TECHNICAL REPORT

"Analysis of the Data Collection and
Statistical Methods
used for Departmental Survey
and Sampling Activities"
Revised and Enlarged Version

Department of Communications, Canada
For the DOSP-C Section
By: François Théberge
Ottawa, April 1991

TECHNICAL REPORT

"Analysis of the Data Collection and
Statistical Methods
used for Departmental Survey
and Sampling Activities"
Revised and Enlarged Version

Department of Communications, Canada
For the DOSP-C Section
By: François Théberge
Ottawa, April 1991

Acknowledgements:

The production of this report was made possible thanks to the collaboration of many employees of the Department of Communications. I would like to thank Gontran Bolduc for his active participation in this report, and for his encouragement throughout my two internships at the Department.

Furthermore, many comments and information were provided by Gary Steckley (DOSP-C), François Guillot, Octave Bélanger, and Colette Tremblay (DAA). Also I would like to thank the staff at the Pacific, Quebec, and Ontario regions who answered my questions.

Table of Contents

Formulas

1. Introduction

The main purpose of this report is to perform a detailed analysis of the following points:

(a) Land-fixed sampling (surveys)[1]. To review data collection and statistical processing as a whole.

(b) Spectrum surveillance. To evaluate the usefulness of this activity.

(c) General public and radiocommunication "radiocom" investigations. To revise the current coding and data presentation methods, to maximize the usefulness of the information provided.

(d) To compare already studied survey and sampling activities, and determine whether there is any correlation between these activities.

In this second version, many paragraphs was modified or added, especially in Sections 4 and 5. Sections 2 and 3, which deal with surveys, was reorganized and completed, and examples was added.

Finally, several appendices was added to provide the maximum amount of information on the technical aspects of the survey and sampling activities, without having to burden the text excessively.

---

[1]The current term used for these activities is "sampling," which refers to the choice and collection of samples. The term "survey" refers to the activity as a whole, including statistical analysis of the results. We will try to follow these definitions, but, as this document only discusses one type of activity, there should not be any confusion.

2. Sampling - Analysis of the Situation

The sampling activities (or surveys) of the Department of Communications fall into two distinct categories: **land-fixed sampling** and **spectrum surveillance.** The common purpose of these activities is to evaluate the quality of radio environment and to attempt to identify problems at various levels. As indicated in the past DOSP-C (spectrum control) annual reports, these activities have not yet produced all the anticipated results. Thus, we will review these activities, both to ensure that they are statistically correct, and to attempt to draw the maximum amount of information at the minimum cost. In this section, we will try to describe the various problems, while section 3 will list the recommendations.

The analysis followed the activity from "start to finish," in other words, from data collection to the processing of the results. Historical data and documents describing land-fixed sampling (RIM, IPC) were the main sources for data analysis. Comments and recommendations from various parties involved were also considered. Finally, a complete revision of the process of report production by the SCORE (Spectrum Control Output Report Editor) program was carried out.

2.1 Land-Fixed Survey

Land-fixed (or "on site") sampling is carried out by inspectors who go to the premises of randomly chosen stations, and whenever possible, the 15 discrepancies listed in Appendix 3 are checked. This activity can be broken down into the following steps:

1 - Definition of the population (stations) to be surveyed. Geography and exclusions are taken into account.

2 - Determination of the size of the sample to be analyzed using Arkin's formula. This formula is used to obtain a sample size that will guarantee the desired margin of error and confidence level. This will be discussed in the next section.

3 - Random choice of stations to be analyzed.

4 - Collection of data by on-site inspection and reception.

5 - Generation of reports using the SCORE and SMIS systems.

6 - Evaluation of the situation by senior management and planning of corrective measures.

2.1.1 Arkin's Model

The purpose of a survey is to draw conclusions about a population as a whole based on data about only part of that population. Thus, the results obtained cannot be 100% accurate, this is why a sampling plan has a confidence level and a margin of error. For example, a survey with a confidence level of 95% and a margin of error of 4% indicates that, in 95% of the cases (19 times out of 20), the results obtained with the survey will differ from the real value by only 4% or less. This margin of error is obviously valid only if the data are collected correctly.

In the case under consideration, we wish to measure proportions, which is to say, the various discrepancy percentages. After having chosen the confidence level and the margin of error of our survey, we must determine the size of the sample to be obtained (at random) from the population. The following formula, from Hubert Arkin's book [Arkin, p. 96], provides a simple method of making this calculation:

$$n = \frac{p(1-p)}{\left(\frac{SE}{t}\right)^2 + \frac{p(1-p)}{N}}$$

p : anticipated discrepancy rate, with $p \in (0,1)$.
SE: margin of error
N : population size
n : sample size
t : normal distribution parameter corresponding to a given confidence level, for example:

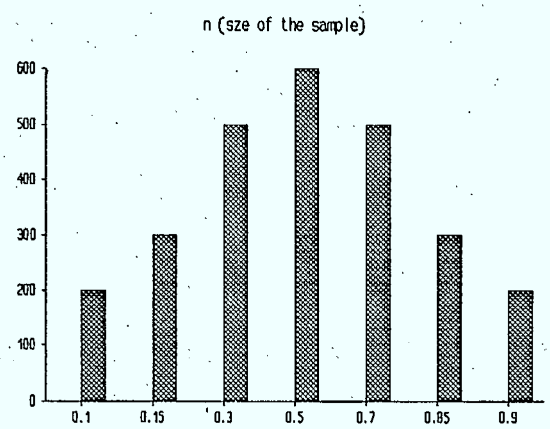
Conf.	t
90%	1.645
95%	1.960
99%	2.575

Arkin's Formula

Various sampling plans using this formula will be analyzed in the next paragraph. In terms of the formula itself, we should keep the following points in mind:

1) In order to make comparison of the results easier, the margin of error (SE = .04 = 4%) and the confidence level (95%) should remain the same from year to year.

2) In order to be able to apply Arkin's formula, which allows us to determine the size of the sample, we must be able to approximate discrepancy rate that will be measured (expressed as "p".) Though there are several different types of discrepancies, and each can be detected in various proportions, it is easier to use a single value for "p", based on results from the previous year. This value has a major impact on the sample size, as the following example, which uses SE=4%, a 95% confidence level, and an infinite population, shows:



Value of "p"

2.1.2 Involvement Level

To ensure that each region participates in the land-fixed sampling activity, several sampling plans are currently[1] offered, based on available resources. These three plans (A, B, and C) represent the use of Arkin's formula at the regional, district, and universe levels respectively. (N.B., the term "universe" is used here to designate any metropolitan or non-metropolitan area, or sub-universe defined by the district under consideration.) Furthermore, these plans are embedded within one another, so that stations in Plan C are also part of Plan B and A, and stations in Plan B are part of Plan A.

