{2}$ 、B 形式 (択肢肢は 4 つ) では  $\frac{1}{4}$ 、A 形式では  $\frac{1}{5}$  (B 形式の択肢肢と「その他」) と考えられる。従つて C は 50%、B は 25%、A は 20% はでために回答をするとしても同じ回答をする割合である。これと実際の % を比べてみて各形式の判定効率を次式で求める:

$$e_C = \frac{71.8 - 50}{50} = 0.436$$

$$e_B = \frac{60.6 - 25}{25} = 1.424 \quad (A) > (B) > C$$

$$e_A = \frac{66.1 - 20}{20} = 2.305$$

従つて実際には A 形式の方が安定度が高いので相関係数による結果とも

よく一致する。しかし勿論このことはサンプル数が少いか一つの問をのべたので今後の研究に俟たねはならない。

何れにせよ *retest* によっても安定度が高いと考えられるので信頼性はあるものと考え本調査の結果を次にのべることにしよう。

サンプルは全体で 1008 であるが各群の *free* の得点、知能点の得点は第 3, 4 表のようであった。 *free* の得点は第 4, 5 群に差があるように見えるが 1% 水準では差はない。知能点は何れも差がない。

第 1 表 [ *free* 得点 ]

群	$\bar{x}$	$s$
1	1.149	1.981
2	1.190	1.993
3	1.095	1.943
4	1.363	2.071
5	0.994	1.885
6	0.911	1.861

第 4 表 [ 知 能 点 ]

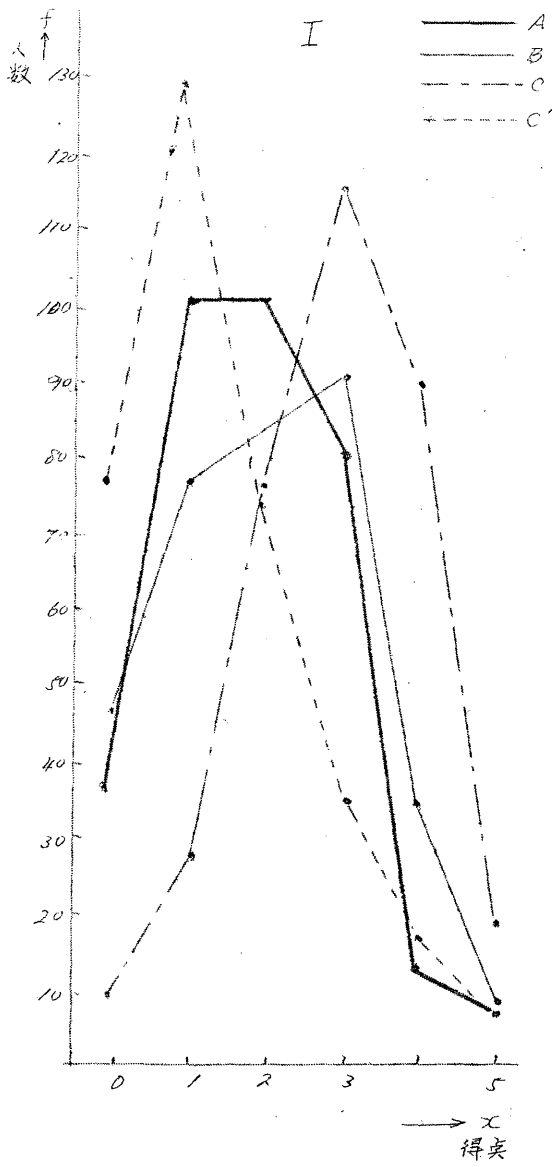
群	$\bar{x}$	$s$
1	68.07	10.580
2	67.34	10.305
3	67.18	10.210
4	69.27	10.365
5	67.52	10.400
6	67.74	10.110

また各問について、例えば I 問なら第 2, 3 群は共に A 形式であるがこの両者の差は全く見られず、比較には両群を合して A, B, C の 3 群にまとめて行うこととした。

