

# 面接調査法の諸問題

西 平 重 喜

(1955年3月受付)

## On the Interviewing Method

Sigeki NISHIHARA

### §0. Introduction

In this paper, we shall prefer to the interviewing methods on the research surveys which we have been investigated by random sampling method.

### §1. Analysis of "non response"

#### 1.1 Non response ratio.

In these surveys, we utilized ration passbooks, poll books, and lists of resident registrations for the sampling lists. As a rule, students were employed as interviewers.

The non response ratio on each survey was generally about twenty percent.

#### 1.2 Reason for non response

The reasons for non response were 'change of address' accounts for 18 to 24 percent, 'address unknown' composes 10 to 24 percent, both totals amount to about 40 percent. For the rest 'not at home,' or 'on a journey' accounts for about 40 percent.

#### 1.3 Predeterminant items

(1) The non response ratio of males is larger than that of females.  
(2) The people in their twenties and more than seventy years of age have higher non response ratios than others.

(3) The ratio of non response showed a variance in each district and the urban ratio of non response showed higher than the rural.

(4) If we substitute the non responding sample the responding sample which is chosen at random from the responding sample of the same sample spot, we shall have the distributions shown in Table 11, 12. These distributions are classified into response group (first line), non response group (second line), real sample (third line), collected sample mentioned above already (fourth line) and census data (fifth line).

#### 1.4 Dynamic items

In the case of research survey by mail, we can see the difference between distribution of the mail response from the response group of the interviewing survey and the distribution of the mail response from the non

response group of the interviewing survey (cf. Table 13, 14).

## § 2. Call back

2.1 The sample who gave an interview at the first call was about half of the entire responding samples. At the first and second call, interviewed samples accounted for 85 percent. Including the third call, the ratio covers more than 95 percent.

However, as to all the samples, including non responding ones, the ratio of those interviewed becomes less than 50 percent of those at first call, then 70 percent at first and second call, and about 80 percent at 1~3 times call.

### 2.2 Predeterminant items

The mean of the number of calls on males was larger than that of females. In addition to the above, there were significant differences in compliance by age and formal education.

2.3 In order to compare the response procured at the first call with that obtained after having called more than twice we made the following test on the whole category. (See page 67.) In this formula  $p_1$  stands for the percentage of the responding samples given at the first call in a certain category,  $p_2$  is the percentage of the responding samples obtained after making a call more than twice percentage in the same category,  $p$  is the one of the total responding samples.

According to this test, a significant difference of the percentage between the two responding groups is seen in the 14 categories among the 300 categories.

Besides, in these 14 categories a significant difference also can be seen between male and female.

## § 3. Falsification by interviewer

3.1 In order to find any cheaters on the survey, three research methods were utilized, mail survey, test-retest and faked samples.

As a result of the investigation through the above methods, we found that from 6 to 20 percent of the interviewers (students) had cheated. But, the ratio of the samples falsified by interviewers was 5 to 10 percent of the total sample. There is some possibility that a sample may not answer giving the actual facts in the mail survey method but as far as the samples interviewed by the members of our institute are concerned, we found that no one declared they had no interview.

3.2 The error caused by falsification is corrected by the method shown in Table 22. In this table,  $A_i$  is the distribution is corrected by the method described under § 1.3 (4),  $C_i$  is the estimation of the distribution of false samples from the result by mail survey.

$$D_i = A_i - C_i, E_i = D_i / \sum_i D_i \text{ and } F_i = (\sum_i A_i) \times E_i$$

where  $i$  is category.

By this method, we tested with six questions (cf. Table 23), and the-

result of the tests showed that the influence of cheating on the analysis of the result of the survey was not so strong.

#### § 4. A clerical error by interviewer

4.1 We had the eighteen interviewers check the answers on the questionnaire, on playing the record of the actual interviewing scene. Consequently, the mean of correct filling in was 78 percent of the total questions. In the above mentioned record, we used vague expressions intentionally.

But in the light of this result at the main survey, we withdrew the questions which were apt to be misleading.

4.2 In another linguistic survey, four expert interviewers tried to the same kind test, the correlation coefficient between interviewers was about 0.9 (cf. Table 30).

4.3 In the case of surveying by the false sample method mentioned (§ 3), the ratio of concordance between the interviewers and members of our institution amounts to about 85 percent.

At that time, the false sample answers so vague, then in the case of an ordinary survey, the ratio of concordance would be higher than 85 percent.

#### § 5. A description of interviewer types

5.1 Personality test of interviewer. Table 31 shows the correlation coefficient. *a* means non response ratio. *b* is the average number of calls. *c* is the appointed time for interviewers attendance. *f* is the evaluation on the interviewer by supervisor.

According to this table, there is no notable mutual relation among these dimensions, but at least it can be said that the interviewer who showed abnormal reactions in more than two dimensions was unsuitable.

5.2 The influence of the interviewer's opinion. In order to know whether the interviewer's opinion exerts a bad influence upon the result of the research survey, we investigated the opinions of interviewers, but no influence was found.

5.3 Experience of interviewer. The rank in Table 40 shows interviewers' experience. The first row expresses student interviewers with previous experienced, the second row is inexperienced students, the third row is experienced interviewers (non-student) and the fourth is inexperienced interviewers (non-student) *a, b, e, f*, indicate the same as in Table 31. *c* is the ratio of concordance between the interviewer's opinion and sample's opinion. *d* is the rank pertaining to the ratio of concordance of supported political party by interviewer and sample. *g* is the rank on false report. Table 41 is the mean rank deduced from Table 40 and the coefficient of concordance.

#### § 6. Bias except by interviewers

6.1 Reliability of sample's response. We tried the 'after-survey'

(retest) six months after the 'before-survey' had been carried out, and in that 'after' 15 percent of the cases made different responses from their 'before' answers on the attitudinal question.

In the case of supporting a party 50 percent of the samples support the same party as before, and only 8 percent changed to the opposite-party support.

#### 6-2 Validity of interviewing method

On our linguistic survey, after explaining some scene, we asked the person being the question, "what would you say in such a case?" at the interviewing survey. Before interviewing survey, we recorded secretly sample's conversation at the real scene by the recording machine. The real scene is the same scene that we use at the interviewing survey.

Comparing sample's answer at interviewing with his conversation at real scene, we find very higher degree validity of our question.

#### 6-3 A false reply by the interviewed person

After finishing the elections, we asked a question of the persons interviewed as to whether they voted or not, then after that, checked the responses with the records of the pollbooks.

Table 43 is the contrast between answers and records. In this table, the first line refers to the one who actually voted as answer indicated, the second line concerns the one who abstained from voting, likewise his response. The third line lists the one who voted while his response was "abstention from voting", the fourth refers to the one who abstained, but gave his answer as "I gave my vote to so and so." The fifth refers to the obscure cases.

Furthermore, the first row is based on the election of the chief of a certain ward, the second row is the case of the election to the House of Representatives in 1949, the third row is the case of the election of the Governor of Tokyo in 1951.

The questions on the elections of chief of the ward and the representatives of the Lower House were asked of the same samples. (See the cross tabulation Table 44). In this survey, it has been proved that 74 percent of the samples responded truthfully on both elections and only 4 percent falsified their answers to both questions.

#### 6-4 Is the question understood by the interviewed person?

The methods of the research on this matter are as follows (the findings shall be omitted in this paper).

- (1) To make the sample choose from many categories, the one category which describes the right meaning of the question.
- (2) To investigate whether his response is in accordance with reason.
- (3) First, a certain concrete example was shown, then, the person was made to choose the category in which the same thing was written in an abstract way.

## §7. The Others.

7-1 The influence attributed to the research survey. The voting rate of the samples who received the survey, shows a higher percentage than the others (universe).

In Table 45, the upper score indicates the election of governor of Tokyo, the lower one concerns the election of the chief of a certain ward. The first line shows voting rate. The figures on the left side from the thick line are the result of the survey which was done before the voting date.

In this table, from the left, each score describes the survey which was done from ten to twelve days before the voting date, from three to five days before, the preceding day, the evening of the voting date, four to six days later, another random sample which was unsurveyed, and the last row indicates the voting rate of all voters (universe).

7-2 To make an interviewing survey accurate. Table 46 is the comparative table of the survey about the rate of each political party's poll in the elections of the House of Representatives. (We mentioned here 4 large parties only.)

The top line is the actual rate of vote obtained by each party, and the other lines are the estimated rate of each party's poll surveyed by a certain calculated method based on the public opinion research surveyed by newspapers.

In this table, the first line shows the ballots counted on February 28, 1955, the second, third, and fourth line shows the expected vote, the fifth line is the vote counted on April 19, 1953, the next two lines indicate the anticipated vote, the eighth line is the result of the election on October 1, 1952, and the next two lines refer to the forecast.

The party is from left to right, Liberal, Democratic, Social-Right-Wing and Social-Left-Wing.

We have described several kinds of error on the interviewing survey. The total of these error may be larger than sampling error. But the fact of table 46 shows us that interviewing survey gives fairly good information.

### Appendix: Bibliography of Research.

As the theses listed here are only those which the writer has read so far, this list cannot be said to be a complete one.

Further, in this paper, there is no part quoted from any of these theses.

§0 はじめに	§5 調査員の種類
§1 調査不能	5.1 性格
1.1 調査不能率	5.2 意見
1.2 理由	5.3 経験, 学生
1.3 決定的要因	§6 調査員に原因のないゆがみ
1.4 ダイナミックな項目	6.1 サンプルの答の信頼性
§2 追求	6.2 面接調査の妥当性
2.1 訪問回数	6.3 サンプルのつくウソ
2.2 決定的要因	6.4 質問は理解されたか
2.3 関係ある項目	§7 その他
§3 調査員のインテキ	7.1 調査の影響
3.1 発見法と比率	7.2 調査は正しいか
3.2 インテキの分析	§8 材料とした調査の説明
§4 調査員の記入のあやまり	付録 文献

## §0 はじめに

最近の統計学は、サンプリングや、数量化の理論の上でめざましい発展をとげてきた。この結果各方面で社会調査がおこなわれるようになり、そのデータはこれらの理論によつて処理されている。しかしこれらの理論は、完全な調査から得られたデータにだけ適用されるべきものである。ところが、統計学の一つの大きな分野である、調査技術についての研究は、ひとつひとつの問題に対する研究はあつても、それをまとめたものはあまりないようである。このためには十分くり返えされた調査とつみかさねられたデータが必要である。

われわれは、ここ数年いろいろな調査をおこなつて来たが、いつもこの調査技術ということを考えてきた。そこで、これらの結果を一応まとめてよい時期になつたと思ひ、いままでの調査のデータを縦断的にまとめてみることにした。ここでは、問題を面接調査に限つたが、調査技術に関する他の諸問題——質問法、調査法、特殊技術などについても、適当な機会に、このような形にまとめてみたいと考えている。

ここで、面接調査というのは、調査員がサンプルの家を訪問して、サンプル自身に会い、一定の調査票に従つて、サンプルの答をとつて来ることをさす。この論文で材料につかつた調査の主なもの、§8に説明をつけた。このため、本文の中では調査の略称を〔 〕でかこんで示した。ここに使つた調査はいずれも筆者が直接参加したものに限つた。また、この中でのべるような形に整理される前のいろいろな調査にはふれなかつたが、この中で述べていることと、むじゆんするような結果は出ていない。