Plan A is currently the minimum level of involvement. The determination of the sample size for these 3 plans is currently[1] done as follows:

1 - Determination of the population to sample, specifying exclusions* at the national level, and definition of the universes in each district.

2 - For each universe, use of Arkin's formula to calculate the size of each Plan C sample, and random choice of stations. The value $p=0.1$ (anticipated discrepancy rate) was used for the year 1990-91 at all levels. Moreover, the confidence level was set at 95%, and the margin of error at $\pm 4\%$.

3 - For each district, determination at random of the stations that will form part of plan B, based on those already chosen for plan C.

4 - For each region, determination at random of the stations that will form part of plan A, based on those already chosen for plan B.

5 - Sending the lists of stations in the various sampling plans to the regions and districts.

* "Exclusions" are all those radio stations that are excluded from the sample, for whatever reason (eg. National Defence, RCMP, etc.).

---

[1]This is in reference to fiscal year 1989-90.

A summary of the 1990/91 sampling data is included in **Table 1** (*n.b., the data tables may be found in Appendix 2.)* The data in the table are shown by region. The first column shows the universe under consideration. A universe is represented by a six figure code, in which the first number represents the region (1,2,4,5 or 6), the next two the district number, and the last three are in accordance with the following convention:

001 - district metropolitan area
002 - district non-metropolitan area
other - sub-universe defined by the district

Tables 2, 3, and 4 contain the results obtained for the country, a region, and a district respectively. All this information will be used for the analysis of data, and the production of statistical reports.

(a) Distribution of Stations

Sampling for Plan C is carried out using Arkin's formula directly, but the method used for plans B and A could be improved. Indeed, for Plan B, we can see (**Table 1**) that all the districts with two or more universes were assigned a sample of 220 stations, INDEPENDENTLY of the population of the district, while Arkin's formula shows that the size of the sample depends upon the size of the population under consideration. Furthermore, we can see that the **distribution** of these 220 stations is based on the size of the Plan C samples for each universe, rather than on the population of each universe. District 22 is the most striking example. The situation is similar in the case of sampling Plan A, at the level of each district and region:

Example, District 22 (90-91):

Data[2] (see Table 1).

	Universe 1 (metropolitan)	Universe 2 (non-metropolitan)
Population	1286	7451
Plan C	185	210
Plan B	101	119

Analysis, Plan B:

1) There are 220 stations in the district, while Arkin's formula requires 211.

2) Proportion of Plan B stations per universe = 119:101 ≈ 1.18.
Proportion of populations per universe = 7451:1286 ≈ 5.79.
Proportion of Plan C stations per universe = 210:185 ≈ 1.14.

[2] The population is an approximation, as indicated on page A-4.

The proportion of Plan B sample sizes should approximately follow that of the populations of the two universes in the district. Thus, it would be normal to have about 5.79 times more stations in universe two than in universe one. Instead, the ratio between the universes in terms of the number of Plan B stations is similar to the ratio for Plan C (1.14 vs. 1.18). Thus, Plan B is not statistically correct, since the two universes have almost the same weight in Plan B, which does not correspond to reality. Moreover, an analysis of **Table 1** shows that the two problems raised in the example are generalized:

1) Misapplication of Arkin's formula in plans A and B.

2) Poor distribution of Plans A and B stations among the districts and universes.

(b) National Level (Plan ZERO)

The descriptions of plans A, B, and C show that these various levels depend upon the resources available in each region, and that Plan A participation is considered to be the minimum required to obtain a detailed national report. In fact, if all the regions participate in plan A, we will have a sample that is 5 times larger than that required for the desired confidence level for the national report. This is due to the fact that there are 5 regions, and Plan A assumes that the confidence level will be met at the regional level (for the regional reports).

Example:

The following are the 1990/91 data, with p=0.1 as a "forecast", a margin of error of ±4%, and a confidence threshold of 95%:

Data (see Table 1).

Region	Population	Plan A Size
Quebec	10718	214
Ontario	16586	214
Atlantic	7163	210
Pacific	10579	213
Central	32257	215
TOTAL (National)	77,303	1,066

n.b., The population is commercial private, after exclusions (Armed Forces, RCMP, municipalities).

According to Arkin's formula, a sample of 216 stations at the national level would be enough to obtain a confidence level of 95% and an error of ±4%. This is what we will define as sampling Plan 0 (zero).

2.1.3 Definition of Universes

Each district may be divided into universes. This is currently done by separating the metropolitan and non-metropolitan areas in each district. Furthermore, circular sub-universes may be defined by districts, but, in all cases, **national** exclusions should be used.

The current distribution of universes lacks flexibility, especially concerning exclusions, which should be the same as those used nationally. A new definition of Plan C will be created, in order to solve this problem.

2.1.4 Quality Control

Quality control is currently attained by using district reports generated by the SCORE software. This report shows the number of inspections carried out, and the number provided for each discrepancy.

An analysis of the data obtained from surveys in previous years shows that, depending upon the districts and the inspectors, the verifications carried out are not always the same, which leads to the suspicion that the sampling process is not understood in the same way by everyone. Such a phenomenon biases the results, especially if some discrepancies are verified in some districts and virtually ignored in others.

Another problem is ensuring that the invalid data are properly identified, as opposed to verifications that were carried out without any problems. There are two different codes which make this distinction, but an analysis of the data from some districts shows that the "analysis not carried out" code is often completely ignored. This obviously reduces the accuracy of our data, as it is almost impossible to check everything during each visit.

These variations in point of view and practice between the districts show that the data collection process should be revised, and that an additional effort should be made at the training level.

2.1.5 Statistical Reports

The results we wish to verify are the 15 discrepancy percentages and the associated errors, at the universe, district, regional, and national levels. The two formulas on the following page can be used to calculate the percentage of occurrence of a discrepancy [Arkin, p. 606].

Formula [1] is used in cases involving a single territory from which samples are obtained at random. However, in the current case, higher-level reports (regional, national) use data collected in different districts. The "weighted" formula [2], taken from [Schaeffer, sec. 5.6], was used in the technical document dealing with sampling [IPC 3.21.01]. This document proposes using the universe as the basic entity (or "stratum"), which is to say that we propose that the discrepancy rates should be calculated for all the reports by weighting on the basis of the rates in each universe.