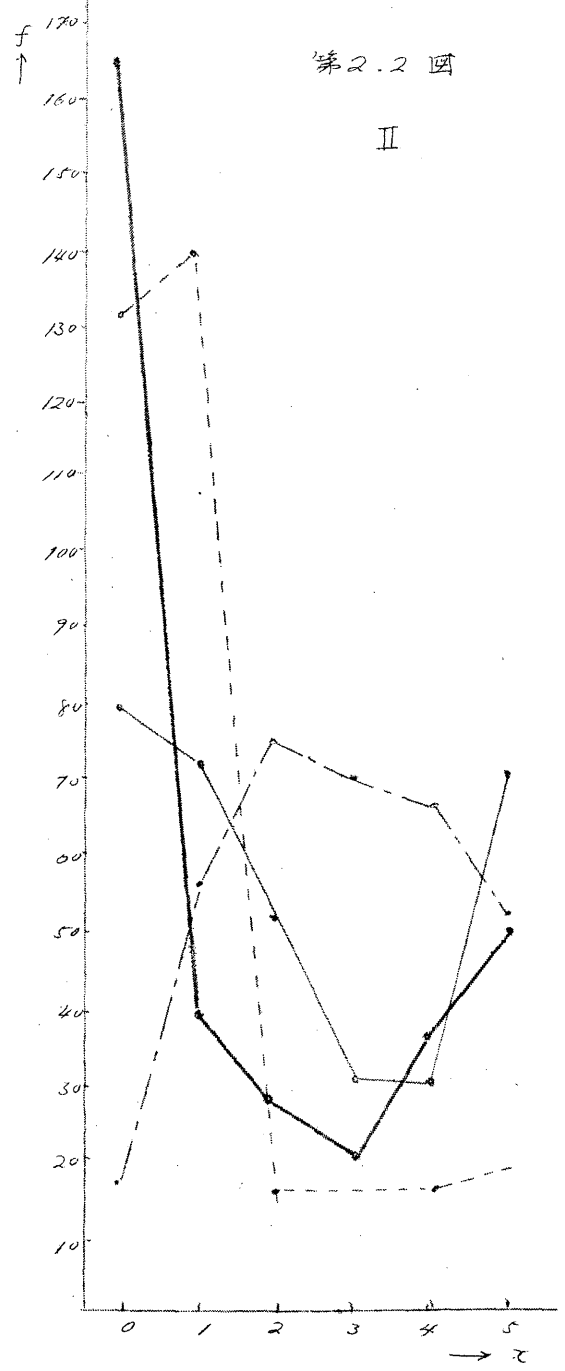
各問の通過率及び誤答の分布、各問の得点分布は第 2, 3 図のようである。これより分ることは