われわれの研究は、ほとんどいつでも研究所の内部又は外部との共同研究である。そうして始めの計画から終りの分析まで、徹底的にディスカスされるので、どれだけが誰のアイディアによるものか、誰からどんな有意義なサジェスションを受けたかはつきりしない。また地味な仕事と、時には不愉快きわまりないことまでやつてくれた研究補助員達なしには、とてもなしとげることができなかつたことが多い。この中では扱つた題材が多いので、深い感謝の意を表わすために、ひとりひとりの名前をあげられないことを、大変に残念に思う。

付録にかかげた文献は、1940年代のものが大部分である。しかも、筆者が直接または間接に本文に当ることができたものだけに限つたので、リストとしてかならずしも完全なものとはいえない。

### §1 調査不能

ふつうわれわれはサンプリングの台帳として、物資配給台帳、選挙人名簿および住民登録の3つを使つてきた。しかし、このうち配給台帳は現在不正確となり、東京では米の販売店ごとにまとめてあるので、実用に適さなくなつてしまつた。以下で引用する調査のサンプリングは、いずれもこれらの台帳にもとづくものである。

ここで、調査不能というのは、面接調査ができなかつたサンプルと定義をしておく。

#### 1.1 調査不能率

まず、各調査を通じて調査不能がどのくらいあるかをみよう。第1表は東京の23区での調査、第2表はそれ以外の範囲での調査の結果である。ここに、調査不能率というのは、調査不能者数の割当サンプル数に対する比である。

これらの表によれば、どの調査でも、不能率は20%前後を示している。

第1表 調査不能率(東京23区)

〔調査名〕	都 知 事			国民性	国民性	効 果
	I (区)	II (区)	III (区)	(準 備)	(東 京)	
割当サンプル数	987	911	344	906	500	1800
調 査 不 能 率	18	18	24	23	17	21
サンプリング台帳	選 挙 人 名 簿			住 民 登 録		

第2表 同 上 (他の地域)

〔調査名〕	都 知 事			国 民 性			
	I (郡)	II (郡)	III (郡)	6 大 市	市 部	郡 部	全 国
割当サンプル数	137	127	48	305	579	1370	2254
調 査 不 能 率	18	16	27	17	18	16	17
母 集 団 の 地 域	東 京 都 の 市 , 郡 部			区' 部	一 般 の 市	町 , 村	合 計
サンプリング台帳	選 挙 人 名 簿			住 民 登 録			

注: 表の中のイタリック数字は % をあらわす。以下同様。

#### 1.2 調査不能の理由

面接調査ができない理由を考えてみると、

- 1° サンプリングに使つた台帳の不完全さ: 例えば記入のちがひ, up-to-date でないこと。
- 2° サンプルの事故: 例えば病気, 旅行, 不在。
- 3° 調査員の責任にかかるもの。

になる。しかし調査不能の理由は、一般には調査員の報告からしか知ることができないのであるから、3° に分類されるような理由は出てこない。それでは、調査員の報告による、調査不能の理由はどのくらい信用できるものであろうか。〔効果〕では約半年後にパネル調査をおこなつたが、前調査で調査不能であつたサンプルも、後調査のときに再び訪問をさせてみた。この結果は第3表のようになる。ただし、前調査で死亡と報告されたものと、余りにも明らかに移転と判定したサンプルは、後調査のとき訪問させなかつた。

この第3表の中で左上の×印があるものは、前と後の調査の報告に矛盾があるものである。この矛盾があるものは26% (= 16/61) である。しかし、その中で2度とも調査不能の9人は、調査員

第3表 前調査で調査不能サンプルの後調査での調査情況〔効果〕

前 後	移 転	尋ね 当らず	病 気 な ど	旅 行	不 在	拒 否	そ の 他	計
移 転	6	1						7
尋ね当らず	1	3		× 1	× 1			6
病 気 等				× 1				1
旅 行	× 1			4				5
不 在	× 3	× 1		2				6
拒 否		× 1						1
そ の 他								0
回 収	× 3	× 4	1	4	19	3	1	35
計	14	10	1	12	20	3	1	61

の責任と考えるよりは、家族や近所の人が変わつたことをいつたと考えた方がよいようである。このような事情から、調査員の努力不足にもとづく調査不能はごく少ないと考えてよいであろう。

前と後の調査の間での回収（調査できた）サンプルと調査不能サンプルの数は第4表のようになる。後調査の不能の方が23人多いが、これは第3表で明らかのように、サンプリングしてから時間がたつたせいであつて、調査員の責任ではない。

また、〔都知事〕で調査不能となつたもののうち、不能の理由から考えると、当然投票できないもの、あるいは投票するはずがないと思われるもの——死亡、病気、移転、尋ね当らず——について、実際に投票したか棄権したかを、選挙人名簿の投票用紙を交付した印によつてしらべてみると、第5表のようになる。

第4表 前・後調査の調査状況〔効果〕

前 後	回 収	調 査 不 能	計
回 収	301	35	336
調 査 不 能	58	26	84
計	359	61	420

第5表 不能理由別の投票・棄権率〔都知事〕

不能理由	死 亡	病 気	移 転	尋ね当らず
投 票	—	33	15	51
棄 権	100	67	55	43
不 明	—	—	30	6
計	100 15	100 130	100 69	100 49

この結果‘死亡’では投票しているものがないし、‘移転’も投票しているものは少ない。ただ‘尋ね当らず’の半分が投票していることは、調査員の努力がたりないことを示しているといえよう。

さて以上のように、調査員が報告した調査不能の理由は、あやしい場合がないとはいえないが、大たい信用できるようである。しかも、これらの不能理由のうちには、調査員がいかげんに報告したのではなくて、サンプルの家の人や近所の人などが、いかげんな返事をするとも考えられるので、すべてが調査員の責任とはいえないであろう。



第6表 不能理由

理由 調査		理由										
		対象外	死亡	移転	尋ねず 当らず	病気	老衰	旅行	不在	拒否	その他	不能の計
〔都知事〕	I	—	2.5	26.6	14.8	10.8	0.5	12.8	28.5	1.0	2.5	100.0 203
	II	—	2.8	30.7	16.7	8.4	0.6	17.3	18.5	0.6	4.4	100.0 179
	III	—	5.2	20.9	10.4	9.4	1.0	10.4	34.4	2.1	6.2	100.0 96
〔国民性〕	全国	0.4	2.5	18.4	6.4	20.9		13.1	27.7	6.5	4.1	100.0 379
	東京	1.2	1.2	34.1	9.4	17.6		11.8	12.9	4.7	7.1	100.0 85
〔効果〕		0.2	1.6	26.1	23.6	6.4	0.5	12.1	25.4	3.0	1.1	100.0 373
〔鶴岡〕		—	1.6	25.0	0.0	21.9	—	25.0	—	6.2	20.3	100.0 64

そこで、調査員によつて報告された調査不能の理由の分布をしらべてみよう(第6表)。これらの諸調査を通じて、不能の理由の分布がいつも同じ割合であるかどうかという問題がある。調査を比較するためには、サンプリングの台帳や、調査した地域が同じでなければならない。

このため〔都知事〕の3回の調査を $\chi^2$ 検定で比較してみた。ただしカテゴリをまとめて6つにした。この結果は第7表のようになり、第II回の調査と第III回の調査とでは、理由の分布にくいちがいがあがるが、I, II; I, IIIの間ではそうは考えられない。

このことと、第6表の数字とをあわせて考えると、不能の理由は概して安定しているとみることができよう。

1.3 決定的要因別の調査不能

調査不能となつたサンプルについて、われわれが正確に知ることができるのはサンプリング台帳に記入されている性と年齢と住所だけである。この3つについて、調査不能のサンプルを検討してみよう。

調査不能が性別によつてかたよりがあるかどうかをみるために、調査不能率を性別にしたのが、第8表である。第8表はたとえば、〔鶴岡〕の男のサンプルのうち、13.8%が調査不能であつたことを示すものである。この表によれば、どの調査の場合でも男の方が女より調査不能が多いことを示しており、この差はサンプリングの誤差以上となるものが多い。

しかもこの性別の不能率が、各調査を通じて一定であるかどうかをみるために、ふたたび〔都知事〕の3回の間で $\chi^2$ 検定を試みた。この結果、男の不能率も、女の不能率も、いつも安定しているものと考えてよいことが判つた。

つぎに年齢別にした場合の不能率をみよう(第9表)。この表によれば20才代と70才以上では、調査不能率が高く、年齢によつては有意な差となる。ここでもまた〔都知事〕の3つの調査の間で $\chi^2$ 検定をすると、各調査を通じて若い年齢層と、老年層でたしかに調査不能が多いことを示している。

第7表 不能理由の分布の比較〔都知事〕

〔都知事〕	$\chi_0^2$	自由度	$P_r(\chi^2 \geq \chi_0^2)$
I と II	7.29	5	0.10~0.20
I と III	7.42	5	0.10~0.20
II と III	14.79	5	0.01~0.02

第8表 性別調査不能率

調査		男	女	計
〔都知事〕 (区部)	I	19.8	16.3	18.3
	II	20.0	14.4	17.5
	III	30.6	18.4	24.1
〔効果〕		25.4	15.8	20.7
〔鶴岡〕		13.8	8.8	

第9表 年令別調査不能率

年令		20~29	30~39	40~49	50~59	60~69	70~	計
〔都知事〕 (区部)	I	23.8	14.6	15.8	12.3	19.0	23.7	18.3
	II	24.6	13.0	13.0	12.5	17.9	27.2	17.5
	III	31.0	22.4	16.4	13.8	40.0	46.1	24.1
〔効果〕		24.2	17.2	17.9	12.2	16.7		20.7
〔鶴岡〕		12.0	13.8	5.9	11.8	10.6		
年令		15~24	25~34	35~44	45~54	55~64		

つぎにサンプルの住所から、調査不能に地域的な差があるか、市郡別に差があるかどうかをみよう。条件をそろえるために、全国調査である〔国民性〕についてみると、第10表のようになる。

この結果、北海道、中四国、九州は全国平均より調査不能が多いが、このような傾向は、他の郵便調査でも出たことがある。

また区部（6大都市）、市部（他の市）、郡部別では、郡部は調査不能が少ない。

第10表 住所別調査不能率〔国民性〕

	北海道	東北	関東	中部	近畿	中四国	九州	全国
割当 サンプル数	117	255	505	401	334	293	349	2254
不能率	22.8	13.7	15.4	14.6	17.6	20.8	20.6	17.8
	区部*	市部	郡部					
割当 サンプル数	1107	579	1370					
不能率	19.7	19.2	16.8					

\* 東京、大阪はサンプルをよけいとつてある。

以上で見えてきたところによれば、調査不能者は決定的要因の各カテゴリを通じて一定ではない。しかも各調査をとおして、各カテゴリの不能率はほぼ安定している。そこで、調査できたサンプルだけを集計することは、当然のことながら、危険である。

