

Arbeidsnotater

T A T I S T I S K S E N T R A L B Y R Å

ronningensgt. 16, Oslo - Dep., Oslo l. Tlf. 41 38 20, 41 36 60

WORKING PAPERS FROM THE CENTRAL BUREAU OF STATISTICS OF NORWAY

IO 71/6

4. juni 1971

TO NOTATER OM FEILKILDER VED VALGUNDERSØKELSEN 1969

Av

Ib Thomsen

I N N H O L D:

	Page
A. On the effects of non-response in the Norwegian election survey 1969	2
1. Introduction and summary	2
2. Outline of the sample design	2
3. The postal stage of collection	3
4. How non-response affects estimates	5
5. Conclusions	7
6. References	7
B. Ikke-tilfeldige feil i valgundersøkelsen 1969	8
1. Innledning	8
2. Kort om utvalgsmetoden	8
3. Feiltyper	9
4. Utvalgsfeil	9
5. Svarfeil	10
6. Virkningen av introduksjonsbrevet	10
7. Forslag om nye erstatningsregler	11
8. Litteratur	12

Not for further publication. This is a working paper and its contents must not be quoted without specific permission in each case. The views expressed in this paper are not necessarily those of the Central Bureau of Statistics.

Ikke for offentliggjøring. Dette notat er et arbeidsdokument og kan siteres eller refereres bare etter spesiell tillatelse i hvert enkelt tilfelle. Synspunkter og konklusjoner kan ikke uten videre tas som uttrykk for Statistisk Sentralbyrås oppfatning.

A. ON THE EFFECT OF NON-RESPONSE IN THE NORWEGIAN ELECTION SURVEY 1969

1. Introduction and summary

In 1969 the Central Bureau of Statistics of Norway carried out an interview survey in order to study the relationship between voting behaviour and other variables, such as sex, age, income, education and occupation. The results are published in [2].

Persons not reached by the interviewer after at least three calls were contacted by mail. By this the non-response was reduced from 14.5 % to 9.9 %. By checking the final non-response against the voting register it was possible to find whether a person was a voter or a non-voter and measure the effect of the final non-response on the percentage of voters in the sample. The results of this checking is described in section 4 below.

2. Outline of the sample design

The sample design is described in detail in [3] and I shall only give an outline of the design. The sample is designed to be representative of the population of Norway aged between 20 and 79 years. The sample is selected in two stages with stratification at the first stage. At the first stage 93 primary units are selected after the following stratification: All primary units are first grouped by region; then within each region primary units in towns with a population of 100 000 or more are grouped separately; finally primary units within each of the resulting groups are stratified and 2 are selected.

At the second stage persons between 20 and 79 years are selected within the selected primary units. This selection was made from the Central Person Register. The population was defined as all persons entitled to vote i.e. persons listed in the voting registers. The sample from the Central Register was checked against the voting register. One person selected was not entitled to vote and therefore considered a blank as defined by Kish [1]. The size of the sample after removal of the blank was 2 999 persons. As in all other surveys of this kind a special problem arose because of movers, and the following procedure was used to solve this: Persons that were living at one address but were registered at another in the Central Register, such as students and sailors, were contacted by mail. Persons that had moved to another address were not followed. To all selected persons a letter was sent before the survey, asking them to co-operate in this survey. The interviewers were instructed to make at least three calls. The response rate was 85.5 %.

3. The postal stage of collection

Among the persons from whom no answer was received 85 were excluded from further investigations. To contact the remaining non-respondents three different letters were issued according to the reason for non-response.

In all 347 letters were sent out and we received an answer from 136 persons. The costs of this was kr. 7.00 per answer. The cost per interview was kr. 20.00.

Table 1 shows the percentage of answers in the postal stage of collection among refusers and others.

TABLE 1. Percentage of answers in the postal stage of collection among refusers and others

Reason for nonresponse before the postal stage of collection	Number of letters sent out	Number of answers	Response rate in the postal stage of collection
Refusers	181	40	22.1
Others	166	96	57.8

The most surprising result is the large response rate among refusers. Probably these persons are what is often called temporary refusers [1], but it seems as if the interviewer is unable to make this distinction during the visit.