1. 問題が困難になるにつれて A 形式では その他 という回答が多くなる。 B 形式ではそれが 4 つの選択肢のどれかに押しこまれて了つて

第2.1 回

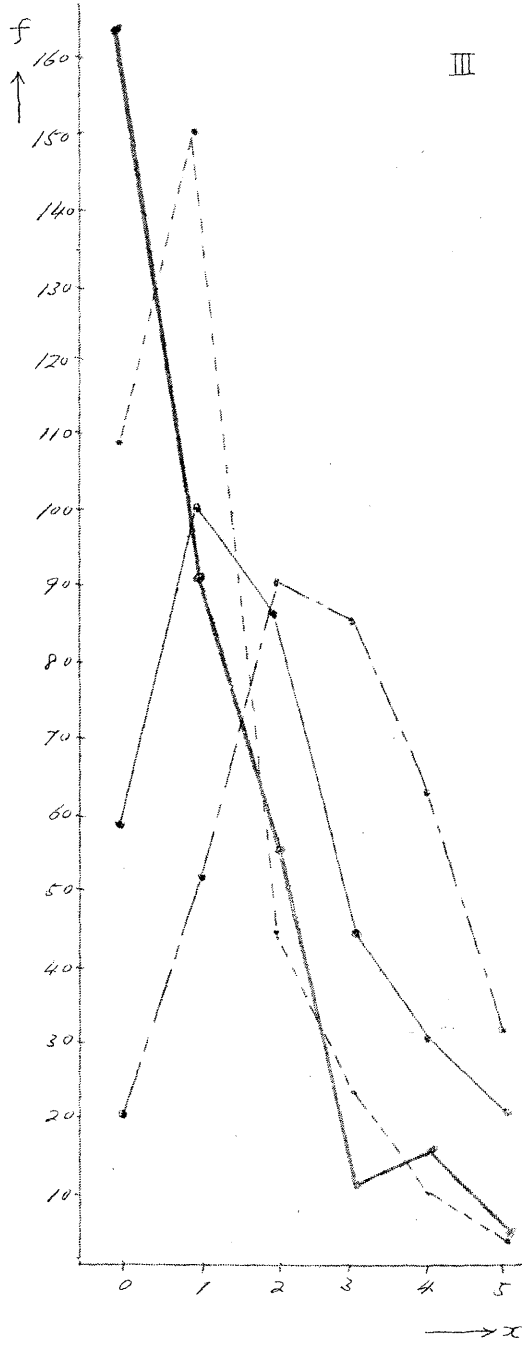


第2.2 回

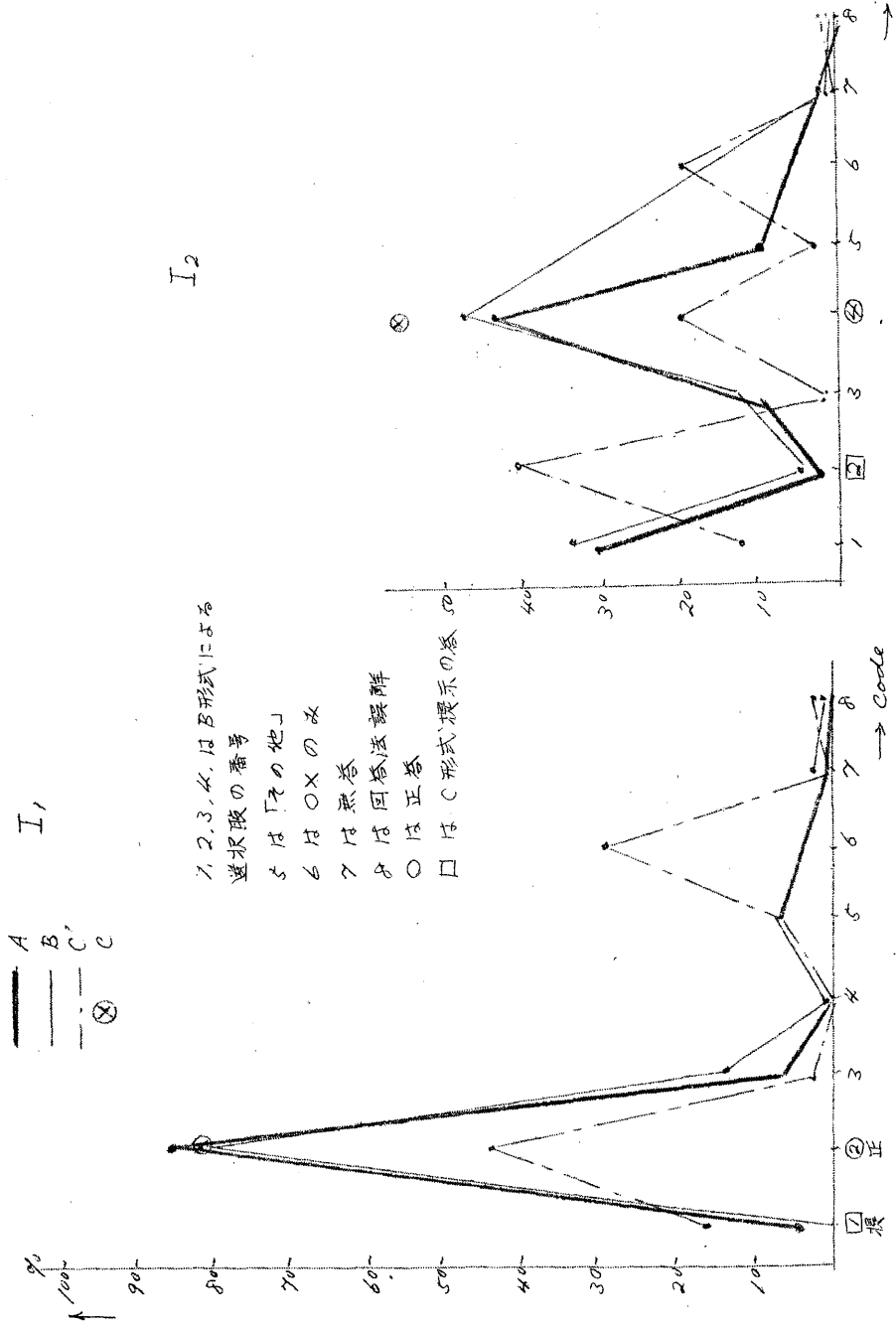


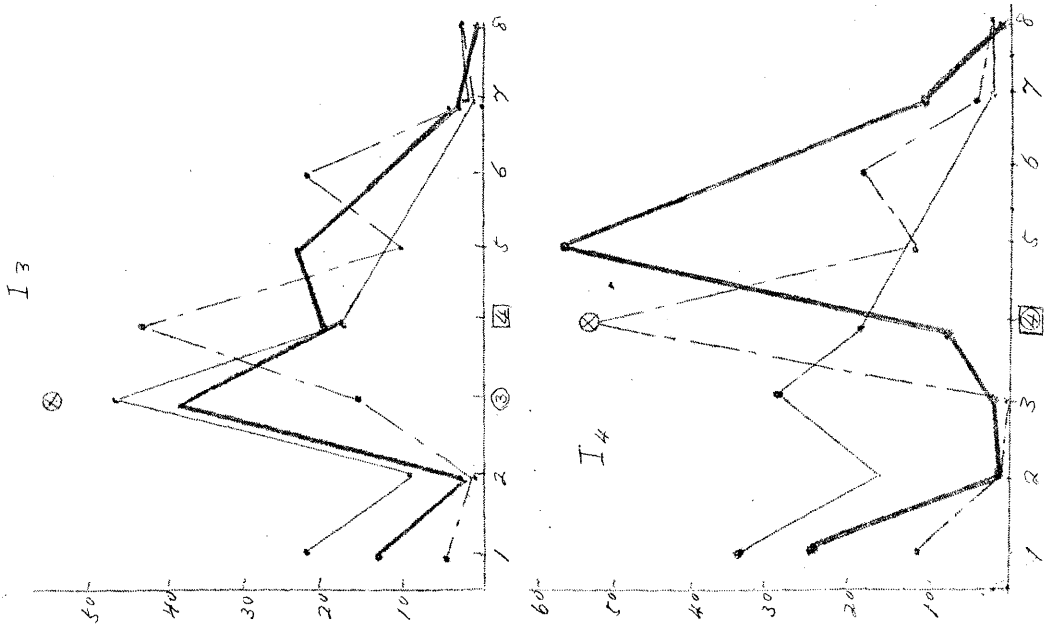
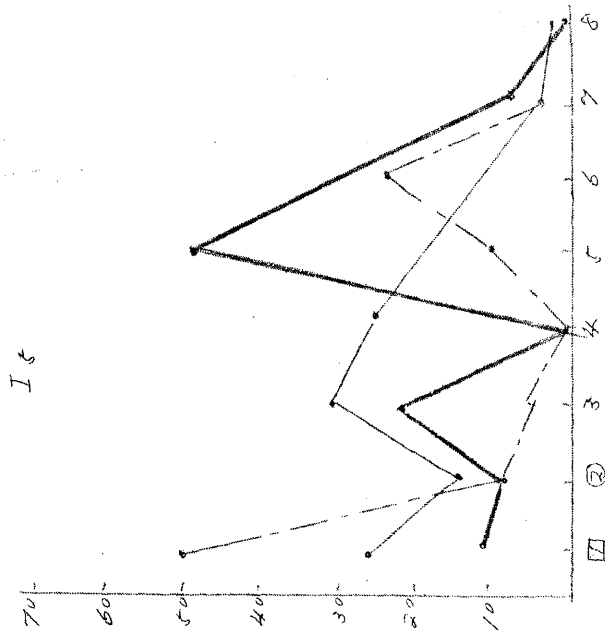