このことを考えて、〔国民性〕では調査不能サンプルの修正をするための、ひとつの方法として、各調査地点（市、町、村）ごとに、調査できたサンプルの中から、ランダムに調査不能者の数だけえらび出して代用してみた。これをここでは**修正サンプル**とよぶことにする。第11, 12表では、面接できたサンプル、調査不能だつたサンプル、割当サンプル（以上のサンプルのほかに、調査員がインチキな報告をしたサンプルをふくむ）、それに上記の修正サンプル、さらにこの調査より約2年半前の国勢調査の10%集計の結果とをならべてあげておいた。

第11表の性別をみると、調査不能者の構成だけが、ほかとちがつている。しかし調査不能者は17%たらずなので、面接できたサンプルと割当サンプルの構成はほとんど同じ

第11表 サンプルの性別〔国民性〕

	男 Male	女 Female	計 Total
面接できたサンプル	46.5	53.5	100.0 1856
調査不能サンプル	51.7	48.3	100.0 375
割当サンプル*	47.5	52.5	100.0 2251
修正サンプル	46.9	53.1	100.0 2254
国勢調査**	49.0	51.0	100.0

\* インチキをふくむ。

\*\* 昭和25年10%集計。

である。修正サンプルは割当サンプルにさらに近い値を示している。また、これらの結果と国勢調査の結果も非常に近いものであり、そのくいちがいは一地点から平均 15 のサンプルをとつたことから考えれば、サンプリング誤差の中に入るであろう。

第 12 表 サンプルの年齢別〔国民性〕

Age	20~24	25~29	30~34	35~39	41~49	50~59	60~	不明	計
面接できたサンプル	18.6	15.2	11.5	11.1	17.9	15.0	10.6	0.1	100.0 1856
調査不能サンプル	21.9	21.9	8.0	9.0	11.4	11.6	15.1	1.1	100.0 378
割当サンプル*	19.1	16.3	10.9	10.9	16.9	14.3	11.3	0.3	100.0 2254
修正サンプル	18.9	15.3	11.4	10.9	18.1	14.9	10.4	0.1	100.0 2254
国勢調査**	17.2	13.6	11.4	11.2	18.7	13.6	14.3	—	100.0

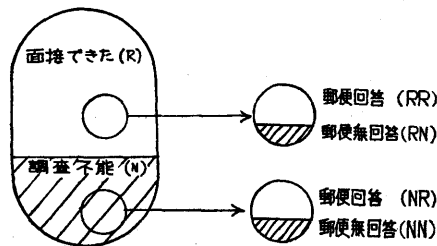
\*, \*\* cf. 第 11 表.

第 12 表の年齢別についても、性別のときと全く同じことがいえる。そこで以下とくに断わらない限り、〔国民性〕の集計用のサンプルとしては、この修正サンプルを使うことにする。

1.4 ダイナミックな項目と調査不能

われわれが調査することは、一般に前節のような決定的要因についてではなくて、ダイナミックな項目——意見、態度、行動など——である。もしも決定的要因で片よりがあつても、ダイナミックな項目で調査不能者の示すはずの分布と、面接できたサンプルの分布とが一致するならば、調査不能者を切り捨て、集計できるわけである。しかしながら、調査不能者がどんな分布をするかは、面接では分からないのであるから、郵便調査をおこなつて、比較をする。ところが、面接のときと郵便調査とは、サンプルの反応のしかたがちがうので、第 13 表のようなデザインのもとに調査

第 13 表 郵便調査のデザイン〔国民性〕



( ) 内はグループの記号 R: response, N: non response

をした。すなわち〔国民性〕では面接調査できたものからと、できなかつたもの（調査不能者）から、それぞれさらにランダム・サンプルをとり、郵便調査をおこなつた。その結果は第 14 表のようになる。ただし郵便調査ではつぎのような 2 問しか調査しなかつた。

問2 あなたはこの市(町, 村, 東京)に住んでいることに、満足していますか?

問10 以前は、女は家の仕事ばかりしていましたが、近ごろでは世間の仕事にたずさわるようになりました。あなたはこれは好ましいことと思えますか、それとも好ましくないことと思えますか?

第14表 郵便調査による不能者の分析〔国民性〕

質問(Q)	Mail survey		郵便調査の回答		面接できたものの全体	NR と RR の間の $\chi^2$ 検定
	Category	Group	不 能	で きた		
			(NR)	(RR)	(R)	
2 住 ま い	満 足		52.9	63.4	75.4	$\chi_0^2 = 8.224$ $D.F. = 3$ $P_r = 0.02 \sim 0.05$
	時 による		19.6	19.3	6.6	
	不 満		16.3	12.0	12.0	
	そ の 他		11.2	5.3	6.0	
	計		100.0 114	100.0 312	100.0 1856	
10 女 の 仕 事	好 ま し い		62.2	73.4	63.4	$\chi_0^2 = 6.177$ $D.F. = 2$ $P_r = 0.02 \sim 0.05$
	好 ま し く な い		14.7	13.1	24.4	
	そ の 他		23.1	13.5	12.2	
	計		100.0 114	100.0 312	100.0 1856	

ここで比較ができるのは、グループ NR とグループ RR であるが、この両グループの間で  $\chi^2$  検定をすると、2問とも統計的に有意な差が出てくるのである。すなわち、面接できたサンプルから来た郵便回答の結果と、面接調査不能のサンプルから来た郵便回答との間には差がある。郵便調査で回答をよこさなかつたサンプルについては判からないが、この結果は、面接できたサンプルと、できなかったサンプルとはダイナミックな項目についてちがいがあつたことを示している。

しかし、つぎのようなことを考えよう。われわれは、面接調査をできたサンプルを使つて、面接調査不能のサンプルの回答を推定し、修正サンプル (cf. 1-3) について集計をすすめている。これは、面接できたグループ R と調査不能のグループ N とが同じ性質とみなしていることになる。ところが、両グループからの郵便回答は、第14表のとおり、同じではなかつた。

しかし、調査不能のグループ N は、全体の約 17% であるので、もしグループ R とグループ N のちがいが小さければ、修正サンプルによる集計で十分によいはずである。このことを、検証するために、第14-a表のような計算をおこなつた。(問10については、途中を印刷することは省略してある。) すなわち、グループ R とグループ N を同質のものと考え、グループ R で推定するというのが、修正サンプルの考えである。この表の D 行は全サンプル 2254 人の郵便回答を、面接できたグループ R の中で、郵便回答もよこしたグループ RR の分布 (cf. 第14表) によつて、推定したものである。いいかえると、この D 行は修正サンプルの考え方に当る、郵便回答の推定である。

一方、グループ R の 1860 人は、RR によつて推定し、グループ N の 394 人は NR によつて推定し、その和をもつて全体の郵便回答の推定としたのが C 行である。

この C 行と D 行を、C 行をもとにして  $\chi^2$  検定した結果が、下の方に書いてある。すなわち、C 行と D 行とはその数字から見ても、同じ分布と考えるとよいであろう。

このことから、われわれの修正サンプルを否定するような結果は出てこないわけである。

第 14-a 表 郵便回答の推定による比較

	住いに					計	備 考
		満 足	時による	不 満	その他		
A	1860 × (RR の %)	1179	359	223	99	1860	グループRの郵便回答の推定分布
B	394 × (NR の %)	209	77	64	44	394	グループNの郵便回答の推定分布
C	A + B	1388	436	287	143	2254	全体を R, N 別別に推定した郵便回答
D	2254 × (RR の %)	1429	435	270	120	2254	全体をグループRとみなして推定した郵便回答
E	(C - D)²/C	1.211	0.000	1.007	3.699	5.917	χ² の計算

	χ₀²	D. F.	P <sub>r</sub> (χ² ≥ χ₀²)
Q 2 住 い	5.917	3	> 5 %
Q10 女 の 仕 事	5.642	2	> 5 %

§2 追 求 (call back)

2.1 訪問回数分布

われわれの調査では、ふつうランダム・サンプルを指定して2日から3日の間に本人に面接して調査することを要求している。そのためサンプルが不在であれば、この期限内であれば何回も訪問をしながらさなければならぬ。何回目に調査できたかを、各種の調査を通じてみよう(第15表)。

第 15 表 訪 問 回 数

[調査]	回 数							計	x̄	σ
	1	2	3	4	5~	不明				
都知事 I 区部	49.1	36.0	12.5	1.2	0.6	0.6	100.0	806	1.8	0.7
" II "	49.3	31.0	15.7	2.5	0.6	0.9	100.0	752	1.7	0.9
" III "	53.2	37.2	8.4	0.8	0.0	0.4	100.0	261	1.6	0.7
国民性(準備) "	57.8	29.8	9.7	2.2	0.4	0.1	100.0	698	1.6	0.8
" (東京) "	48.5	33.3	12.9	3.9	1.2	0.2	100.0	412	1.8	0.9
効 果 "	60.4	27.7	8.4	2.2	1.1	0.2	100.0	1426	1.6	0.8
国 民 性 全 国	56.2	31.9	8.3	2.2	1.1	0.3	100.0	2254	1.6	0.8

この結果をみれば、1回の訪問で調査ができたのは調査できたサンプルの約半分であり、2回まで訪問すれば85%位、3回では95%以上調査できる。しかしこの%は調査できたサンプルに対するものであり、すでに見たように調査不能が20%前後あるので、全サンプルに対しては、1回の訪問では50%ならず、2回までで70%位、3回では80%近くのサンプルを調査できることになる。

ところで、1回あるいは2回、3回までの訪問で調査できたサンプルが割当サンプルからランダムにとれていれば、いいかえれば指定したランダム・サンプルの2段抽出(subsampling)になっていれば、追求を途中でやめてもよいわけである。これを検討してみよう。

2.2 決定的要因別の訪問回数

性別・年齢別・学歴別に訪問回数をしらべてみよう。まず第16表の調査できたサンプルに対する平均訪問回数をみれば、男の平均訪問回数は女より多いが、この差は両調査ともサンプリングの誤差より大きくなる。年齢別も両調査とも同じ位の値を示し、学歴別では多少ようすがちがうが、[国民性]は全国調査であり、[効果]は東京の区部であるので厳密には比較できない。

第16表 要因別の面接回数

		計	男	女	年齢						
					20~24	25~29	30~34	35~39	40~49	50~59	70~
サンプル総数	[国民性]	2247	1053	1194	423	345	256	245	406	506	63
	[効果]	1423	684	739	294	239	166	146	269	284	28
平均訪問回数	[国民性]	1.62	1.73	1.52	1.78	1.69	1.59	1.60	1.59	1.49	1.46
	[効果]	1.58	1.71	1.46	1.71	1.67	1.47	1.50	1.66	1.45	1.36
1回で面接できた率	[国民性]	56.3	48.5	63.4	48.7	49.2	60.5	57.9	58.4	62.7	63.5
	[効果]	60.4	42.7	66.6	53.1	56.1	65.4	65.1	58.0	66.9	64.3
同上、割当サンプルに対して	[効果]	47.9	39.9	56.1	36.9	42.9	54.1	54.6	48.0	58.4	53.0