We shall define three different samples:

A_1 : The sample collected by interviewers, sample size 2567

A_2 : The sample collected by mail, sample size 136

A_3 : The non-response, sample size 296

In the following $A_i + A_j$ means the sample consisting of observations from sample A_i or A_j . In table 2 is shown the effect of sample A_2 on the rate of voting.

TABLE 2: Rate of voting in sample A_1 and $A_1 + A_2$

	Age					
	20-24 years	25-29 years	30-49 years	50-69 years	70-75 years	All ages
Rate of voting in sample A_1	88.4	88.3	92.1	92.0	85.5	90.8
" " " " " $A_1 + A_2$.	85	88	92	92	85	90

It is seen that the effect is large for very young people. The reason for this is that the sample A_2 consists of a large number of young people, who live at one address and are registered at another. A person in this category has to go back to his home to vote or must go to the local police which is probably the reason for the low percentage of voters.

Table 3 and 4 shows the effect of sample A_2 on the voting behaviour e.i. the rate of voting for different parties.

TABLE 3. Voters in different age groups, by party. Percentages. Sample $A_1 + A_2$

Age	Party								Total	Number of respondents
	A	H	KrF	K	Sp	SF	V	Un-known		
BOTH SEXES										
20-24 years	49	17	7	-	13	4	9	1	100	220
25-29 "	47	24	5	1	17	2	4	-	100	180
30-49 "	49	17	9	-	12	2	9	1	100	994
50-69 "	50	17	10	-	12	1	9	1	100	836
70-79 "	44	17	13	-	15	1	9	1	100	209
All ages	48	18	9	1	13	2	8	1	100	2 439
MALES										
20-24 years	42	23	3	-	18	5	7	2	100	125
25-29 "	49	22	2	1	19	3	4	-	100	95
30-49 "	52	17	6	1	12	3	7	2	100	516
50-69 "	51	16	7	1	14	1	9	1	100	428
70-79 "	51	16	6	-	19	2	6	-	100	101
All ages	50	18	6	1	14	2	8	1	100	1 265
FEMALES										
20-24 years	57	8	12	-	7	2	13	1	100	95
25-29 "	45	27	8	-	14	1	5	-	100	85
30-49 "	45	18	11	1	12	2	10	1	100	478
50-69 "	48	18	15	-	9	1	8	1	100	408
70-79 "	36	19	19	-	11	1	12	2	100	108
All ages	46	18	13	-	11	1	10	1	100	1 174

TABLE 4. Voters in different age groups, by party. Percentages. Sample A₁

Age	Party								Total	Number of respond- ents
	A	H	KrF	K	Sp	SF	V	Un- known		
BOTH SEXES										
20-24 years	53	14	7	-	14	3	8	1	100	190
25-29 "	47	24	5	1	17	2	4	-	100	173
30-49 "	49	18	9	1	12	2	9	1	100	955
50-69 "	50	17	10	1	12	1	8	1	100	808
70-79 "	44	18	12	-	15	2	9	1	100	206
All ages	49	18	9	1	13	2	8	1	100	2 332
MALES										
20-24 years	47	19	4	-	20	4	5	1	100	105
25-29 "	49	20	2	1	20	3	3	-	100	89
30-49 "	52	17	6	1	12	3	7	2	100	490
50-69 "	52	16	7	1	15	1	9	1	100	413
70-79 "	52	16	6	-	19	2	6		100	101
All ages	51	17	6	1	15	2	7	1	100	1 198
FEMALES										
20-24 years	59	7	12	-	7	2	12	1	100	85
25-29 "	45	27	7	-	14	1	5	-	100	84
30-49 "	45	18	11	1	12	2	10	1	100	405
50-69 "	48	18	14	-	9	1	8	1	100	359
70-79 "	36	19	18	-	11	1	2	2	100	105
All ages	47	18	13	-	11	1	9	1	100	1 134

4. How non-response affects estimates

In the great majority of surveys one is not able to say much about the effects of the non-response on the estimates. In most cases one is limited to tabulate the age-sex distribution of the non-response. In this survey we checked all selected persons by means of the voting registers. By this control it turned out that 30 persons answered that they were voters in the election 1969 but were not marked as voters in the voting registers. 2 persons answered "not voter" in the survey but were voters according to the voting registers. The answers from these persons were corrected according to the information in the voting registers. By checking the non-respondents in the same way it was

possible to study the effect of the non-respondents on the rate of voting.