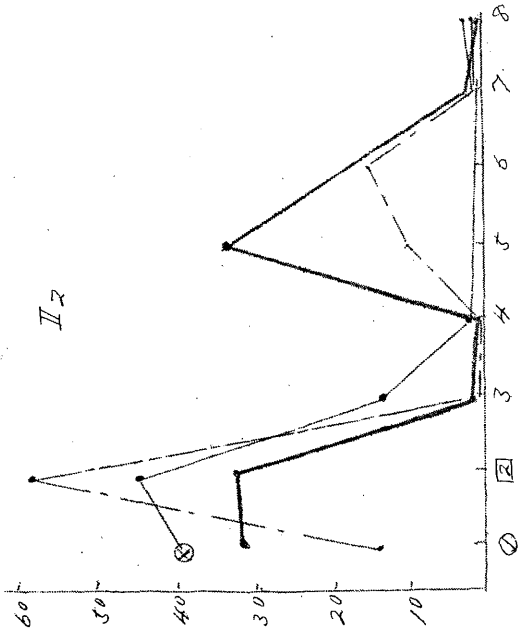
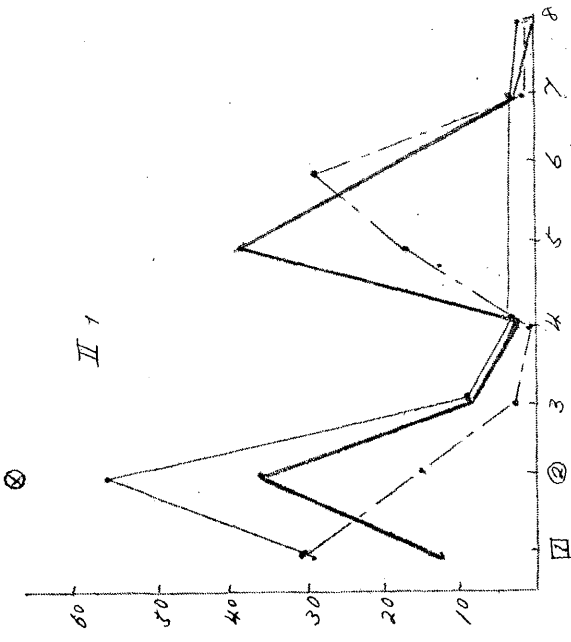
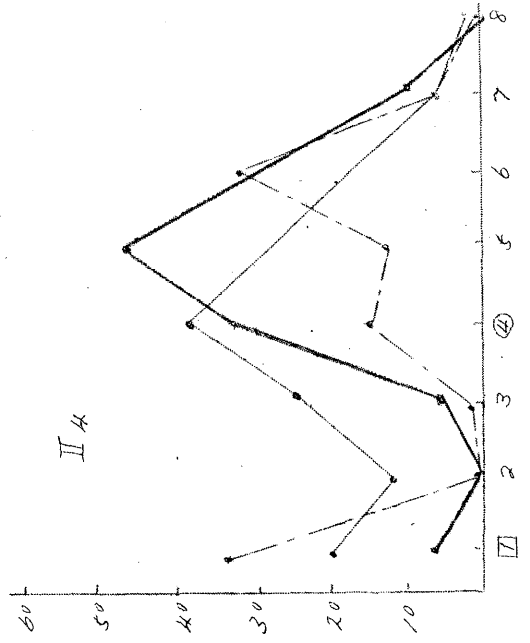
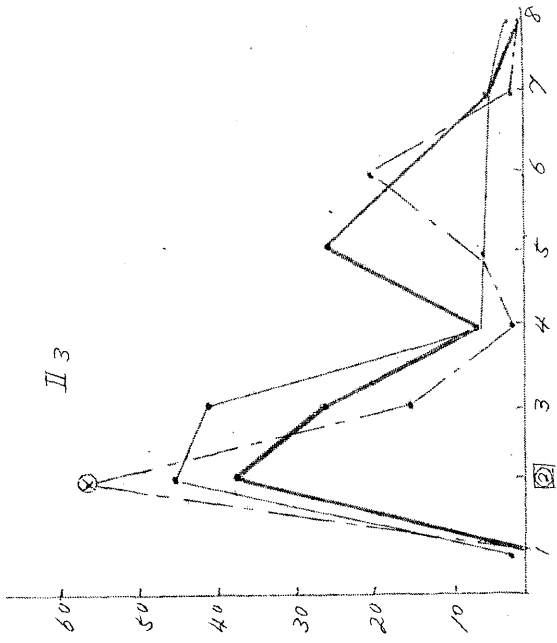
第 2.3 圖