		学歴なし 小学中退	小学卒	高等小学 新中学	旧中学 新高校	旧専大 大学
		サンプル総数	[国民性]	129	608	788
	[効果]	42	233	415	458	259
平均訪問回数	[国民性]	1.33	1.47	1.67	1.70	1.83
	[効果]	1.57	1.49	1.66	1.54	1.60
1回で面接できた率	[国民性]	74.5	65.0	53.3	51.6	42.3
	[効果]	59.5	65.9	56.9	61.0	59.5

つぎに、面接できたサンプルのうち、1回の訪問でどのくらい面接できるかをみるために、面接できたサンプルを100として、そのうち、1回で面接できたものの率をみよう。性別では、男は1回の訪問で40~50%であるが、女は60%以上であり、この差はサンプリング誤差より大きい。年令別・学歴別でもカテゴリによりちがいがあがる。

〔効果〕では調査不能を含めた、割当サンプルを100としたとき、1回の訪問で調査できる率もしらべてみた。こうすると、ただ1回では50%たらずのサンプルしか調査できない。しかも男はわずか40%にすぎないことがわかる。

さて、以上を通じて、決定的要因別にみると、調査をするための訪問回数にはカテゴリによつて大分ムラがある。このことは調査不能とも関係があり、追求を厳重におこなわなければ、片よつたサンプルの調査票しか回収できないことになる。しかし追求をかさねる調査員の労力は大変なものであるし、その調査員がかつてに調査票をデッチ上げてしまつたり、指定したサンプルのかわりに他の人を調査するというようなことも起つてくる。これに対して、調査票をくばつて歩き、あとから回収する留置調査方法がある。労力としては配布と回収の2回でほとんどすむし調査不能率は大変少なくなる。しかし、この方法はサンプルのその時の知識、意見、態度などのほかにいろいろな要因が働らくので、面接調査とは全く別の調査となつてしまう。この点については統計数理研究所輯報第4号でふれておいた。

### 2.3 訪問回数と関係のある項目

〔国民性〕の項目は全部で58あるが、これらの各項目はいくつものカテゴリをふくんでいるのでカテゴリの数は全項目を通すと、約300に達している。訪問回数の多少によつて、この各カテゴリに対する回答のあらわれ方にちがいがあがるかどうかという問題がある。〔国民性〕の場合、1回の訪問で調査できたサンプルは1267人で、2回以上の訪問をすることによつて調査できたサンプルは987人であつた。そこでサンプルをこの2つのグループに分けてみた。そうして、グループごとに各項目での各カテゴリのあらわれる%を計算し、両グループの%の差が、サンプリング誤差より大きいかどうかを検定した。ただし、この検定はつぎのようにした。

ある項目のあるカテゴリでの、両グループの % をおのおの  $p_1, p_2$  とし、全サンプルのそれを  $p$  とする。このとき、

$$|p_1 - p_2| > 2\sqrt{p(1-p) \times \left(\frac{1}{1267} + \frac{1}{987}\right)}$$

であれば、両グループの間に差があるとした。

この結果は、第 17 表のようになる。すなわち、差の出たカテゴリは約 300 のうちのわずか 10 項目の 14 カテゴリにすぎない。しかもそのうち 10 カテゴリはフェース・シートであり、調査のネライであつた質問の中ではわずか 4 カテゴリにすぎない。しかも、これらの性別をのぞく 12 カテゴリは性別や支持政党でも差がでる。このことは、訪問回数に深い関係のある性別によるゆがみによつて引き起こされたとも考えられるが、さらに支持政党による差と質問の内容を考えあわせると、open-closed な態度も影響しているようである。

第 17 表 訪問回数により差のあるカテゴリ [国民性]

項 目	続 柄		性 別		結 婚		学 歴	産 業	新 聞	
	世帯主	妻	男	女	未婚	既婚	小卒以下	無職	よくよむ	よまない
1 回 訪 問 $p_1$	32.6	33.5	40.4	59.6	13.8	72.0	38.8	38.8	52.6	22.8
2 回 以 上 $p_2$	40.5	22.7	55.4	44.6	21.3	66.7	25.2	25.2	60.5	13.9

項 目	Q22 競 輪	Q34 自 然	Q38 労 組	Q50 赤穂浪士	サンプル数
カ テ ゴ リ	やむを得ない	を利用	無答	あの時はよい	(100.0%)
1 回 訪 問 $p_1$	19.7	37.9	13.2	30.8	1267
2 回 以 上 $p_2$	26.6	45.7	7.4	37.0	987

これらの事実から、つぎのような予想ができるが、どれが妥当であるかただちに決定はできない。

1° 差の出たカテゴリは、300 のうちの 14 にすぎないので、信頼度 95% では当然こういうことは起りうる。そうだとすれば、全体のサンプルの % をみるかわりに、1回の訪問で調査できたサンプルだけを集計してもよいことになる。特に、上で述べたような性別や、open-closed な態度で、バランスをとれるように注意すれば、十分であるということになる。

2° 性別・学歴などのフェース・シートで差が出ても、調査のねらいであつた質問項目で差があまりでなかつたことは、フェース・シートが調査のねらいの質問内容との関係が薄いことを示している。

3° (2° の対隅) われわれの質問内容がフェース・シートと深い関係があるはずとすれば、サンプルの答の信頼性 (reliability) がうすいことを示している。

### §3 調査員によるインチキ

#### 3.1 インチキの発見法とそのインチキ率

調査員が指定されたサンプルに面接をしないで勝手に調査票をデッチ上げたもの、すなわち故意におこなつた不正を調査員のインチキということにする。調査員はこのほかに、不正をおこなう意志がなくても、あやまりをおかすことがあるが、これについては §4 で述べる。

まず、どのくらいのインチキがされるかをみよう (第 18 表)。これによれば、回収した調査票のうち数 % はインチキであり、調査員の 10% 近くがインチキをすることが予想される。

しかも調査員の中には、受持サンプルのうち、いくつもインチキをするものがあるが、あとから

のインチキ発見のための調査は、一部分をえらんでいるので、はつきりしないが、1人で3~4部ものインチキ調査票をつくつたものもある。

第18表 調査員のインチキの率

調査名	国民性 準備	国民性	張り込み	都知事I	都知事II	効果
	インチキ発見方法	郵便	郵便	張り込み	再調査	再調査
対象としたサンプル数	136	371	32	?	?	420
そのうちのインチキ率	6.6	11.0	3.1	約5	約5	5.2
対象とした調査員数	85	239	8	118	143	80
そのうちのインチキ率	7.1	15.9	12.5	12.7	6.3	21.5

つぎに、インチキ発見の方法をみよう。

1° 郵便調査 面接調査がすんでから、郵便(往復ハガキ)で面接をうけたかどうかを尋ねる方法である。この方法によると、サンプルが故意または忘れてしまつてウソをついてしまつたことまで、インチキとしてしまうことがある。

このような、サンプルの方のウソ(事実とのくい違い)のていどをしらべるために、絶対に信頼のできる——研究所の職員で、あとからこのような郵便調査をすることを承知していた——調査員が、面接したサンプルにこの郵便調査を試みた(第19, 20表)。

第19表 職員の訪問したサンプルの郵便調査〔効果〕

面接	郵便		
	発信	持戻り	返信
面接した	121	0	100*
調査不能	18	4	5

第19, 20表の\*は対応する。

第20表 左表のうち面接したサンプルで返信のあつたもの

調査員来ない	0
途中まででやめた	1
調査票をおいていつた	1
調査を自分でうけた	98
計(面接した, 返信)	100*

この表から分かるように、面接できたサンプルへの郵便調査用ハガキが持戻り(移転先不明, 尋ね当らずなど)になつたものはなく、84%の100通の返信があつた。また調査不能(面接できなかつた)の18サンプルのうち4通は、ハガキが届かずに持戻られたし、返信のあつた5通はいずれも、調査できなかつたことはやむを得ないと思われることが記入されていた。

さらに面接したサンプル中、返信のあつた100通のうち、‘途中までしかうけなかつた’‘調査票をおいて行かれて、記入した’というのが1人づつあつたが、調査員に問い正してみても、そのような事実はなく、サンプルの誤解と考えられる。

すなわち、心配をしたような、郵便調査に関するサンプルのウソは、返信のうちの2%位にすぎない。そこで、第18表のようなインチキ率も多少わり引いて考えなければならぬにしても、ほとんどかわりないであろう。

2° 再調査 パネル調査などで一定期間ののちに、再調査したときに、前回面接をうけたかどうかをしらべる方法である。ただし、この場合、後の調査員の始めの挨拶の中で、‘この前別の調査員がうかがつたはずですが、たびたびおじやまをして、……’というようなことをいつたとき、サンプルの方から、‘来なかつた’とか、‘私は会わなかつた’というような返事のあつたものを、前の調査員がインチキをしたと考えた。



この方法では、サンプルのウソのほかに後の調査員のインチキも考えられるが、ともにごくわずかであろう。

3° 張り込み調査 これは §4.2° のべるように、研究所の職員をニセのサンプルにしておき、これを訪問させる方法である。この方法は確実にに訪問したか否かがたしかめられるし、そのほか質問をとばすとか、インストラクションを守つたか否かがはつきりとわかる。

### 3.2 インチキ調査票の分析

インチキな調査票というのは、上で述べたように、サンプル本人の意見や態度でないものが、記入されているわけである。

われわれがこれらのインチキをされたサンプルを追求してみた経験から判断すると、全く本人に会おうとしないで‘作文’をしてしまつたものは少ない。多くのインチキ調査票は始めの訪問ではサンプルに会えずに、2回目あたりで、そのサンプルの家の人から、サンプルの学歴、職業などをきき出し、あとは調査員がかつてに‘作文’したか、その家の人に‘サンプルならどういふでしょうか?’というような聞き方をしていることが多い。

では、インチキ調査票はどうつくられているかを、[国民性]についてしらべてみよう。

この[国民性]のインチキは郵便調査法で発見したのであるが、郵便を発信したのは、回収したサンプルの  $\frac{1}{10}$  であつた。ところでこの郵便調査に返事をくれたのは 82% に達した。そこで返事をくれたものの中で発見できたインチキの5倍をもつて、推定インチキ調査票とする。

まず、インチキ調査票となつたサンプルの性別と年齢別をみよう(第21表)。これはサンプリングのとき、住民票で分かっているのである。インチキされる率というのは、各カテゴリーの全サンプル数に対する推定インチキ調査票の数の比率である。

第21表 性別・年齢別のインチキをされた数、およびインチキをされた比率 [国民性]

	男	女	計	20~24	25~29	30~34	35~39	40~49	50~
インチキされ数	20	17	37	6	4	5	6	10	6
インチキされ率	9.5	7.0	8.0	6.5		11.0		8.0	