Methods of calculating confidence intervals:

(a) Simple Sampling, Formula

$$p = \left(\frac{a}{n}\right) \qquad \epsilon = \pm t\sqrt{\frac{(N-n)p(1-p)}{N(n-1)}}$$

a : number of discrepancies observed
n : total number of verifications carried out
p : estimate of the discrepancy rate
∈ : estimation error of the discrepancy rate
N : size of the population
t : confidence level (see Arkin's formula)

Formula [1]

(b) Stratified Sampling

$$p = \frac{\sum_i (N_i \times p_i)}{N} \qquad \epsilon = \pm\frac{t}{N}\sqrt{\sum_i\left(\frac{N_i(N_i-n_i)p_i(1-p_i)}{(n_i-1)}\right)}$$

p : estimate of the discrepancy rate
N_1 : population of the i^{th} stratum (or other division)
p_1 : estimate of the discrepancy rate for the i^{th} stratum using formula [1]
N : total population of the territory
∈ : estimation error of the discrepancy rate
n_1 : size of the sample in the i^{th} stratum
t : confidence level (see Arkin's formula for meaning)

Formula [2]

There are four types of reports, depending upon the geographic entity under consideration. These are the universe, district, regional, and national reports. Each is analyzed separately here, on the basis of results obtained with the SCORE software, and the technical document [IPC 3.21.01].

(a) National Report

Table 2 (in Appendix 2) is a list of the data obtained using the SCORE software for each of the regions, **for discrepancy #1 alone.** Please note that the Atlantic region was not taken into account, as no data was included in the SCORE data base for this region for the 1989-90 sampling.

Analysis

For the national report, if formula [1] is used, the result obtained is p = 67/1786 ≈ 4%.

Using formula [2], based on the regions, produces a figure of 7%. SCORE produces the same figure.

Thus, this software seems to use formula [2], weighting as a function of the regions. Furthermore, it is easy (but time consuming) to verify that the error associated with this rate is well under ± 1%, as the SCORE report and formula [2] indicate. The other 14 anomalies were verified, and, **in all cases**, formula [2] was used, based on the regions. Now it remains to be seen how the discrepancy rates are calculated at the level of each of the regions.

(b) Regional Reports

To check the validity of these reports, the 1989/90 data for Quebec were used, and discrepancy #9 was checked. The data are summarized in **Table 3**.

Analysis

If formula [1] is used, a result of 222/508 = 44% is obtained. Formula [2] weighted as a function of the **universes** produces a figure of 35%, and if this formula is weighted as a function of the **districts** (the 3rd and 4th figures of the universe number) a result of 48.6% is obtained.

The regional report produced by SCORE shows a value of **44%** for discrepancy #9, and the only way to arrive at this result is to use formula [1]. The reports for other regions were also checked, and in all cases, the discrepancy calculation was made by simple division. Thus, the reports produced by SCORE do not meet the requirements of the technical document [IPC 3.21.01, 1990 draft], which recommends using formula [2], taking the universes as the basic element. This is a problem in cases where certain districts of the region have had many more surveys carried out than others. This problem exists at the regional level, and also at the level of the national report, since the latter uses the regional discrepancy rates to obtain its results.

(c) District Reports

The data for the Montreal district will be analyzed in detail (**Table 4**). Once again, we will try to determine whether the calculation of discrepancies was done using formula [2] or directly.

Analysis:

Formula [2] produced a result of 35%, and the direct calculation a result of 65/256 ≈ 25%. The district report obtained with SCORE (unit 5672) shows a value of 25%. Thus, formula [2] was not used, and, again, this conclusion is generalized.

(d) Universe Reports

This type of report was not really analyzed, as the universe is the smallest geographical element considered. Thus, the only way to calculate discrepancy rates in universes is to use formula [1], or simple division.

2.2 Spectrum Surveillance

Spectrum surveillance, or monitoring, is another way of "measuring" the compliance of spectrum users in Canada. This activity is carried out from control centres across Canada and from specially equipped vehicles. It involves listening to a certain number of frequencies randomly selected from all **allocated** frequencies in Canada. The number of frequencies to be monitored for a given territory is determined using Arkin's formula, as described above. Any discrepancies detected are noted, and the inspector should then attempt to remedy them. Obviously, it is impossible to verify all 15 anomalies listed in the Appendix, as some require the inspector to be present at the station, which is not the case in spectrum surveillance. For the purpose of data collection, universes are created in order to divide the territory of a district. These are circles with varying diameters. The size of the population (that is, the allocated frequencies) of each universe is then determined, taking into account "exceptions," and Arkin's formula is used to determine the number of frequencies to be randomly selected and monitored. The following facts cast doubt on the continued existence of this activity in its current form:

- Since not all districts are equipped in the same way, and not all areas of the country are easily accessible, it is impossible to obtain a valid national, or even regional portrait.

- Several discrepancies are difficult, if not impossible, to verify using this method, as they require the presence of an inspector on the site.

- Since inspectors are required to take CORRECTIVE actions during this activity, they could easily tend to <u>first</u> check those frequencies where they expect to find problems, which could bias the results.

- Also due to the corrective nature of the activity, the actual time spent monitoring a station is quite variable, as inspectors are required to communicate immediatly with the licensee when a discrepancy is detected, and also because one can never be sure of the times when a station will be on the air.

- An analysis of data from previous years shows that the various regions do not check the same types of discrepancies using this method. There are even reports that show detection of discrepancies that can only be verified on site (#13 - station installation, and #14 - certification of operational personnel), which does not inspire confidence in the accuracy of the data!

- These same data show that, other than operational irregularities, few verifications are actually made using this method. Furthermore, these operational discrepancies could also be verified by monitoring the station in question before arriving at the premises to carry out the on-site sampling activity.

- Finally, experience from past years has demonstrated that this activity is of little use, since the data have never been used to produce reports.

For all of these reasons, it will be recommended, in the next section, that this activity be abandoned, at least at the national level.

3. Sampling - Recommendations

Following the discussion in the last section, we will provide a series of recommendations here. The paragraphs are numbered in the same way as those in Section 2, except for a few points that have been added.

3.1 Land-Fixed Sampling

A new land fixed sampling plan, plan 0 (zero), will be introduced and used as a minimum national standard. Furthermore, the process of choosing stations will be done in order, from Plan 0 (most general) to Plan C (most precise.) Other recommendations are made in order to add to the accuracy and flexibility of the activity.