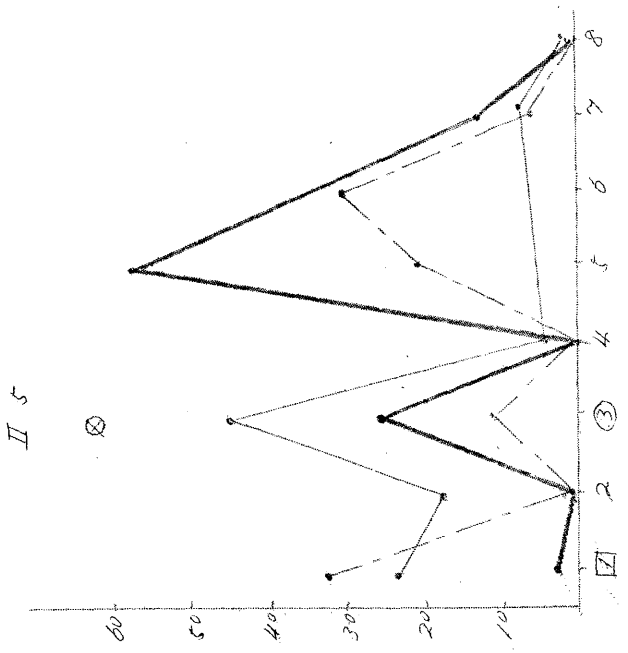
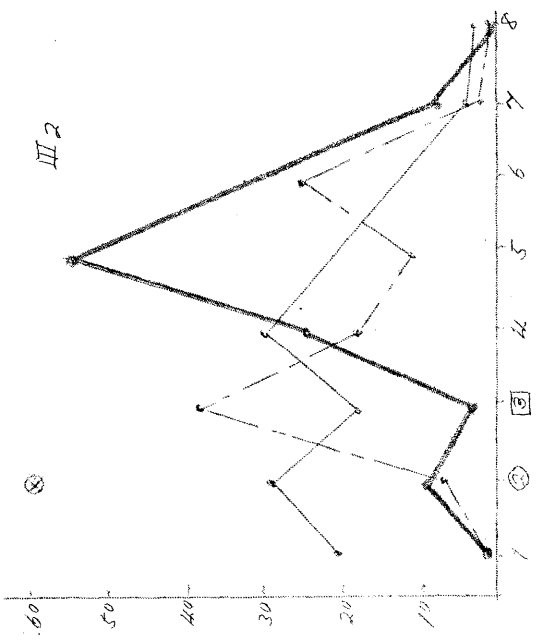
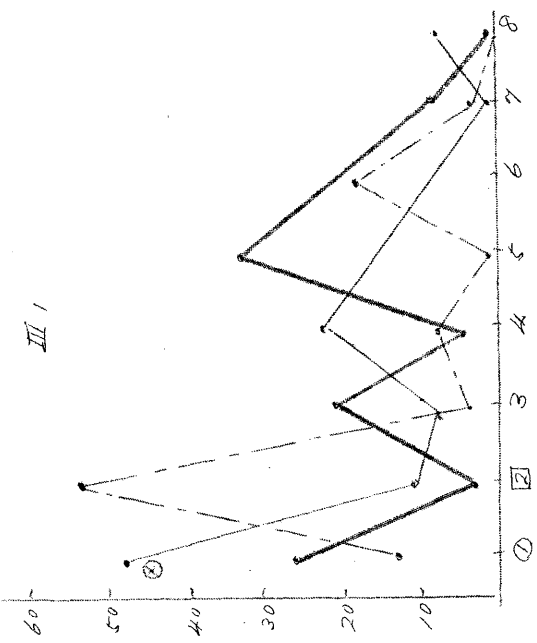