すなわち、性別や年齢別にみて、どういうサンプルがとくにインチキされやすいというようなことは、これだけの結果からはいえない。

つぎに、インチキ調査票の‘作文’の影響をみるために、[国民性]の質問の中から、質問形式や質問の内容からいつて、代表的な質問について、2つの方法で検討をしてみた。

第1の方法 郵便調査で判かつた 37 枚のインチキ調査票を、われわれは、集計からのぞき調査不能の扱をした。しかし、上で述べたように、発信したのは  $\frac{1}{10}$  にすぎなかつたので、回収サンプルの中にはまだインチキ調査票が入つているはずである。その未発見のインチキ調査票も、発見されたものと同じ傾向を示すものと考え、しかも少し多目に見つるために、調査不能を考へてウェイトをつけた修正サンプル (cf. 1.3) の反応から、発見されたインチキ調査票の反応の5倍をとり、つぎのような  $\chi^2$  検定をおこなつてみた。

$i$  はカテゴリーをあらわし、 $A_i, B_i, \dots, F_i$  はそれぞれのカテゴリーの度数をあらわす。

$A_i$ : 修正サンプル

$B_i$ : 発見したインチキ・サンプル

$C_i$ : 推定インチキ・サンプル

$D_i = A_i - C_i$ : 修正サンプルから推定インチキ・サンプルをのぞいたとき

$E_i = D_i / \sum D_i$ : 同上比率 (%)

$F_i = (\sum A_i) \times E_i$ : 集計用サンプルの  $D_i$  にとずく推定

$$\chi_0^2 = \sum_i \frac{(A_i - F_i)^2}{F_i}$$

これをカテゴリの数より1少ない自由度で検定をした。

このことを, 問 10 [女の仕事] で例示しよう。(第 22 表)

第 22 表 計算例 (女の仕事) [国民性]

Category	好ましい	好ましくない	その他	計	$\chi_0^2 = \sum_i \frac{(A_i - F_i)^2}{F_i}$ $= 0.707$ $D.F. = 2$ $Pr(\chi^2 \geq 0.707) \approx 0.70$
$A_i$	1425	550	279	2254	
$B_i$	22	12	3	37	
$C_i = B_i \times 5$	110	60	15	185	
$D_i = A_i - C_i$	1315	490	264	2069	
$F_i$	0.636	0.237	0.127	1.000	
$F_i = 2254 \times E_i$	1434	534	286	2254	
$A_i - F_i$	- 9	16	- 7	± 16	

さて, このような方法で検定した結果は, 第 23 表のように, どの質問でも,  $A_i$  と  $F_i$  の間に有意な差はあるとはいえない。すなわち, インチキ調査票は全体の集計に影響を与えないと考えてよいであろう。

第 23 表 インチキ調査票の影響 [国民性]

Q	内 容	$\chi_0^2$	D.F.	$Pr(\chi^2 \geq \chi_0^2)$
2	住所に満足か	0.066	3	$Pr \geq 0.99$
10	女が仕事をする事	0.707	2	$Pr \approx 0.70$
14	首相の伊勢神宮参拜	5.778	6	$Pr = 0.30 \sim 0.50$
25	日本人と西洋人はどちらがすぐれているか	5.233	5	$Pr = 0.30 \sim 0.50$
36	世話になつた人に, もつて行くカシはどこで買うか	1.886	3	$Pr = 0.50 \sim 0.70$
42	父の死のとき, 重大な用があつても帰るか	1.798	2	$Pr = 0.30 \sim 0.50$

第2の方法 上記の質問のうち, 問2と問10は, 郵便調査のときも質問した。そこで推定インチキ調査票を, この郵便調査の結果で修正してみる。この方法はつぎのようにした。

$i$  はカテゴリをあらわし,  $A_i, B_i, \dots, I_i$  はそれぞれのカテゴリの度数をあらわす。

$A_i$ : 回収サンプル

$B_i$ : 推定インチキ・サンプル

$C_i = A_i - B_i$ : 正しいサンプルの推定

$D_i$ : 郵便返信結果による %

$E_i = \{(\sum G_i) - (\sum C_i)\} \times D_i$ : 調査不能と推定インチキの郵便調査による推定

$F_i = C_i + E_i$ : 同上による修正集計用サンプル

$G_i$ : 集計用サンプル (Cf, §1.3)

$$\chi_0^2 = \sum_i \frac{(F_i - G_i)^2}{F_i}$$

この結果は, 第 24 表のようになり,  $F_i$  と  $G_i$  は同じ分布でないことを示してしまう。しかし

このことは、郵便調査で面接調査を推定するという、調査法のちがいもあるので、インチキの影響が大きくて、集計用サンプルにくるいがおきたのかどうかはつきりしない。 $F_i$  (郵便調査による修正集計用サンプル) と  $G_i$  (集計用サンプル) の % をみれば、その差は非常に大きくはないことが判かる。(第 25 表)。

第 24 表 調査不能およびインチキ調査票の郵便調査による修正の影響〔国民性〕

	$\chi_0^2$	D.F.	$P_r(\chi^2 \geq \chi_0^2)$
問 2	40.30	3	$P < 0.005$
問 10	32.06	2	$P < 0.005$

第 25 表 調査不能およびインチキ調査票の郵便調査による修正〔国民性〕

問 2	満 足	時による	不 満	その他	計
$F_i'$ 修正集計用サンプル	70.1	9.8	12.7	7.4	100.0
$G_i'$ 集計用サンプル	75.4	6.6	12.0	6.0	100.0

問 10	好ましい	好ましくない	その他	計
$F_i'$	63.9	20.5	15.6	100.0
$G_i'$	63.4	24.4	12.2	100.0

§ 4 調査員の記入のあやまり

ここでとりあげることは、調査員の記入のあやまりであるが、これは誤解とミス・コードの両方をふくめる。これをしらべるために、つぎの 2 つの方法をもちいた。

1° 同時調査 一人のサンプルの答を数人の調査員が同時に、調査票に記入する。この方法では調査員自身が実験をしている——調査をされている——ということ意識してしまうので、ふつうの面接調査のときより、あやまりは少ないであろう。

第 1 には〔国民性〕の前調査のとき、調査員 (学生) が余つたので、この学生と、そうとうに調査になれている研究所の職員 (女) とに対しておこなつた同時調査について述べよう。

調査の方法は、他の実際に面接調査をする調査員と同時に、調査の指示を与えてから別室に集め、あらかじめ調査の場面を録音しておいたものを聞かせて、調査票に記入をさせた。このため質問は選択技法のものに限ぎつた。すなわち、答のミス・コードをしらべるのが、この目的であつた。

調査をうけた調査員は 18 人である。テープにはスクリプトに従つて 2 人のサンプルの分を、吹込んである。この 2 人のサンプルはそれぞれ 41 の質問をされているが、このうち 14 題は共通の質問である。そこで同時調査をうけた質問の数は 70 題 (= 41 × 2 - 14) であるが、調査員の記入は、各人 82 回 (= 41 × 2) である。

まず、各調査員の記入がどのくらい正しかつたかをみよう (第 26 表)。すなわち全記入 (82 回)

第 26 表 正答率別の調査員度数分布〔国民性〕

A 正しい記入	82	81	80	79	78	77	76	75	74	73	72	71	...	65	計
B 正 答 率	100.0	98.6	97.5	96.4	95.1	93.8	92.6	91.4	90.2	88.0	87.8	86.5		78.2	
C 調査員数	0	0	1	0	1	1	1	4	4	1	2	2		1	18

ただし B = A/82

が正しかつたものはなく、一番正しかつたものでも 82 回の 97.5% に当る 80 記入であつた。一番まちがつたものも 78.2% に当る 65 記入はあつていた。18 人の調査員の正答率の平均は 78.2% で、82 記入のうち 73.9 記入に当る。なお標準偏差は 3.1 記入となる。

こんどは、質問の方からながめてみよう (第 27 表)。すなわち、全調査員の記入が正しかつたものが約半分の 43 題であり、 $\frac{1}{4}$  以上の調査員がまちがつて記入したのは 10 題にすぎない。

第 27 表 正答率別の質問の度数分布〔国民性〕

A 記入の正しかつた調査員の数	18	17	16	15	14	13	12	11	10	9	8	計
B 正しい記入率	100.0	94.4	88.9	83.8	77.8	72.2	66.7	61.1	55.6	50.0	44.4	90.8
C 質問の数	43	11	7	4	7	2	1	0	1	3	3	82

$$B = A/18$$

もともとここにとりあげたのは、〔国民性〕の準備調査で、全国での本調査のときは質問をへらさなければならなかつた。そこで正答率の低かつた質問を捨てたり、修正したりしたが、こういう分類によると正答率は第 28 表のようになる。

第 28 表 準備調査の質問の分類別正答率〔国民性〕

本調査のとき	そのまま使つた質問	修正して使つた質問	小計
質問数	34	16	50
調査員の平均正答率	95.6	83.3	91.7
調査員の正答の標準偏差 (%)	7.9	17.1	13.0

ただし、調査員は 18 人。

すなわち、これから考えると本調査ではおそらく各質問とも調査員の 95% 前後が正しく記入して来たであらう。

さて、上でも述べたように、14 質問は 2 人のサンプルの録音がある。もちろん、その答は同じではない。この質問が 2 人のサンプルでどう記入されたかをくらべてみよう (第 29 表)。すなわち、両サンプルとも——おそらく、いつでもといいかえることができよう——正しく記入されにくいというような質問は余りない。一方だけ記入がまちがうというのは、この録音が極めてイジワルなスクリプトに基づいているからである。そこで全質問を通じて、実際の調査の場合は正答率は、これより高いと期待できる。

第 29 表 2 人のサンプルの正答数による共通質問の分布〔国民性〕

第 1 サンプルの答を正しく記入した調査員数	18	18	18	18	18	18	13	17	14	9	計
第 2 サンプルの答を正しく記入した調査員数	18	17	16	13	10	9	18	14	17	16	
上記のように正しく記入された共通質問の数	4	2	1	1	1	1	1	1	1	1	18

第 2 に「上野言語」の場合をみよう。このときは、調査員は国立国語研究所の所員 3 人と筆者の 4 人である。4 人ともふつろの調査や言語調査を何回もおこなつており、しかもお互いに気心も知れあつているし、年齢も同じ位であり、調査員をこれほどそろえることはめつたにできないといえるであらう。

その同時調査は、1人がサンプルを調査するのを、他の3人も横で聞いていて記入する方法によつた。

まず、多肢選択法の質問について、4人の調査員の関係をみよう。選択肢はだいたい一定の順にならべてあるので、そのコードをスコアとして相関係数を計算してみると、第30表のようになり、各調査員のコーディングはよく一致している。

またサンプルの答を速記する質問は、同時に速記したので、これと各調査員の速記が、どのくらいあつているかをしらべてみた(第31表)。この結果は約70%

しか正しくなかつた。そこでこのような、こまかなところまでねらうのには、どうしても録音器を必要とすることになる。しかし、各調査員の速記から判定した敬語の段階と、録音から判定した段階とをくらべると、第32表のようになり、各調査員とも95%近く一致している。

第30表 調査員間の相関係数  
(多肢選択法)〔上野言語〕

	K	U	N <sub>o</sub>	N <sub>t</sub>
K		0.95	0.89	0.92
U	0.95		0.89	0.91
N <sub>o</sub>	0.89	0.89		0.98
N <sub>t</sub>	0.92	0.91	0.98	