3.1.1 Arkin's Model

Arkin's formula, as described in Section 2, is a good method for the estimation of the sample size corresponding to the desired margin of error. In terms of the choice of p[3], it would be best to analyze the preceding year's report, and set the value of "p" as a function of that discrepancy that, among the 15 discrepancies, is closest to 50%. Given that, in practice, it is possible to have some discrepancies that are much higher than others, it is normal to estimate "p" on the basis of the majority of the discrepancies, as all exaggerated increases in this estimation lead to a major expansion in the necessary resources. On the other hand, it must be remembered that an underestimation of "p" will automatically lead to a drop in the accuracy of the results. For example, in 1989-90, six discrepancies were above 10%. Thus, we recommend increasing this value to 15%, and seeing whether the new data suggests another change in this value.

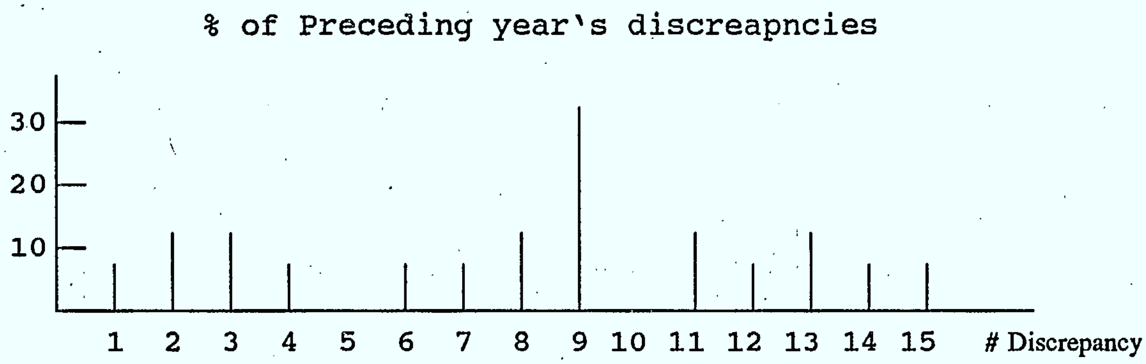
Furthermore, if participation in Plan A becomes sufficient, it would be desirable to set this "p" parameter by region, especially if major differences between the regions have been discovered. The same principle applies to the districts, if Plan B is completed.

At the universe level, Plan C should now be independent of all the other sampling plans. Given the local character of this activity, it would be desirable for this parameter to be estimated by the district office, given their experience working in their area.

---

[3] Expected discrepancy rate

Example:



If p is set at 30%, good accuracy is assured for all the discrepancies, but major resources are required. If (for example), accuracy for discrepancy #9 is not too important, p could be set at 10%, which ensures good accuracy for all the other discrepancies, and cuts back on the resources required.

3.1.2 Involvement Level

At the central administrative level, sampling Plan 0 (zero) is enough to produce an annual national report. Thus, it is possible to require only Plan 0 (zero) as the minimum data collection level, while explaining to the regions that if they do not participate in Plan A, they will not obtain a detailed report of their territory.

The following are recommendations for a better application of sampling Plans A, B, and C, and the implementation of Plan 0 (zero) at the national level. The main difference from the earlier method is that the stations for the various plans will be chosen going from the most general (plan 0) to the most specific (plan A), "adding" as many stations as required.

Furthermore, we wish to allow the districts to define their own universes, and their own exclusions within these universes. The national exclusions are used for the choice of stations, unless otherwise indicated.

Step #1 : Plan 0 (NATIONAL)

All stations :

↓

##########

Sample, Plan 0 (zero).

The size of the sample to be obtained is determined by using Arkin's method, adding 15% as a safety margin.

Step #2 : Plan A (REGIONAL)

For Plan A at the regional level, we obviously wish to use the stations selected for Plan 0 (zero) from this region. Thus, the necessary sample must be completed, in order to obtain the size required by Arkin's formula.

```
*        *
   *  *       *
 *  *     *
```

All stations, LESS those already chosen for Plan 0 (zero).

↓

```
*   *
```

Stations necessary to complete Plan A according to Arkin's model.

Thus, the sampling for Plan A is based on:

(1) Stations chosen in Step #1 from this region.
(2) Stations chosen in this step.

Step #3 : Plan B (DISTRICT)

For Plan B at the district level, we obviously wish to use the stations selected for Plans 0 (zero) and A from this region. Thus, the necessary sample must be completed, in order to obtain the size required by Arkin's formula.

```
*        *
   *  *      *
 *  *      *
```

All stations, LESS those already chosen for Plans 0 (zero) and A.

↓

```
*   *
```

Stations necessary to complete Plan B according to Arkin's model.

Thus, the sampling for Plan B is based on:

(1) Stations chosen in Steps #1 and 2 from this district.
(2) Stations chosen in this step to complete the sample.

Step #4 : Plan C (for each UNIVERSE)

As for Plan C, it is desirable to give the districts free scope to make their own analyses according to local needs. Thus, there are two possible scenarios.

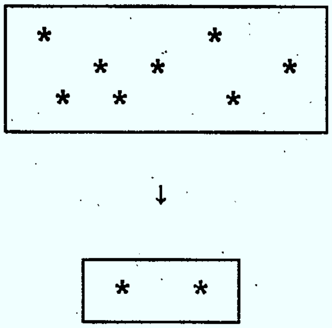
(1) Plan C is carried out using the national exclusions. In this case, the data from Plans 0, A, and B may be used, and the sample only has to be completed in accordance with Arkin's formula.

(2) New exclusions are defined for this universe. In this case, the previous data (Plans 0, A, and B) may NOT be used, due to the incompatibility of data. Thus, a COMPLETE sample must be obtained, using the locally defined exclusions.

These two scenarios are illustrated on the next page.

(1) National Exclusions

The stations chosen for Plans 0, A, and B from this universe may be used. Thus, the sample must be completed in order to obtain the size required by Arkin's formula.



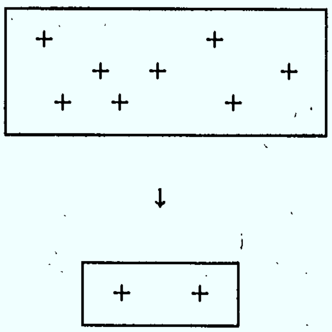
All stations, LESS those already chosen for Plans 0 (zero), A, B and C.

Stations necessary to complete Plan C according to Arkin's model.

Thus, the sample for Plan C is based on:

(1) Stations chosen in stages #1, 2, and 3 from this universe.
(2) Stations chosen in this stage to complete the sample.

(2) Local Exclusions



All stations in the universe after local exclusions.

Sample, Plan C.

The size of the sample is determined by Arkin's method, adding 15% as a safety factor.