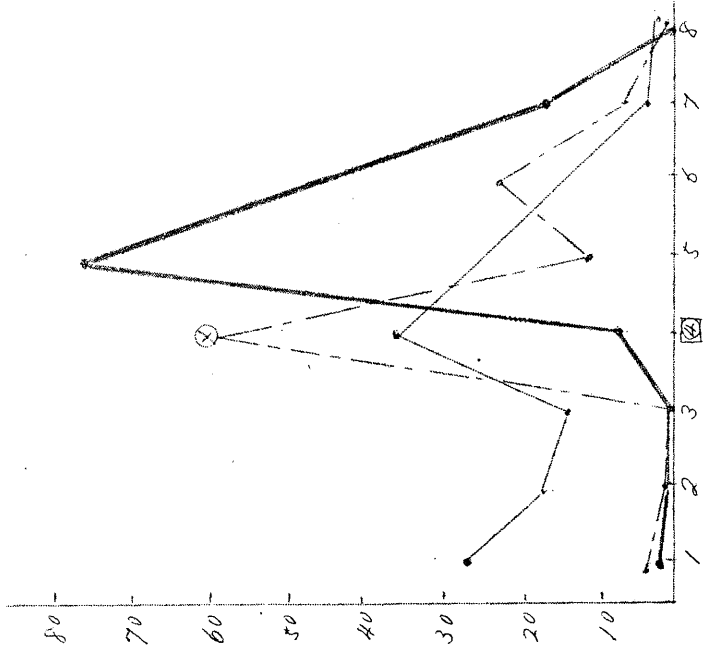
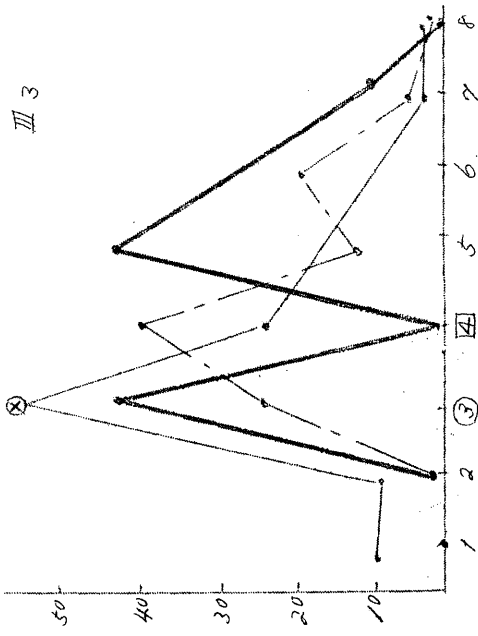
第3図〔正答誤答分布表〕



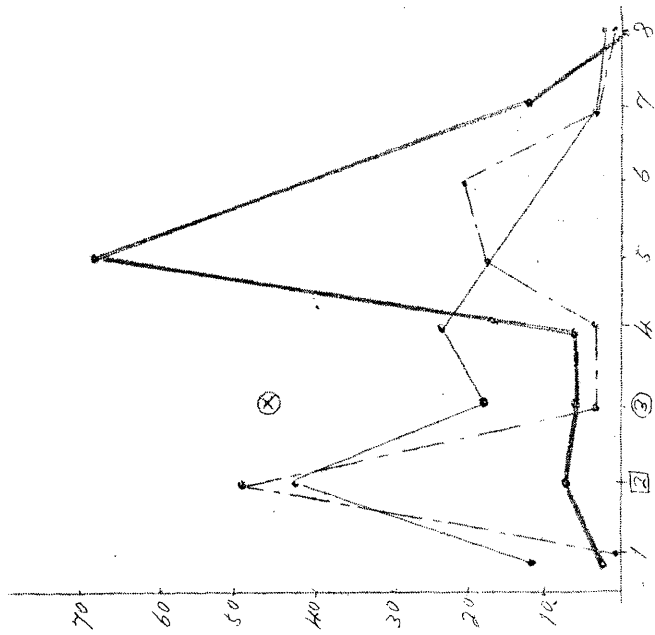








III 5



いる。従って選択肢にその他の欄をつくることはテストに於ても必要であらうと考えられる。

2. C形式とA形式を比べるとその他はC'の方が少い。それは答を予め提示したため(或る意味での *bias question*)であらうと思われる。
3. C形式では提示されたものを *accept* し易い。このことは問題が、困難となるにつれて顕著である。これは面接調査でいうと *bias* にかかった回答と考えられる。
4. 通過率はC形式の正答提示以外は殆んど常に  $C > B > A > C'$  であり、得点についても第5表のように  $C > B > A > C'$  の順である。
5. B形式, C形式は真の学力を知らうとする立場からは必ずしも妥当とはいえない。勿論問題が変わるときは更に検討を要するであらう。(問題 *source* からの *random sample* が必要となる)。参考のため輿論科学協会の行った共産党に対する態度調査の結果と比較しておこう。これはA群を *free*, B群を多項選択法, C群を二項選択法, によって質問したもので, Aは「共産党に対してどんな感じをもっていますか」Bは「共産党に対してどんな感じをもっているか」と尋ねて7つの選択肢を記入した紙をみせてえらばせ,