K, U, N<sub>o</sub>, N<sub>t</sub> は調査員  
反応は延 55

第31表 速記式回答の正しさ〔上野言語〕

		調査員				計
		U	K	N <sub>o</sub>	N <sub>t</sub>	
注敬語 目文節 す節の べう きち	まちがつて記入した文節 数率	7.3	7.3	8.7	9.3	8.2
	記入もれの文節数率	23.3	18.7	24.0	17.3	20.8
	(以上の小計)	(30.6)	(26.0)	(32.7)	(26.6)	(29.0)
	正しい記入の文節数率	69.4	74.0	67.3	73.4	71.0
	合計	100.0 150	100.0 150	100.0 150	100.0 150	100.0 600
かつてにつけくわえた敬語文節数		1	1	0	1	3

第32表 速記回答と録音の判定〔上野言語〕

		調査員				計
		U	K	N <sub>o</sub>	N <sub>t</sub>	
録音と速記が 同じ敬語段階と判定できる反応		36	36	35	35	142
不明		1	1		1	3
ちがう敬語段階と判定される反 応				2	1	3
合計		37	37	37	37	149

2° 張り込み調査 われわれの関係者(研究所職員)をニセのサンプルにしたてておき、調査員に‘このサンプルを調査しろ’と命ずる。ニセのサンプルの答はあらかじめきめておくが、調査員のきき方によりきめた通りにはゆかないので、ニセのサンプルのほか、別の関係者をカゲにかくしておき、答を記入させる。

この方法の調査を、〔国民性〕のときおこなつた。調査員は8人、ニセのサンプルは7人にしておいた。これが調査員に気付かれなかつたことは、§3.1のインチキでも述べたように、調査員の1人がニセのサンプルを1人調査しないくせに、記入した調査票を持つて来たことでもわかる。

この結果、調査員の記入と、カゲの人の記入の一致のぐあいをみると、第33表のようになる。質問形式の‘リストなし’というのは、サンプルの答を調査員が判断をして、あらかじめ印刷さ

第33表 調査員とカゲの人の記入の一致率

質問の形式		質問数	延質問数	記入の一致率
リストなし	2肢選択	9	288	82.6
	3肢選択	17	544	87.9
	4肢以上	8	256	83.1
	計	34	1088	85.5
リスト使用	2肢選択	6	192	84.4
	3肢選択	6	192	85.9
	4肢以上	4	128	80.5
	計	16	512	84.0
自由記入法		16	414	88.6

第34表 あやまり率

	数	平均	標準偏差
調査員別	8人	19.4	4.8
ニセ・サンプル別	7人	17.6	9.3

第35表 あやまり率(質問形式別)

	延質問数	あやまり率
リストなし	1216	21.2
リスト使用	512	22.1
自由記入法	366	11.5
計	2094	19.7

つぎに調査員の態度の印象を、ニセ・サンプルとカゲの人が判断した結果をみよう(第36表)。

第36表 調査員の態度

カテゴリ	リストの見せ方			まふふ		きふ		調ふた			にるふぶ			早ふの			
	よ	渡	不	じ	つ	ど	つ	査	つ	た	こ	こ	す	つ	み	つ	ろ
	い	し	明	め	う	め	や	レ	う	い	こ	こ	す	つ	み	つ	ろ
回	23	8	1	21	11	0	1	0	22	10	16	14	2	5	21	6	

カテゴリ	リふ		能ふこ			礼ふ無			感ふ理		
	ク	つ	つ	ば	儀	つ	情	つ	性	的	的
	ツ	ヤ	弁	う	な	し	う	礼	的 <td>的 </td>	的	
回	0	32	5	21	6	13	19	0	1	29	2

これも調査員8人が、ニセ・サンプル1軒を訪問するたびに判断された結果である。この結果をみても、調査員の態度は大体よかつたといえよう。

れた、選択肢にチェックする方法である。リスト使用というのは、選択肢のリストをサンプルに見せて、どの選択肢に当るかを、サンプル自身に指し示めさせる方法である。自由回答法というのは、サンプルの答の要点を記入してくる方法である。

延質問数というのは、8人の調査員が約4人のニセ・サンプルを調査しているので、それと質問数の積である。

一致率は、調査員とカゲの人の記入が一致している場合の延質問数に対する比である。この結果、一致率はどれでも85%前後である。

ここで注意しなければならないことは、ニセ・サンプルの答は非常に複雑な、シツコイものにしてあつたので、一般のサンプルの場合は恐らく90%以上の一致率が期待される。10%ぐらいのあやまりは、調査員の過失と考えられる。

つぎに、調査員がインストラクションを守つたかどうかをしらべてみた。細いことは省略して、誘導尋問、調査票を見せる、質問文の一部をいわない、勝手な判断などの許せない誤りをおかした回数を、延質問に対しての%でみることにしよう。

調査員別では1人が平均20%近いが、ニセ・サンプル別では18%位である。質問の形式別にみると、自由記入法が一番少ない。

## §5 調査員の種類

ここで調査員の種類というものは、広い意味であつて、性格、意見、経験のありなし、学生か否かなどの分類をさすことにする。この§で述べる例はいずれも〔国民性〕の準備調査のときにおこなつたものである。

## 5.1 調査員の性格

調査員の性格 (personality) は、調査に大きな影響を与えるであろうということは、十分に予想されることである。そこで、われわれは調査員に対して、性格検査をおこなつた。

この性格検査は、新に作つたもので、後にはこれを改良して、岡崎市での敬語の社会心理学的研究のときにも使つた。一口でいえば、The R-S Reaction study (by G. W. and F. H. Allport) や The personality inventory (by R. G. Bernreuter) などに似た形式のもので、選択肢法自記式のペーパー・テストである。

われわれは、Ascendancy (出しやばり性)、Neurosis (神経性)、Cyclothymia (躁鬱性) の3つのスケールしかつくりすることができなかつた。

一方、調査員のよいか・わるいかの判定を、不能率、平均訪問回数、集合時刻、回収したときの総合評価などの面からみるために、それぞれ相関係数を計算してみた。この結果 (第 37 表) のように、ほとんど関係がなかつた。

しかし、このテストで2つ以上のスケールで極端な傾向を示す数人の調査員をみると、満足できない態度を示していた。また性格調査そのものにも問題があるので、さらに、くわしい研究が必要である。

なお、サンプルと調査員の接触時の態度は §4 の 2° で少し観察してある。

第 37 表 性格検査 (相関係数) [国民性]

	Ascendancy	Neurosis	Cyclothymia
a 調査不能率	0.015	-0.132	0.200
b 平均訪問回数	-0.003	0.111	-0.127
e 集合時刻	0.096	0.008	-0.408
f 評価	0.049	0.020	-0.065

## 5.2 調査員の意見など

調査員の意見や態度やイデオロギイなどが、調査票への記入や、サンプルの答に対して、ゆがみを与えはしまいかということも問題となる。

これをしらべるために、調査にさきだつて、調査員自身に、サンプルを調査するのと全く同じ調査をしてみた。これによつて、調査員の意見が調査におよぼす影響をみることにした。

まず、調査員の支持政党と、その調査員のしらべたサンプルの支持政党がどのくらい一致しているかをみた。その一致の度合を、記号的に定義すると：

$$O = \frac{1}{\sum_i n_i} \sum_i \sum_{s=1}^{m_i} \delta_{is}$$

ただし  $\delta_{is}$  は調査員がしらべたサンプル  $s$  の支持政党と、 $i$  自身の支持政党が一致すれば 1、そうでなければ 0 をとるとする。 $n_i$  は調査員  $i$  のしらべたサンプルの数である。

この結果は  $\sum_{i=1}^{94} n_i = 245$  に対して、 $O = 17.9\%$  であつた。この支持政党の一致というのを多少ゆるめて、保守派か革新派かさえ一致していればよいとすれば  $O = 37.1\%$  である。

いずれにしても、調査員の意見はサンプルの意見にゆがみを与えているとは考えられない。

こんどは、一般の質問の調査員の意見の影響を、つぎのようなインデックスでしらべてみた。

$$H_i = \frac{1}{n_i \times q_i} \sum_{Q=1}^{12} \sum_{s=1}^{n_i} \delta_{is}^*(Q)$$

ここに、 $i=1, 2, \dots, 94$  は調査員をあらわし、 $n_i$  は調査員  $i$  が面接したサンプル数、 $q_i$  は調査員  $i$  がハッキリとした意見をのべた質問の数、 $Q$  は質問番号をあらわす。

$\delta_{is}^*(Q)$  は  $Q$  という質問で、調査員  $i$  がハッキリと意見を述べていて、その意見と調査したサンプル  $s$  の意見が反対であれば1をとり、そうでなければ0をとるものとする。

このインデックス  $H_i$  を 94 人の調査員について計算すると、

$$\bar{H} = \frac{1}{94} \sum_{i=1}^{94} H_i = 35.9\%$$

$$\sigma_H = \sqrt{\frac{1}{94} \sum_{i=1}^{94} (H_i - \bar{H})^2} = 16.6\%$$

となる。

この結果からも、調査員の意見は調査に影響をおよぼしているとは考えられないといえよう。

### 5.3 調査の経験と学生か否か

調査員に経験が有るか無いかによつて、調査にちがひがあるか、また調査員が学生かどうかによる影響はどうかをしらべるために、同じ調査地点に、これらの4種類の調査員を入れて調査してみた。ただし、経験のある一般(非学生)の調査員をうまく集めることができなかった。

まず、経験のあるなしによつて、各質問の各カテゴリへの記入のしかたがちがうかどうかをみるために、両グループのつけた各カテゴリの%を、つぎのようにしてしらべた。ある質問を経験ある調査員がしらべたサンプルの  $p_1$  % があるカテゴリであつたとし、経験のない調査員では  $p_2$  % だつたとする。また経験ある調査員がしらべたサンプル総数を  $n_1$  人、経験ない調査員の方は  $n_2$  人とする。

$$p = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2}$$

とおき、

$$|p_1 - p_2| > 2 \sqrt{p(1-p) \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}$$

となるカテゴリは全部で100のうち11であつた。これはもし、調査員の経験は問題がないと仮定すると、11%のエラーとなるので、経験によつて少しは差があると考えたがよいであろう。

ちなみに、全サンプルをひとりおきに2組に分けたときの両組の結果を、上記の  $p_1, p_2; n_1, n_2$  としてみると、エラーは3%であつた。

つぎに同じようなことを学生か否かと経験とについてみる。このとき計算を機械的にするため、上記の  $p_1, n_1$  は調査員のある種類についてであるが、 $p_2, n_2$  には全サンプルの価をとり、しかも  $p = p_2$  とした。

この結果は第38表のようになり、経験のない一般(学生でない)調査員がやや多いが、それほど大きな差はみられない。

第38表 有意差のあのカテゴリ

調査員	有意差あるカテゴリ				計
	経験・学生	未経験・学生	経験・一般	未経験・一般	
有意差あるカテゴリ	5	3	1	8	4

つぎに、これらの調査員を比較するために、よいわるいの判定の基準として、調査不能の率、平均訪問回数、調査員自身と反対の意見にどれだけチェックしてきたか、調査員自身と同じ支持政党