3.1.3 Definition of Universes

As the result of implementing the recommendations for plans A, B, C, and 0, it will become much easier to give the districts the freedom to define their own universes without reducing the validity of the regional and national reports. Indeed since the data are collected "from the top down" in terms of geographical areas, different universe splits will have NO effect on the distribution of stations for plans 0, A, and B. However, it is essential that the exclusions defined at the national level for plans 0, A, and B be maintained. For plan C, the national exclusions will be used "by default"; it is possible for a district to redefine these exclusions to meet specific needs, but only for plan C, which is different from plans 0, A, and B.

Despite the freedom that will be given at the district level, we do NOT recommend the creation of a large number of universes to carry out plan C, as sampling all these universes will require a large number of resources (more than 200 data samples per universe).

3.1.4 Quality Control

In order to give more credibility to the results, which already have a margin of error, the collection of data should be as homogenous as possible throughout the country. Thus, the following points are recommended:

1) Revision of the different types of discrepancies by a group of inspectors and managers, in order to produce a homogenous method of data collection.

2) During the revision of the discrepancies, it would be possible (and desirable) to suggest that some "reception" verifications be made. This method of operation would complement the tests that should be carried out in the field. The purpose of this proposal is to allow the maximum verification for each station sampled.

3) Based on recommendations 1 and 2, a training program should be set up for inspectors, describing the data collection methods used and the objectives of the statistical plan in which they are participating.

4) Apply "overall quality control" at the national and regional levels by district office (not by inspector). A method of carrying out this control is described here.

"Overall Quality Control"

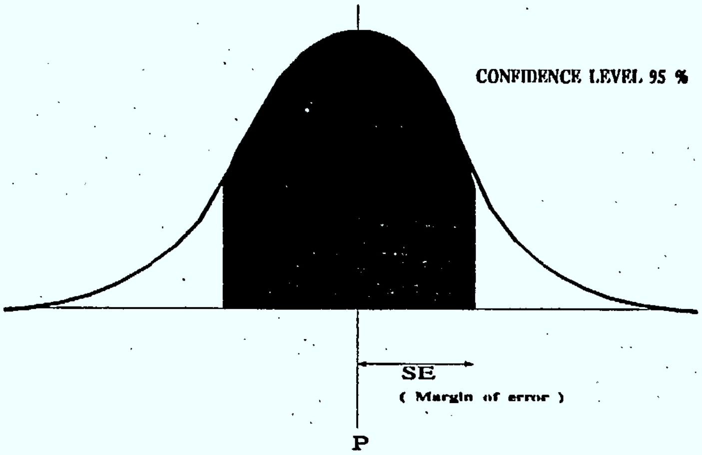
(a) Main Points of the Method

- control by district office. No data at the inspector level.
- comparison of discrepancy rates, district vs. national average.
- comparison of discrepancy verification rates, district vs. national average.

- analysis of data collection and of data as such.

(b) Methodology

At the national (or regional) level, for all discrepancies, the value of the discrepancy rate and the associated error should ideally follow a normal distribution:



Where p_1 is the discrepancy rate in a district, p is the national rate, and SE is the associated error. With a 95% confidence level, the probability of having $p_1 > (p+SE)$ is 2.5%, and that of having $p_i < (p-SE)$ is also 2.5%. We will define these as "extreme" conditions.

The following data are compiled for each discrepancy:

- National discrepancy rate, and associated error (%).
- National verification rate of the discrepancy studied (%).
- Number of inspections done in the district*
- Number of verifications done in the district per discrepancy studied*
- District verification rate and comparison **(i)**
- District discrepancy rate and comparison **(ii)**

* The number of inspections should be at least 30 in order to obtain a significant value for **(i)**. The number of verifications per discrepancy should also be at least 30 in order to obtain a significant value for **(ii)**.

(i) Verification rate per discrepancy

Here we compare the number of times a discrepancy is actually verified in the district as a function of this rate at the national level. Let us call this quantity "r". Thus:

$$r = \frac{\text{\% of verifications in the district}}{\text{national \% of verifications}}$$

Ex:

If r=1, then the % of verifications in the district = that of the country.
If r=.5, the % of verifications in the district = half that of the country
If r=2, the % of verifications in the district = twice that of the country

Thus, an r value that is less than 1 indicates a less than average level of verification, and vice versa.

(ii) Percentage of discrepancies of the district

The discrepancy rate for a district is calculated and:

(a) If the rate is higher than the national average, we enter ">", or ">>>" in an extreme case[2].
(b) If this rate is lower than the national average, we enter "<", or "<<<" in an extreme case[2].
(c) Equality is represented by "=".

(c) Example

Discrepancy considered: #8

National Discrepancy Rate: (8.6 ± 2.0)%

National Verification Rate: 63.6%

Some Districts (1989-90)

DISTRICT	INSPECTIONS	% TEST	(i)	% DISCREPANCIES	(ii)
5672	277	75.8	1.2	2.9	<<<
5641	22	*	*	*	*
5623	416	11.5	0.2	12.5	>>>
2640	183	89.6	1.4	1.8	<<<
1640	33	84.8	1.3	10.6	>
2630	288	100.0	1.6	1.1	<<<

[2]As defined at the end of the previous page.

Meaning of terms:

*	:	indicates insufficient data
%Test	:	indicates the percentage of verifications of this discrepancy
%Disc.	:	indicates the percentage of discrepancies in this district.
(i)	:	see description on previous page.
(ii)	:	see description on previous page.

Note:

The above table shows information per discrepancy. The same thing may be done for a district, by presenting all the discrepancies on a single table identical to the previous one, except that, in the first column, we would include all discrepancies considered. The following is an example of such a table:

Some Discrepancies (hypothetical data)

DISCREPANCY	INSPECTIONS	% TEST	(i)	% DISC.	(ii)
1	200	75.8	1.2	2.9	<<<
2	25	*	*	*	*
etc ..					

d) Comments

1) We can see that the situation is very variable between districts, even with only a few cases. In fact, many are at the extreme points of the normal distribution, which may indicate differences in data collection between districts and regions.

2) We may also question the choice of Code "A - test done with no problems" as a default value during data entry. The use of Code "0 - test not carried out" as a default would probably be safer. See document [IPC 3.21.01, 1991] for details.

3.1.5 Statistical Reports

After analyzing the 1989-90 reports, we can say that there are certain problems with the SCORE software. Furthermore, if we use the technical on-site sampling document [IPC 3.21.01] as a basis, we can see that the method suggested to obtain the confidence intervals consists of using formula [2], taking each **universe** as the basic element (what statisticians call "strata"). However, **the following points suggest that this method should be revised:**

1) In the production of regional and national reports, there will be relatively few stations in some universes, as these represent a small proportion of the total population. For the smallest universes, it is even possible to have only one or no stations sampled in an area. If only one station is sampled, the calculation of errors using formula [2] will produce a division by zero, which is to be avoided at all costs! In his book, L. Kish, deals with various sampling techniques, and arrives at the same conclusion, which is to say that "...at least two samples must be obtained from each stratum to be able to calculate an unbiased variance estimate" [Kish, p.101].