I	$\bar{x}$	$s$
A	1.854	1.136
B	2.045	1.2775
C	2.898	1.1269
C'	1.420	1.2025

第5.1表

II	$\bar{x}$	$s$
A	1.622	1.9311
B	2.232	1.8612
C	2.807	1.4642
C'	1.107	1.3780

第5.2表

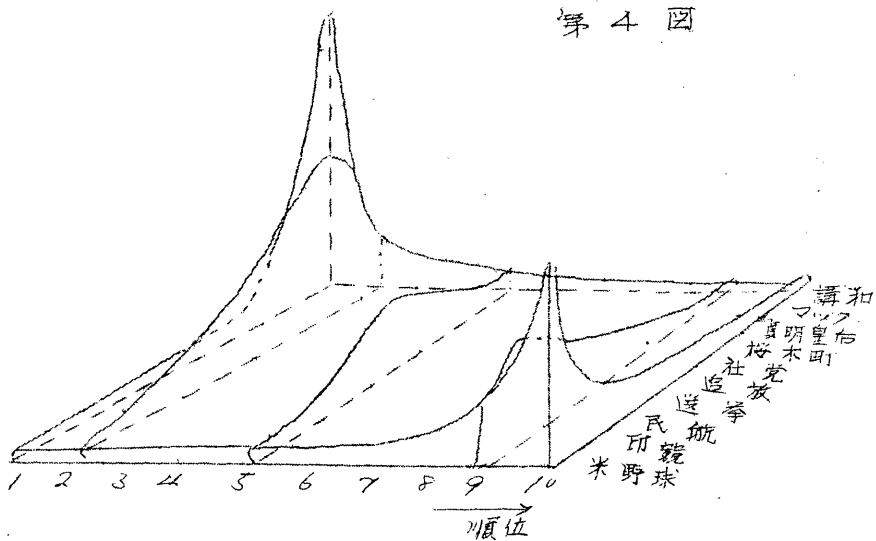
III	$\bar{x}$	$s$
A	0.920	1.1610
B	1.848	1.8986
C	2.628	1.3323
C'	1.057	1.0464

第5.3表

Cは「共産党に好感をもちますか、もちませんか」という質問をしたものである(第6表)



次に第2部の ranking (A群) と best one (B群) multiplied choice (C群) の比較をのべてみよう。この3群の等質化は知能点の外資に control として十大ニュースに1から10までの順番をつけさせた。

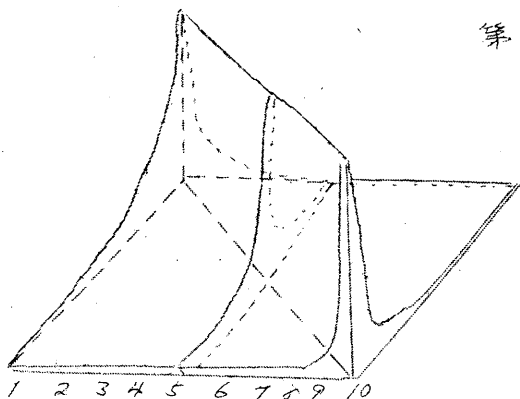


第 6 図

	好意的	非好意的	同比較ぬ	分らぬ	その他	無答	すく答ぬ	漸く考える	なかなか答えぬ	答えなし
A	7.5%	61.4	8.1	12.2	2.1	9.3	59.5	21.9	8.7	9.9
B	11.6	65.6	15.2	4.6	1.3	1.7	51.6	28.0	16.6	5.0
C	8.0	73.1	14.7	—	2.6	1.6	76.4	16.3	3.8	3.5

その結果第1群は第7表のようである群とも同一の結果を示した。これを図示すると興味深い。(第4図) 理想的に順位が附せられるものなゆ

は第5図のようであらう。第4図から分かることは「講和が結ばれたこと」



第5図

「アメリカの職業野球チームが日本に来たこと」は明瞭に第1, 10位を占めたのであるが中間のものは明瞭な差が示されなかつたことを示す。

第7表

—— 横の mode  
 - - - 縦の mode

項目 \ 順位	1	2	3	4	5	6	7	8	9	10	計
講和	<u>226</u>	26	16	8	5	6	8	4	1	5	305
マカサー	26	<u>114</u>	61	46	25	19	9	4		1	4
貞明皇后	18	56	<u>79</u>	49	39	20	17	13	9	5	3
桜木町	23	40	48	<u>76</u>	40	28	26	8	11	5	3
社会党 分裂	4	17	38	38	<u>48</u>	45	43	30	27	15	3
追放 除名	2	21	22	32	<u>53</u>	<u>54</u>	46	37	22	16	4
選挙	2	21	15	28	31	38	<u>47</u>	39	39	45	4
民権 問答	1	4	14	12	31	42	<u>51</u>	<u>65</u>	52	33	4
野球 競会	2	3	8	12	22	35	38	63	<u>85</u>	57	4
米野球	1	3	4	4	11	18	20	42	59	<u>743</u>	4

このことは split してしるべた問題「今度の日曜日にしたいものに  
 順位をつけよ（一つだけえらべ）、（いくつでもえらべ）」という問題に  
 対しては明瞭な差がでてこない。しかも順位は次表のようになって完全  
 には一致しない。（一致係数は 0.577 で一致性はないとはいえない）こ