にどれだけチェックしてきたか、集合時刻への集まりぐあい、総合評価、インチキをした数について、しらべてみた(第 39 表)。

第 39 表

調査員	学 生		一 般		学 生	一 般	経 験	未 経 験	全 員
	経 験	未 経 験	経 験	未 経 験					
調 査 員 数	29	34	18	36	63	31	47	47	94
調 査 不 能 率	15.2	22.4	23.9	41.5	19.1	31.3	18.5	27.7	23.1
平 均 訪 問 回 数	1.7	1.5	1.6	1.4	1.6	1.5	1.7	1.5	1.6
$\bar{H}$ (反対意見)	33.2	35.3	40.9	38.3	34.3	39.6	35.6	36.2	35.9
O(同じ政党)	21.6	7.5	40.0	18.8	15.2	25.0	24.8	11.7	17.9
集 合 時 刻*	37.9	26.5	68.8	76.9	31.7	72.4	48.9	40.4	44.6
綜 合 評 価	4.6	4.9	3.7	2.7	4.7	3.3	4.3	4.3	4.3
イ ン チ キ	4	1	2	0	5	2	6	1	7

\* 遅刻者の %

サンプルを調査できなかつた率は学生より一般が多く、また経験者より未経験者の方が多い。とくに一般の未経験者は 40% 以上も調査不能であるが、学生の経験者は 15% にすぎない。

サンプルを何回ぐらい訪問しているかを平均でみると、調査不能のときと同じ関係があり、調査不能が多いということは、追求の努力がたりないことを示している。

反対意見率は学生より一般が多い。

同じ政党率は学生より一般が多く、経験者より未経験者が少ない。とくに学生の未経験者は 7.5% できわめて少ないが、一般の経験者は 40% で多い。

集合時刻に遅刻したものは、学生は少ないことが目立っている。

総合評価も学生の方がよい。

インチキは、学生の方が多いが、これはサンプルが少ないのでハッキリしたことはいえない。

以上の結果から、調査員としてより順位をつけると、第 40 表のようになる。これについて順位の平均を出し、また順位相関係数を計算してみた(第 41 表)。

第 40 表 調査員の種類別のよい順位

調査員	学 生		一 般		‘よい順位’の き め 方	
	経 験	未 経 験	経 験	未 経 験		
a	調 査 不 能 率	1	2	3	4	小さいほどよい
b	平 均 訪 問 回 数	1	3	2	4	大きいほどよい
c	$\bar{H}$ (反対意見)	4	3	1	2	大きいほどよい
d	O(同じ政党)	3	1	4	2	小さいほどよい
e	集 合 時 刻	2	1	3	4	早いほどよい
f	綜 合 評 価	2	1	3	4	高いほどよい
g	イ ン チ キ	4	2	3	1	少ないほどよい

第41表 平均順位と順位相関係数(W)

調査員	学 生		一 般		W	Pr
	経 験	未 経 験	経 験	未 経 験		
a ~ g	2	1	3	4	0.143	0.40
a, b, e, f	1	2	3	4	0.773	0.01

$$W = \frac{12 s^2}{m^2(n^2 - n)}$$

この7項目全部の順位相関係数は 0.143 で, しかもこのようなことの起る確率 Pr

$$Pr = Pr\{s^2 \geq m(n-1)W\}, \quad D.F. = n - 1$$

でみると, Pr = 0.40 であるから, 余りよく一致しているとは考えられない。

しかし, 定義のはつきりしなかつた  $\bar{H}, O$  とサンプルの少なかつたインチキをのぞいた4項目については, 順位相関係数は 0.773 で, しかもこのようなことの起る確率 Pr

$$Pr = Pr\{s^2 \geq s_0^2\}$$

から計算すると, Pr = 0.01 である。すなわち, よく一致しているといえる。

そこで, 4項目の平均順位をみると, 学生の経験者, 学生の未経験者, 一般の経験者, 一般の未経験者の順となる。

### §6 調査員に原因のないゆがみ

いままでみてきたことは, 調査員に原因のある面接調査のゆがみであつた。ここでは, 調査員には全く(といつてよいほど)責任がなく, サンプルの方にその原因のあるゆがみをしらべることにする。ただし, サンプルに原因するゆがみがあるということは, 調査の妥当性そのものをおびやかす危険がある。このことは, 調査法の問題として, 適当な機会にくわしくふれるつもりである。

#### 6.1 サンプルの答の信頼性

サンプルの答がいかげんなものでないか, すなわち信頼できる (reliable な) ものであるかどうかということは, 面接以外の調査でも重大な問題である。

この信頼性は一番かんたんには, 全く同じ方法で再調査をし, 前と後の調査を比べることである。この方法による結果を, [効果] でのべよう。前調査というのは 1954 年 3 月, 後調査というのは同じ年の 10 月末である。ここでとりあげた質問は, 第 42 表のとおりであるが, 前後の調査の間で, この質問の答の分布がかわるような, 変化があつたとは考えられない。それにもかかわらず

第42表 信 頼 性 [効果]

問 あなたは, 自分が正しいと思えば世のしきたりに反しても, それをおし通すべきだと思いますか, それとも世間のしきたりに従つた方がまちがいないと思いますか?

後	前				計	後の %
	おし通せ	場合による	従がえ	その他		
おし通せ	○57	27	×19	7	110	36.6
場合による	35	○28	17	7	87	28.9
従がえ	×25	32	○30	8	95	31.5
その他	5	1	3		9	3.0
計	122	88	69	22	301	100.0
前の %	40.6	29.2	22.9	7.3	100.0	

前と後の答の分布を検定すると、同じ分布とは認められない。しかも前と後の答を個人的にみれば全く同じ答のもの（○印）は115人で38.2%にすぎないし、全く反対（×印）は44人で14.6%におよんでいる。

一方、この前と後の調査の間で政治情勢は大分変わったのであるが、支持政党をみると、つぎのようになる。

全く同じ政党をあげたものは92人である。これは前調査ではつきり支持政党をあげた195人に対して47.2%にあたり、後調査の171人に対しては53.8%にあたる。

政党を保守・革新の2派に分け、前後とも同じ派のものおよび、前後とも支持政党のなかつたものは163人で全体（301人）の54.1%に当る。

この2派の反対派に移つたのは24人で、全体の8.0%に当る。

以上のように性質の違った問しか検討できないが、信頼性は必ずしも高いとはいえない。しかし吉田内閣がついに総辞職に追い込まれる時期（約1ヶ月前）においても、保守・革新の反対派への移動は8%であることは、注目されてよいであろう。

## 6.2 面接調査の妥当性

調査にさいして、信頼性ととも、妥当性（validity）が問題になる。これは測かろうとしたものをはかっているかどうかということである。

われわれは、〔岡崎言語〕で敬語の使われる場をいろいろ考え、そのおのおのに対応し、しかも、‘こんな場合、あなたはなんといいますか？’という形の質問にできるものだけを面接調査でとりあげた。しかもこれらの質問についても、面接調査で答えたものと、ふだん実際にそのような場での話し方が、ちがつていたのでは、実状をしらべたことにはならない。そこで、われわれの面接調査の質問の答が、どれだけ実状をとらえているか実際の場と対応をつけてみた。

その方法は、スナップ録音（気づかれないように、録音をとる）と、調査票とをつけあわせることによつた。

スナップ録音ができる、つぎの4つの場に対して男女6人ずつのサンプルに調査をした。

- A 電報局で電報用紙をもらう。
- B 道で昔ならつた先生又は友人に会う。
- C 道で見知らない旅人に市役所へ行く道をきかれる。
- D 買いつけの店で荷物をあずける。

6人のサンプルは年齢、階層、職業などとB、Dで共通の人や店の使かえる範囲から、有意的に選んだ、またA、Dという場では、‘こうしてくれ’と頼んでおいたが、B、Cは全くサンプルには予期されていないし、この場面を大部分のサンプルは調査であることに気が付かなかつた。

さて、この結果は、敬語の段階付けがまだ未完成であるが、面接と録音とでは全く同じような敬語が使われていたと判断をすることができた。ただAでは録音の方が多少修飾句がつくし、Dでは録音の方は会話となるので、店主の気サクな会話に引き込まれているようである。

しかし一言でいえば面接調査はうまく実状をつかんでいると判定できるようである。

## 6.3 サンプルのつくウソ

つぎに、これも一種の妥当性をしらべることになるが、サンプルが正直な答をするかどうかは、大きな問題である。しかし態度や意見の調査では、述べられたことは、たとえ論理的にいくちがうことがあつたとしても、ウソであると断定することはできない。そこでサンプルのいつたことがウソであるかどうかは、過去の記録されている事実を、面接によつて尋ね、対応をつけてみるよりほかにない。

われわれは、選挙のあとで、投票したかどうかを尋ね、選挙人名簿とつけあわせてみた（第43表）。この結果によれば、ウソは10%をこえている。しかも実際に棄権したのに投票したというの

がその大部分である。

つぎに、このウソをついたサンプルを性、年齢、学歴、職業、外出しがちかどうか、選挙への関心、支持政党のどの分類をみても、とくにどのようなカテゴリの人間がウソをつきやすいということはいえなかつた。

第43表 ウソツキ率

	面接	事実	港区 1949				都知事	
			区長選挙		衆院選挙			
ホント	投票した	投票	52.2	83.0	56.6	84.7	75.4	89.5
	棄権した	棄権	30.8		28.1		14.1	
ウソ	棄権した	投票	4.0	13.7	1.2	14.0	1.0	10.5
	投票した	棄権	9.7		12.8		9.5	
不明			3.3	3.3	1.3	1.3	—	—
計			100.0 670	100.0 670	100.0 670	100.0 670	100.0 482	100.0 482

さらに、よくウソをつくサンプルがいるかどうかを、1649年の区長選挙、衆院選挙の両方でくらべてみた(第44表)。両方ともウソをついたのは4%にすぎない。これによれば、ウソとはいつたが、大部分の人がついたウソは無意識で(忘れてしまつて)まちがつた答をしてしまつたことによるのかもしれない。

6.4 質問はサンプルに理解されたか

面接法のような、聞きとりの調査で、質問がサンプルによく理解されたかどうかが問題である。

ところで、理解するということの定義が問題になる。たとえば知能検査の順唱・逆唱のように、オオムガエシにいえるだけで、理解したとすることもできる。一方、ソクラテスばりに‘理解した’ということは、行動をおこすことである’というような定義も考えられる。

また理解度は、これらの定義にかかわらず、測定の方法によつてもまた違つた値がでてくる。

われわれは「国民性」で検討した実験について説明しよう。対象としては中学2年生をとつた。これは「読み書き能力調査」の結果、全国民の平均点が中学2年生に一致することや、全国民の学歴の平均が8学年であることによつたのである。ただし読み書きの能力と、問題を理解することとは、必ずしも同じでないかもしれないが。