Table 5 gives the rate of voting in the final non-response, A_3 and in the whole sample $A_1 + A_2 + A_3$.

TABLE 5. Rate of voting in sample A_3 and $A_1 + A_2 + A_3$

	Age					All ages
	20-24	25-29	30-49	50-69	70-79	
Rate of voting in the final non-response A_3	59.0	55.6	72.4	78.3	73.5	71
Rate of voting in the whole sample, $A_1 + A_2 + A_3$	81	84	90	91	84	88

From table 5 it is seen that the percentage of voters is smaller among non-respondents than in the whole sample and that the difference is bigger among young people than among older people.

The effect of the non-response on the estimates of the rate of voting in the population is shown in table 6.

TABLE 6. Rate of voting in sample $A_1 + A_2 + A_3$ and $A_1 + A_2$

	Age					All ages
	20-24	25-29	30-49	50-69	70-79	
Rate of voting in the whole sample $A_1 + A_2 + A_3$	81	84	90	91	84	88
Rate of voting among the respondents $A_1 + A_2$	85	88	92	92	85	90

As expected from the results in table 5 the effect is largest for young people.

We have seen that the rate of voting is lower in the final non-response than in the whole sample, and a question of interest is: Why is it low? To study this a little further the rate of voting is given for three different groups of non-responders according to the reason for non-response.

TABLE 7. Rates of voting in A_3

Rate of voting among refusers	81.0
" " " among not at homes	65.1
" " " among mentally or physically ill persons	55.1

In my opinion the most interesting finding is the high rate among the refusers. One would expect that the reason for refusing to answer is a lack of interest in elections. We have calculated the rate of voting among the refusers for different agegroups and they seem to indicate that among young refusers the rate is very low (40 % - 50 %), while among older refusers the rate is close to the rate in the whole sample. But the number of observations is too small to give more than an indication.

5. Conclusions

When the sample is selected from a central register of persons it is of importance to follow up persons that are not living at the registered address. How to follow up these is still an open question, which will be studied further in the future.

Whatever we do to reduce the non-response there will remain a "hard core" of refusals and others from which we will not receive an answer. The effect of this will be that the percentage of voters in our sample will be overestimated for reasons mentioned in section 4.

6. References

- [1] Kish, Leslie (1965): Survey sampling. Wiley, New York
- [2] NOS. A 340 (1970): Storting Elections 1969. Vol. II. Central Bureau of Statistics of Norway
- [3] Tamsfoss, Steinar (1969): On the use of sample surveys by The Central Bureau of Statistics, Norway. Working Paper.

B. IKKE TILFELDIGE FEIL I VALGUNDERSØKELSEN 1969

1. Innledning

I 1969 utførte Statistisk Sentralbyrå en valgundersøkelse. Resultatene av undersøkelsen er publisert i [2]. Fra et datainnsamlingssynspunkt er undersøkelsen av særlig interesse bl.a. fordi vi har mulighet for å sammenligne svarere på spørsmålet om personen har stemt eller ikke med opplysningene i manntallslistene, hvor alle som har stemt er avkrysset. De spørsmål vi har tatt sikte på å undersøke i dette notatet er av typen: Hvorledes påvirkes utvalget av erstatningsregler? Hva kan gjøres for å redusere de uheldige virkninger? Får vi korrekte opplysninger på våre spørsmål? I slutten av notatet er gitt forslag om nye erstatningsregler. Mange av resultatene fra denne analysen har interesse for andre undersøkelser enn valgundersøkelser både når det gjelder effekten av svarfeil, introduksjonsbrev (IO-brev) og erstatningsregler.

2. Kort om utvalgsmetoden

Utvalgsplanen er reservert i [4] og jeg skal her bare nevne noen enkelte viktige trekk av interesse i denne sammenhengen.