群別 \ 項目	本 読	ハイキング	衣 着	ス ポー ツ	映 画	模 型	ベ ー ス ボ ール
順位づけ	1	2	3	4	5	6	7
一つだけ	2	3	1	6	5	4	7
いくつでも	1	4	2	3	5	6	7

のことは図  
 でいうと対  
 再線上にモ  
 ードが表わ  
 れないこと  
 を示す。従  
 って best  
 scale だけが  
 全体の順位

(注) 項目は(本を読む)(スポーツをする)(野球を見に行く)(映画を見に行く)  
 (ハイキングに行く)(模型を作ったりラジオをいじったりする)(かわいい着な  
 人達のために行く)の7つである。

を推定することは無理であり、いくつでもえらばせる場合も真の順位は  
 異なった結果をもたらすであろう。従つて必要な順位までつけさせるか、引  
 度をしるべる措置をとらなければならぬことが分る。

次に第3部の結果の一部をのべよう。問題は4つで各々は2つに split  
 されている。例えば第4問は「F君とH君とがっれたつて歩いているし  
 た。何うから校長先生がこられました。G君はおじきをしました。先生  
 もおじきをされました。H君は平気な顔で歩いていたG君は怒つて  
 H君をなぐりました。多々の人はG君のしたことはよい行いだといつて  
 います。あなたはG君のしたことについてどう考えますか」という質問  
 では下表の如き結果を得た。

	大変よい行 いと思う	相当よい行 いと思う	よいともわる いともええな い	相当わるい行 いだと思う	大変わるい行 いと思う
(よい)	44.5% (44.5)	44.1% (44.1)	51.3% (49.5)	22.8% (22.7)	17.2%* (19.2)
(ない)	5.2 (4.7)	2.5 (2.5)	47.4 (43.6)	21.3 (22.4)	23.1* (26.9)

(注) (よい)は問題中—の部分のあるもの、(ない)は—の部分のないもの  
 \*印は有意差のあることを示す。

理由になっていないものを含めれば、( )内の名となる。これによつてみても問題の形式によつて、得られた結果の解釈には心理的要素を加味しなければならぬことが分る。

以上ごく大雑把に調査形式のもつ偏りをもつたのであるが、調査がどのような場面で偏りをもつようになるかを一まとめにしておくことも無意味ではなからうと思うので重複をいとわずに列挙してみよう。

## I. 企 画 の 段 階

1. 調査目的は正しく把握されているか。
2. 調査主体は科学的な意図をもっているか、政治的、思想的の意図をもっているか。
3. 標識は正しく定義されているか。
4. *Pretest* は十分行われたか。

## II. サンプルに関するもの

1. *universe* の定義は明確か、*list* と *map* は正確完全か。
2. *Sampling* は質問媒体とよく適合しているか、精度と費用は経済的に考慮されているか。
3. 推定値は *unbiased* か、*biased* か、精度は計算してあるか。
4. *Sampling* は指示通り正確に行われたか。ごまかしと有意的選出が行われていないか。
5. *Non response* による *bias* の程度はどうか。

## III *Field work* と *Response error*

1. *questionnaire* の *bias*、即ち標識の定義に基づくもの、質問の媒体(郵便、面接、留置等)による *bias*。質問の形式による *bias*。
2. *Interviewer* による回答の *bias*、*interviewer* の虚偽 (*cheating*)
3. 回答者の回答による *error* (年令、教育程度、職業の誤り、記

憶遠い、隠蔽)

4. 指示 (*instruction*) は完全か、正しく履行されたか。
5. 時期おくれの回数による *bias*。
6. 実施中の監督不十分。
7. 調査時期と期間の不適切より生ずる *bias*。

#### IV. 集計と分析の段階

1. 集計の *error*, 分析の *error*。
2. 誤った総表と結果の解釈の誤謬

以上のようなことが我々の質問を用いる調査に入ってくる *bias* である。これらは互いにかみ合つて入ってくるものであるから、それらの絶対的な大きさを知ることは中々困難である。要はその *bias* の起るべき所在を明確にして最善の計画を立てる外はないのである批判に際してもこれらの点から客観的に評価して行くことが望ましい。

以上で大要質問を用いる調査研究における誤謬についてのべたのであるが、結局のところ調査の全体的な構成が重大なためであつて推計学の形式を応用したというだけではごく一部の *bias* を減少し得るにすぎないことが分ると思う。我々が統計数理と称するものはこの全体的な行動原理をさすのであつて、企画、形式化、運営実施、分析、数量化、予測の各段階をもつ方法論である。それは常に *dynamic* な現象を最も合理的に、即ち経済的にも、精度的にも最適な方法を以て説明して行くものである。而もそれは嘗ての数理統計学の如く固定した形式の残骸ではなく、現象と共に *optimum situation* を求めて変動して行く柔軟性をもたぬはならない。このような考え方が一般の調査にも用いられ、より良い調査が行われ行くことを切望してやまない次第である。

(1952.6.14)