この実験は、つぎのようにしておこなつた。

1° 質問の内容 質問の中のステートメントを録音しておいたものを聞かせ、その内容はどうであつたかを、印刷されている選択肢の中からひろわせる。

2° 答とその理由 質問の録音を聞かせ、各自の意見とその理由を書かせて、質問、意見、理由の間のくいちがいをしらべる。

3° 選択文の内容 「国民性」で使つたリストの選択文の内容が理解されているかどうかをしらべるため、具体的な説明文をあげてこれに対して(抽象的な)選択文にチェックをさせる。

これらの方法のほかに、選択肢の順序も考え、しかも同じ質問もいろいろの方法でとりあげ、約50人づつ5組で別別の調査をおこなつた。

第44表

		衆院		計
区長		ホント	ウソ	
ホント		74.3	11.3	85.5
ウソ		10.4	4.1	14.5
計		84.6	15.4	100.0

100.0% = 779

この結果のくわしいことは省略するが、結論のあらましをあげておこう。

まず方法別にみれば、第2の方法（答とその理由）による測定方法では理解度が高く出る。第1、第3の方法では選択肢のもつ意義によつて、理解度は高くなつたり低くなつたりする。

つぎに、選択肢が理解度にあたえるえいきよとして、選択肢の文章の長さ、選択肢の内容のむずかしさ、選択肢の数を考えた。これらを、おのおの3つのカテゴリにまとめて、各質問の理解度（正答率）との相関係数を計算すると、つぎのようになった。

	文章の長さ	むずかしさ	選択肢の数	放送のシラブル数
相関係数	0.185	-0.031	-0.437	0.645

すなわち、選択肢の数が多いほど、理解をこんなんにしているが、あとは余り関係ない。

放送のシラブル数というのは、質問文の長さをしらべるためにとりあげてみた。この関係は0.6でやや高い。〔国民性〕では、具体的で、日常生活でよくあるようなことを、質問化したので、質問文が長くなつてしまつた。そのために、理解をこんなんにしているのではないかという疑問もあつたが、これからみると、そんなことはない。

## §7 そ の 他

その他、面接調査について注意すべきことをあげておこう。

### 7.1 調査の影響

調査をすると、サンプルは刺戟を受けたことになり、その意見、態度、行動などに変化をおこすおそれがある。このことは、つぎの事実によつて検証された。

〔都知事〕のとき、投票日前に調査をしたサンプルの（選挙人名簿でしらべた実際の）投票率は投票日後に調査をしたサンプルや、23区民全部の投票率よりも高くなつており、しかも有意な差がある（第45表）。調査することによつて、選挙の重大さを意識したためであろう。前日のサンプルでは10%も高くなつているが、このことは before-after 調査などでは、とくに考える必要がある。ただサンプル数が母集団の数より非常に小さければ、全体に対する影響は小さいわけである。

第45表 サンプルの投票率〔都知事〕

（知事選挙）

	10~12 日前	3~5 日前	前日	当日夜	4~6 日後	調査 せず	有権者 全体
投票率	67.3	69.2	73.2	66.5	64.8	63.8	62.9
サンプル数(100%)	965	886	325	179	332	1561	
有意差*	あり	あり	あり	なし	なし	なし	

（区長選挙）

	3~5 日前	2~4 日後	6日後	7日後	10~15 日後	調査 せず	有権者 全体
投票率	78.9	76.7	77.7	74.6	75.6	73.9	75.4
サンプル数(100%)	949	876	323	189	328	1533	
有意差*	あり	なし	なし	なし	なし	なし	

\* 有意差は有権者全体に対するもの(95%)

## 7-2 調査は正しいか

以上いろいろ述べてきたが、調査の結果が正しいかどうかは問題である。これは筆者がフィールド・ワークに直接関係したことはないが、新聞社の選挙予想調査と開票結果とをくらべてみよう。予想得票数は4大政党だけを、つぎのような式で計算するものである。

$$A \text{ 党の予想得票率} = A \text{ 党の固定得票率} + \text{浮動票グループ率} \times \frac{A \text{ 党の候補者数}}{4 \text{ 党の候補者数}}$$

ただし、固定得票率というのは、全サンプルのうち政党名をあげた率、浮動票グループ率というのは、政党名をハッキリいわなかつたものや無回答、調査不能のサンプルの全サンプルに対す比率の和である。(この式の解釈はここでは省略する。) この結果は第46表のようになり、きわめてよく当てているといえよう。

第46表 新聞の世論調査からの予想得票率

年, 月, 日		自 由	民 主	右 社	左 社	小 計
30, 2, 28	開票結果	28.9	40.0	14.7	16.4	100.0
30, 2, 13	毎 日	29.2	40.9	15.3	14.6	100.0
30, 1, 14	朝 日	30.7	37.1	19.2	13.0	100.0
29, 12, 20	毎 日	24.9	44.3	15.8	15.0	100.0
28, 4, 19	開票結果	51.8	19.4	14.6	14.2	100.0
28, 3, 25	毎 日	43.6	18.5	16.6	11.9	100.0
28, 1, 31	朝 日	56	13	31		100.
27, 10, 1	開票結果	54.9	20.8	13.3	11.0	100.0
27, 9, 16	毎 日	55.3	20.7	13.5	10.5	100.0
27, 9, 12	朝 日	53.1	16.9	30.0		100.0

(注) 28年の自由は分自を含む。28年、29年の民主欄は改進黨である。

面接法について、いろいろの問題を述べ、それぞれのエラーをあげてきた。これらを総合するとサンプリング以外の誤差は相当な大きさになると思われる。しかしながら、面接調査の結果は、最後にあげた選挙予想の例で明らかかなように、十分によい情報を提供しているといえよう。われわれはこの結果にももちろん満足するものではなく、調査技術にさらに改良を加えなければならないことはいうまでもないことである。

## §8 材料とした調査の説明

このなかで扱った調査の主な事項をかたんに述べておく。

## 〔 〕 調査の略称

1. 調査名
2. 発 表
3. 調査の目的と時期
4. 調査の対象, サンプリング台帳
5. 調査員

## 〔国民性〕

1. わが国民性の統計数理的研究
2. 日本統計学会会報 教育統計 30号, 1955年単行の予定.
3. 日本人の意見, 態度, ノルムなどがどうであるか, また国民の種種の要因によつてどんな違

いがあるかをみる。本調査は 1953 年 4 月に全国で実施。準備調査は同年 1 月に東京の 23 区で実施。他に関連した調査をおこなつた。

4. 全国の 20 才以上の日本人、住民登録、全国調査のサンプル数は 2250
5. 各地の大学生に、研究所員が直接説明して面接調査させた。

〔都知事〕

1. 東京都知事選挙の調査 (1951 年)。
2. 数量化と予測 (丸善)
3. 選挙予想方法の検討。調査は 5 回おこなつた。〔Ⅰ〕というのは投票日の約 10 日前、〔Ⅱ〕は約 5 日前、〔Ⅲ〕は前日、〔Ⅳ〕は当日夜、〔Ⅴ〕は数日後におこなつた。
4. 東京都全体の有権者。選挙人名簿。
5. 調査員はほとんど全部大学生。

〔効果〕

1. マス・コミュニケーションの効果。
2. 未発表。
3. トピックについての世論の変化と、新聞の内容分析とを比べて、マス・コンの効果をしらべようとするもの。
4. 第Ⅰ回調査は 1954 年 2 月に、主として市民とマス・コンとの接触をしらべ、第Ⅱ回の調査は、同年 11 月に、第Ⅰ回のサンプルの 1/4 を再調査した。第Ⅲ回以下計画中。第Ⅰ回のサンプルは約 1200、第Ⅱ回は約 300。東京都 23 区内の 20 才以上の市民。住民登録。
5. 調査員は大部分が大学生、第Ⅱ回は一部は研究職員。

〔鶴岡〕, 〔上野〕, 〔岡崎〕

1. 鶴岡は共通語化の調査、他の 2 つは敬語の社会心理学的研究。
2. 地域社会における言語生活など。
3. 題名から明らかなどあり。敬語の場合は面接、集合調査をあわせておこない、極めて多角的に研究を進めている。調査はおのおの 1950 年、1952 年、1953 年におこなつた。
4. 年令 15 才以上の市民。ただし新市域の一部は切りすてた。物資配給台帳を使用。
5. 国立国語研究所員と筆者が当つた。

### Bibliography of Research

- Bain, R.: Stability in questionnaire response. Amer. J. Sociol. 1931.
- Blankenship, A. B.: The effect of the interviewer upon the response in a public opinion poll. J. Consult. Psychol. 1940.
- Cahalam, D., Tamulonis, V. and Verner, H. W.: Interviewer bias involved in certain types of opinion survey questions. Intern. J. Opinion and Attitude Research, 1947.
- Cantril, H. and Research Ass.: *Gauging public opinion*. (Princeton Univ. Press) 1944.
- Chapin, F.: Some problems in field interviews when using the control group technique in studies in the community. Amer. Sociol. Rev. 1943.
- Crespi, L. P.: The cheater problem in polling. Pub. Opin. Quart. 1945.
- Crespi, L. P.: The interviewer effect in polling. Publ. Opin. Quart. 1948.
- Deming, W. E.: On errors in surveys. Amer. Sociol. Rev. 1944.
- Durbin J. and Stuart, A.: Differences in rates of experienced and inexperienced interviewers. J. Roy. Statist. Soc. 1951.
- Durbois, C.: The card sorting or psychophysical interview. Pub. Opin. Quart. 1949.
- Durant, H.: The cheater's problem. Pub. Opin. Quart. 1946.

- Frank, B.: Stability of questionnaire response. *J. Abnor. Social. Psychol.* 1935-6.
- Hansen M. H. and Hurwitz W. N.: The problem of nonresponse in sample survey. *J. Amer. Statist. Assoc.* 1946.
- Hyman, H.: Do they tell the truth? *Publ. Opin. Quart.* 1944.
- Katz, D.: Do interviewers bias poll results? *Publ. Opin. Quart.* 1942.
- Lazarsfeld, P. E.: The controversy over detailed interviews—an offer for negotiation. *Publ. Opin. Quart.* 1944.
- Manheimer, D. and Hyman, H.: Interviewer Performance in area sampling. *Publ. Opin. Quart.* 1949.
- Morgan, R.: Interviewer introspection on “bias”. *Publ. Opin. Quart.* 1947-8.
- Politz, A. and Simmons, W.: An attempt to get the “not at homes” into the sample without call backs. *J. Amer. Statist. Assoc.* 1950.
- Rice, S. A.: Contagious bias in the interview. *Amer. J. Sociol.* 1929.
- Rose, A. M.: A research note of experimentation in interviewing. *Amer. J. sociol.* 1945.
- Sheatslay, P. B.: Some uses of interviewer report forms. *Publ. Opin. Quart.* 1947-8.
- Sheatsley, P. B.: The influence of sub questions on interviewer performance. *Publ. Opin. Quart.* 1949.
- Stember, H. and Hyman, H.: Interviewer effects in the classification of responses. *Publ. Opin. Quart.* 1949-50.
- Stock, J. S. and Hochstin, J. R.: A method of measuring interviewer validity. *Publ. Opin. Quart.* 1951.
- Terris, T.: Are poll questions too difficult? *Publ. Opin. Quart.* 1949.