2) For plan C (only), we wish to leave the districts free to determine their own exclusions. In this case, we must base higher-level reports on the district discrepancy percentages, because, if we use universes, since they are not defined with the same exclusions as those used at the national level, we would be unable to maintain the national exclusions, and would mix two different populations.

Thus, it is desirable to use formula [2], weighted on the basis of **districts**, for the regional and national reports, as long as there are at least two samples per district. Thus, the universe report will be the only one to use plan C data, given that the exclusions there may be different. The production models for statistical reports at all levels are described next, with detailed examples. It is important to remember that formula [1] is a simple **proportion** between the number of discrepancies observed and the number of tests carried out, while formula [2] is a **weighted** calculation of discrepancy rates.

(a) Universe Reports

For a given universe, let:

a : number of discrepancies observed (Plan C only)
n : total number of verifications carried out (Plan C)
N : population size (universe, after LOCAL exclusions)
t : confidence threshold parameter (t=1.96 for 95%)

Thus we have:

$$p = \left(\frac{a}{n}\right)$$

$$\varepsilon = \pm t \sqrt{\frac{(N-n)p(1-p)}{N(n-1)}}$$

Formula [1]

where:
p : estimation of the discrepancy rate
ϵ : error in the estimation in the discrepancy rate

Then, we obtain the necessary values for the discrepancy report of this universe. This calculation is done for EACH discrepancy considered during sampling (ex: 1 to 15).

If we wish to produce a report limited to a particular source, the statistical method is the same, with one exception. The difference that should be considered, is that we know the TOTAL population of the universe (N), but do NOT know the population for a specific source. Thus, this value must be estimated using the results of our survey.

Let:

a, n, N: As above

a_f : Number of discrepancies observed for THIS source
n_f : Number of tests carried out for THE Source under consideration.
N_f : Portion of the population THIS source serves (Unknown data we wish to estimate).

Thus, we estimate $N_f = N(n_f/n)$, and use formula [1] above, using a_f, f_f, and N_f in place of a, n, and N, respectively. Once again, the calculation should be done for EACH discrepancy considered during the survey.

Notes:

(1) Reports for sources who hold only a small market share will have a very large margin of error, and thus should not be used for precise conclusions.

(2) n should be >=2 (respectively n_p>=2) to avoid dividing by zero in formula [1]. Exceptions should be taken into account for coding. In any event, if n<2, no conclusion can be drawn.

(b) District Reports

District Reports are obtained in the same way as universe reports. The only changes to the method above are:

1 - Replace UNIVERSE with DISTRICT
2 - Replace PLAN C with PLAN B

N.B: Stations in Districts registered in Plans 0 (zero) and/or A are automatically part of Plan B.

(c) Regional and National Reports

Given the cost of each data collection, it is desirable to obtain the maximum amount of information possible during the production of reports, in order to reduce the margins of error. Thus, we recommend using data from Plans 0, A, and B for the region (or country), weighting the results by district office. This is possible because Plans 0, A, and B are based on the same exclusions.

Let:

N_i	:	population of the i^{th} district
n_i	:	number of tests performed in the i^{th} district
d_i	:	number of discrepancies found in the i^{th} district
p_i	:	estimation of the discrepancy rate in the i^{th} district
N	:	total population of the region (or country)
t	:	confidence level parameter (t=1.96 for 95%)

Thus, we have:

$$p = \frac{\sum_i (N_i \times p_i)}{N}$$

$$\epsilon = \pm \frac{t}{N} \sqrt{\sum_i \left(\frac{N_i (N_i - n_i) p_i (1 - p_i)}{(n_i - 1)} \right)}$$

where:

$p_i = a_i/n_i$

Σ	:	Sum of the various districts
p	:	Estimation of the discrepancy rate
ϵ	:	Estimation error of the discrepancy rate

Formula [2]

Notes:

(1) It is essential to have at least two samples per district in order to avoid division by zero. If this is not the case, ignore the district in question and indicate that the report is incomplete.

(2) Given the local nature of the source codes, there are no regional or national source reports.

(d) Examples:

Remember: A "1/2" exponent is in fact a square root.

(1) Discrepancy ratio for universe X - discrepancy #1

Where: N = 1000 (population of X after exclusions)
n = 175 (number of tests **carried out**, discrepancy #1 and Plan C)
a = 25 (number of times discrepancy #1 was found)

Formula [1] produces:

$$p = 25/175 = 0.143 = \mathbf{14.3\%}$$

$$\epsilon = \pm(1.96)[\ (1000-175)(.143)(.857)\ /\ (1000)(174)\ ]^{1/2} = \pm .047 = \mathbf{\pm 4.7\%}$$

The probable number of stations involved is thus:

(.143) (1000) $\pm$ (.047) (1000) = **143** $\pm$ 47

(2) Ratio for <u>one source</u> for district Y - discrepancy #1

Where:	N =	2000 (TOTAL population of Y after exclusions)
	n =	250 (TOTAL number of tests carried out, disc. #1 Plan C)
	n_f =	100 (number of tests carried out for THE source)
	a_f =	20 (number of discrepancies for THE source)
Therefore:	N_f =	2000(100/250) = 800 (population estimation)

Formula [1] with n_f, a_f and N_f produces:

$$
\begin{aligned}
p &= 20/100 \\
&= 0.2 \\
&= 20\%
\end{aligned}
$$

$$
\begin{aligned}
\epsilon &= \pm(1.96)[(800\text{-}100)(.2)(.8) \;/\; (800)(99)]^{1/2} \\
&= \pm.074 \\
&= \pm 7.4\%
\end{aligned}
$$

Therefore we have: **(20.0 +/- 7.4)%**

(3) Regional Report - discrepancy #1

Let us assume we have a region composed of 3 districts, which we will represent by indices 1, 2, and 3. And let the data (Plans 0, A, and B):

$N_1 = 500$	$n_1 = 50$	$a_1 = 5$
$N_2 = 2000$	$n_2 = 200$	$a_2 = 10$
$N_3 = 5000$	$n_3 = 250$	$a_3 = 50$