Utvalget er planlagt representativt for velgere i alderen 20-79 år. En velger er definert som en person oppført i mantallet ved valget 1969.

Særlig tre ting er viktige:

- (i) Trekkeenheten er person
- (ii) Vi trekker fra et register som ikke er helt ajour
- (iii) Registeret inneholder personer som ikke er velgere

Målet for trekkingen på 2. trinn er at enhver velger i det uttrukne primærområder skal ha samme kjente sannsynlighet for å bli trukket. På grunn av (i) og (ii) får en derfor problemer med personer som er trukket på 2. trinn, men som er flyttet ut av primærområdet før valgdagen. I denne undersøkelsen valgte en å erstatte en slik person med en annen tilfeldig trukket person. På denne måten ble altså alle som hadde flyttet fra en kommune til en annen like før valget holdt utenfor undersøkelsen. I avsnitt 4 er virkningen av disse erstatninger analysert og i avsnitt 7 er nye erstatningsregler foreslått.

Når det gjelder virkningen av at vi trekker personer som ikke tilhører populasjonen, er det ikke gjort noe forsøk på å beregne virkningen av dette. Noen generelle merknader om problemet, er imidlertid gitt av Leslie Kish i Survey Sampling [1].

3. Feiltyper

I dette notatet skal bare ikke-tilfeldige feil behandles. Problemer i forbindelse med utregning av varianser på estimatorene vil derfor ikke bli tatt opp her. Heller ikke skal vi her ta opp feil som oppstår under bearbeidningen av materialet. Vi står igjen med følgende feiltyper:

1. Utvalgsfeil
2. Svarfeil
3. Virkningen av introduksjonsbrevet

4. Utvalgsfeil

Med utvalgsfeil menes her de skjevheter vi innfører under trekkingen av utvalget. Vi skal spesielt se på virkningen av at alle flyttere erstattes. For å studere virkningen har vi tatt for oss alle flyttere som ble trukket og undersøkt i manntallslistene om disse stemte ved valget. Vi har i hele notatet sett bort fra muligheten av feil i manntallslistene. Resultatet av denne undersøkelsen er gitt i tabell 1.

TABELL 1. Valgdeltaking blant uttrukne flyttere etter alder. Prosent

	Alder					Alle aldre
	20-24 år	25-29 år	30-49 år	50-69 år	70-79 år	
Stemte	64	68	73	70	79	68
Stemte ikke	36	32	27	30	21	32
Tallet på personer ...	43	25	41	20	7	136

Valgdeltakelsen i populasjonen utregnet på grunnlag av opplysninger fra [3] er gitt i tabell 2.

TABELL 2. Valgdeltaking i populasjonen. Prosent

	Alder					Alle aldre
	20-24 år	25-29 år	30-49 år	50-69 år	70-79 år	
Stemte	73	81	88	88	82	85
Stemte ikke	27	19	12	12	18	15

Valgdeltakelsen blant flyttere er signifikant mindre enn valgprosenten i hele populasjonen. Dette kan skyldes aldersfordelingen blant flyttere, men tabell 1 og 2 indikerer en tendens uavhengig av alder. Antall observasjoner er imidlertid for lite til å gi mer enn en indikasjon. Uansett årsaken innføres en skjevhet i utvalget ved å erstatte flyttere.

5. Svarfeil

Ved kontroll mot manntallslistene viste det seg at 30 personer som svarte at de stemte, ikke var avkrysset i manntallslistene. Disse ble rettet i samsvar med opplysningene i manntallslistene. To personer svarte at de ikke stemte ved valget, men var avkrysset i manntallslistene.

6. Virkningen av introduksjonsbrevet

For å undersøke virkningen av introduksjonsbrevet er en nødt til å utføre et eksperiment, der en del av utvalget (tilfeldig trukket) blir tilsendt et brev, mens en annen del ikke får noe. Deretter må innkomne svar bearbeides med tanke på å finne eventuell forskjell.