Therefore, formula [2] produces:

$$
\begin{aligned}
p &= [500(5/50) + 2000(10/200) + 5000(50/250)] \;/\; (500+2000+5000) \\
&= 1150/7500 \\
&= 0.153 \\
&= \mathbf{15.3\%}
\end{aligned}
$$

$$
\begin{aligned}
\epsilon &= \pm(1.96/7500)[E_1 + E_2 + E_3]^{1/2} \\
&= \pm 0.034 \\
&= \mathbf{\pm 3.4\%}
\end{aligned}
$$

With:

$$
\begin{aligned}
E_1 &= 500(500\text{-}50)(5/50)(45/50) \;/\; 49 = 413.2653 \\
E_2 &= 2000(2000\text{-}200)(10/200)(190/200) \;/\; 199 = 859.29648 \\
E_3 &= 5000(5000\text{-}250)(50/250)(200/250) \;/\; 249 = 15261.044
\end{aligned}
$$

The probable number of stations involved is then:
$0.153\ (7500) \pm 0.034\ (7500) = \mathbf{1148 \pm 255}$

3.1.6 Estimate of the Number of Illegal Mobile Units

Discrepancy #11 is currently used to indicate the presence of unlicensed related stations. Thus, the compilation of these results does not allow us to evaluate the number of illegal mobile units, and the resulting loss of earnings. In order to make this information more useful, the following data should be collected for each (base) station sampled:

1) Total number of authorized mobile units (a_i).
2) Total number of mobile units in service (s_i).

Data from Plans 0, A, and B will be used at the district level. Thus, we will estimate the loss of earnings (probable number of illegal units) by DISTRICT, and total the sum for the regions and the country.

Thus, for a district, let:

$A = \Sigma a_i$, the total number of AUTHORIZED mobile units inspected.
$R = \Sigma r_i$, the ACTUAL number of mobile units found.
N = The total number of authorized mobile units in that district.

The proportion of illegal mobile units is : $\mathbf{p = (R-A)/R}$

If $\mathbf{A > R}$, the survey indicates that there is NO loss of earnings. Thus, we obtain a negative value for p, and indicate that the loss of earnings for that district is NIL.

On the other hand, it is much more probable that $\mathbf{A < R}$, which is the case that interests us, as we wish to evaluate the number of illegal mobile units. In this case, we estimate the total number of stations with: $\mathbf{POP = N + I}$, where I is the illegal population. We can evaluate I, since we know that:

$$p = I/(N+I)$$

$$\Rightarrow \quad I = N*p/(1-p)$$

$$\Rightarrow \quad POP = N + N*p/(1-p)$$

Using formula [1], we then have:

$e = \pm\, 1.96*[(POP-R)p(1-p)/(POP*(R-1))]^{1/2}$

where "e" is the error associated with proportion "p".

Thus, with a confidence level of 95%, the number **Q** of illegal mobile units in this district is:

$$\mathbf{Q = (p \pm e)*POP}$$

For a region or for the country, the probable number of illegal units is simply the sum of the probable number of illegal units in each of the districts where A < R, and the associated error is the sum of the absolute errors.

Example:

Total number of authorized mobile units in the district: 10,000 = N
Total number of authorized mobile units sampled: 1000 = A
Total number of mobile units sampled: 1200 = R.

THUS:

$p = (1200-1000)/1200 = 0.167$
$I = 10000*(1/6)/(5/6) = 2000$
$POP = N + I = 12000$
$e = \pm 1.96*[(12000-1200)(1/6)(5/6)/(12000*1199)]^{1/2} = 0.020$
$Q = (0.167 \pm 0.020)*12000 = 2000 \pm 240$

3.1.7 Processing of Multiple Choices

The new description of the survey activity [IPC 3.21.01] includes a certain number of "verifications" dealing with the client's satisfaction. Unlike Boolean (logical) discrepancy measurements, these data allow several possible responses. Using hypothetical data, the type of report to be presented in such a case is as follows:

Title	
Option 1	40 %
Option 2	35 %
Option 3	25 %
Total	100 %

(1) Number of respondents with no opinion: 250
(2) Number of respondents with opinion: 1000.
(3) Total number of responses taken into account: 1250.
(4) Number of responses eliminated: 10.

Meaning of Terms:

(1) Corresponds to code K in [IPC 3.21.01, 1991 draft], or "does not know".

(2) Corresponds to various possible response options. ONLY these data are compiled in order to obtain the percentages described above. In the example above, there were 400, 350, and 250 responses corresponding to options 1, 2, and 3 respectively, which gives 40%, 35%, and 25%.

(3) The sum of (1) and (2).

(4) Corresponds to code L of [IPC 3.21.01], or "not applicable".

Margins of Error:

These margins are calculated using formulas [1] and [2] separately for each option. For the population, only respondents with opinions (1000 in our example) should be used.

3.1.8 Comments on Normal Distribution

(1) In general, any sample of under 30 stations will not have a normal enough distribution to provide useful results, which is indicated by a margin of error. Thus, it is very important to consider the margin of error before making any conclusions about a result.

(2) Confidence intervals obtained for the proportions are of the form ($p \pm e$). In cases where p is very small, ($p-e$) $\langle$ 0. In such cases, the *Poisson* distribution will produce a more precise result, but the calculation method will be very long. In any case, when p is low, the discrepancy rate is low and "everything is OK". Thus, it is sufficient to use the approximation ($p-e$)$=0$, as it is absurd to have a negative percentage. This prevents the method used from getting overly complex.

(3) When $p=0$, we obtain $e=0$, but the comments made in (2) still apply.

3.2 Spectrum Surveillance

The idea behind spectrum surveillance, that is, measures taken to verify some parameters without having to notify the station owner, is certainly valid. Nevertheless, it would be very difficult, both on the practical and theoretical levels, to draft a valid national or even regional plan. Thus, it is recommended that **spectrum surveillance activities be ended**, but that some "reception" verifications be added to the on-site sampling activity.

With the introduction of plan 0, and the cancellation of spectrum surveillance, the available resources could be used to obtain a national picture of the situation by service. Some services among the ones listed in Appendix 10 of this document [Radio Regulations] could then be sampled (on-site) by applying plan 0 at the national level, which does not require an enormous amount of data to be collected in each region. Several different services could be analyzed, and the specific problems of each service could then be easily identified. For example, the services could be initially distributed as follows:

1. Maritime
2. Aeronautical
3. Land Fixed

3.3 Sampling Costs

For the year **1988-89**, sampling costs were as follows[3]:

On-site fixed sampling: $920,600 (Volume=2775)
Spectrum surveillance: $940,665
TOTAL: $ 1,861,265

For the year **1989-90**, sampling costs were as follows[4]:

On-site fixed sampling: $1,018,533 (Volume=3210)
Spectrum surveillance: $849,625
TOTAL: $ 1,868,158

After Modifications:

In order to estimate costs, we will use the 1989-90 data to estimate the cost per inspection: (1018533/3210)=$317.30. Let us assume 250 inspections per complete plan, which is a good approximation. Therefore:

Cost of Plan 0 (volume 250): $79,325, or an average of $15,865 per region. Cost for Plans 0 and A in all regions (volume 14250): $396,625.