Dette ble ikke gjort ved valgundersøkelsen, men jeg vil likevel forsøke å si noe om virkningen ut fra de foreliggende data ved å se på det utvalg vi opprinnelig trakk fra registeret, og undersøke valgdeltakelsen for dette utvalget, og sammenligne denne med valgdeltakelsen for populasjonen. Svakheter ved denne framgangsmåten er store især fordi dette utvalget består av en blanding av personer som har fått introduksjonsbrev før valget (85,5 %) og personer som ikke har fått brev før valget. Og delingen er ikke skjedd tilfeldig.

Dessuten viste det seg vanskelig og arbeidskrevende å finne valgdeltakelsen for det opprinnelige utvalget. Dette er funnet ved å fjerne alle erstatninger fra det innsamlede materiale og deretter tilføye alle flyttere til utvalget. Størrelsen på dette utvalget er 3018 som er 19 flere enn opprinnelig trukket. Dette skyldes at 19 personer opptrer to ganger i utvalget, nemlig som frafall og som flytter. En utskillelse av disse 19 personer ville være meget arbeidskrevende og er derfor ikke gjort.

TABELL 3. Valgdeltaking i "Det opprinnelige utvalg" etter alder. Prosent*

Alder	Stemte	Stemte ikke	Total antall
20 - 24 år	78	22	331
25 - 29 "	82	18	246
30 - 49 "	89	11	1 181
50 - 69 "	90	10	985
70 - 79 "	83	17	275
Alle aldre	87	13	3 018

Ved å sammenligne dette med tabell 2 finner en at også i dette utvalget ligger stemmefrekvensen systematisk over stemmefrekvensen i populasjonen. Selv med de ovennevnte betenkeligheter er det nærliggende å tro at den høye valgdeltakelsen kan ha sammenheng med at en stor del av utvalget har fått tilsendt introduksjonsbrev før valget, noe som kan ha endret valgadferden til de uttrukne personer.

7. Forslag til nye erstatningsregler

Av det foregående framgår at vi ved den anvendte utvalgsplan systematisk har utelatt alle flyttere av utvalget. Jeg skal nå vise hvorledes gruppen flyttere kan komme inn i vårt utvalg. Det første en bør være oppmerksom på er at også de som er flyttet inn er flyttere, og det vi ønsker er at disse personer skal ha samme sannsynlighet for å komme med i utvalget som de øvrige personer innen utvalgsområdet. Dette kan gjøres på følgende måte: Når vi trekker personer med samme sannsynlighet p , har en husholdning bestående av n_i personer sannsynligheten $n_i \cdot p$ for å bli trukket. (Dette er en tilstrekkelig god tilnærming for vårt bruk). Hvis det viser seg at en ny husholdning er flyttet inn, og at denne har m_i medlemmer, kan intervjueren trekke en person tilfeldig. Sannsynligheten for at en bestemt person i den nye husholdningen kommer med i utvalget blir altså $\frac{n_i \cdot p}{m_i}$, som bare er lik p dersom $m_i = n_i$. Ved tabellkjøringer må da opplysningene for disse personene veies med $\frac{m_i}{n_i}$.

Et annet tiltak som bør settes i gang ved neste valgundersøkelse, er at vi undersøker mulighetene for å bruke intervjuere ved oppfølgingen av midlertidig fraværende personer; det er fortsatt nesten 50 % av frafallet vi bør forsøke å få inn.

* De "opprinnelige utvalg" består av 3018 observasjoner.

Sist, men ikke minst, bør vi vel bruke introduksjonsbrevet på en annen måte. Enten ved å sende det etter valget, eller unngå at vi opplyser om hva vi ønsker å spørre om.

8. Litteratur

- [1] Kish, Leslie (1965): Survey Sampling. John Wiley and Sons. New York
- [2] NOS. A 340 (1970): Stortingsvalget 1969. Hefte II. Statistisk Sentralbyrå
- [3] NOS. A 354 (1970): Stortingsvalget 1969. Hefte III Statistisk Sentralbyrå
- [4] Tamsfoss, Steinar (1969): Om bruken av stikkprøver ved kontoret for intervjuundersøkelser. Artikler 37. Statistisk Sentralbyrå