Thus, the supplementary cost to go from Plan 0 to Plan A is $317,300, or an average cost of $63,460 per region.

For Plan B, the highly variable district size makes the application of such an average less significant, but the supplementary cost for a district that goes from Plan A to Plan B should be about $60,000 (about 200 extra data collections).

Finally, the cost of a Plan C independent of Plans 0, A, and B would be about (volume 250): $79,325.

N.B.: Plan 0 is the new minimum national standard, which will allow the sampling activity to be maintained, even if a major drop in resources occurs.

---

[3]1988-89 DOSP-C Annual Report.

[4]1989-90 DOSP-C Annual Report.

4. Investigation

In this section, we will concentrate on the investigations carried out when a complaint is received. These are **radiocommunication** (or "radiocom") investigations and **general public** investigations. The primary purpose of these investigations is to solve problems as and when they happen, within a reasonable time frame. The data collected during a survey are written on a form and compiled at various levels. Statistics can then be obtained in order to detect problems and trends in order to help plan any actions that should be taken.

4.1 Current Method

Report 16-902 (or its electronic equivalent on SCOMS) is divided into three parts (A, B, and C). Section A contains details such as the address of the complainant, the type of problems found, etc. Section B is a description of the actions taken by the inspector, and his comments throughout the survey. The section we are most interested in is Section C, as it is used to generate the survey reports. It contains the results of the survey in coded form, which are later available from a database. The following is an example of the format of this section, followed by a discussion concerning certain fields we no longer wish to retain.

INVESTIGATION REPORT SECTION C CONTROL: 99999
RESULTS NAME: XXXXXXXXXXXXXX

Discrepancies: 01 02 03 04 05 06 07 #T 08 09 #O 10 11 12 13 14 15 #R
Discrepancies: 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00

Stns. with discrepancies 0 Stns. examined 0 Sources

City: Svc: Intensity: Solution: No:

Company or Manuf:
Model:
C.O.S:

Time T: 0.0 D: 0.0

Investigation carried out by: Approved by
Inspector: Supervisor:
Date: Date:

Discrepancies (#1 to 15 T.O.R.):

In the document given to the inspectors (IPC 3.1.4, Appendix A), they are asked to write, for each discrepancy, a number that represents the number of stations where they have identified the discrepancy in question during the course of the survey.

For example, let us assume that, during a survey, we have to visit two stations, and that discrepancy #3 (excess power) is found at both stations, and discrepancy #1 (off-frequency) is found at only one station. We would thus write 2 in the "discrepancy #3" field, and 1 in the "discrepancy #1" field.

On the other hand, the document does not mention that the original purpose of these fields was to report all the discrepancies <u>**that were the CAUSE of the complaint**</u>, and not ALL the discrepancies found.

Even though it is desirable to inform the licensee about any irregularity, we should emphasize those that are the SOURCE of the problem, <u>**but there is already a space where that can be entered**</u>; that is, in the "source" field (which will be discussed soon). Thus, there are redundancies between the "discrepancy" and "source" fields. Moreover, due to its ambiguities, the "discrepancies" field is not useful when trying to obtain significant statistics. Thus, we recommend dropping the "discrepancy" field, and entering the problems caused by radio stations in the "source" field.

Stations with Discrepancies:

The total number of radio stations where discrepancies were found is written here. This field is currently compiled in some reports (SMIS Report V1), but this information is not useful for reaching any conclusions. Thus, we recommend dropping it.

Stations Examined:

This is the total number of radio stations examined. Once again, this field should be dropped, as it is of very little use in the various reports.

No:

The number of homes affected is supposed to be written here. However, in practice, it is not possible to obtain this number for obvious reasons. Thus, this field serves no purpose and should not be retained.

4.2 Analysis and Recommendations

In this section, certain specific points about investigations will be analyzed one by one, and recommendations made if necessary. All the recommendations apply both to radiocommunication investigations and general public investigations. On the other hand, some types of codes may apply specifically to one of these two types of investigations, but it is still desirable that coding be uniform.

4.2.1 Investigation Without Visits

The documents describing procedures to follow during investigations encourage solving problems by telephone. In many cases, sending a simple information brochure may be enough to solve a problem. If we add a simple field:

"Number of Visits? ____"

we can then obtain the following information:

1) No-visit solution rate.
2) Impact of information brochures as a solution.
3) Help in analyzing survey-related costs.

4.2.2 Symptom Code

Taking into account the fact that we wish to be able to analyze these reports in order to detect problems effectively, it is important to have coherent source codes, and they should actually represent the **sources**, and not symptoms of the problem, which is often the case at present.

Example:

For example, it is possible to note an intermodulation problem (new signal produced by the "mixing" of two or more other signals). However, if one of the stations causing the intermodulation is in a location other than that indicated on its license, the REAL source of the problem for this station is discrepancy #5 (DO5 - wrong location). Thus, the intermodulation is a **symptom** of the problem. With the current codes, it would be possible to enter either DO5 or D23 (intermodulation) as the **source** of the problem, which is ambiguous.Thus, a new "symptom" field will be introduced for survey coding. This field describing the symptom of the problem independently of the source is new. Thus, it is not in the reference list. The purpose of this report is not to draw a complete list of all the symptom codes. Nevertheless, we have included a "starting" list in Appendix 7, after consultation with the regions.

The addition of a new field may seem to be an additional complication at the data collection level, but this is not actually the case. In fact, once a list of symptoms is accepted, it will be advantageous to automate the entry of the description of a problem into the SCOMS program, by allowing the user to "choose" from this list, while allowing entry of different text for exceptions. Thus, symptom coding could be done automatically, according to the description chosen in the list.

4.2.3 Source Code Z[1]

There are two very different cases in which we use source code Z: "Z001 - unjustified complaint" and "Z009 - unidentified source". There are many investigation with this source code, which may seem surprising:

1987-88:	2879 source Z investigations
1988-89:	2678 source Z investigations
1989-90:	2502 source Z investigations

It would seem important to separate the two code Z types, as many source compilations only keep track of the first letter, which is confusing in the case of source code "Z". For example, the majority of source Z complaints in previous years (more than 2000) have been from unidentified sources, or "Z009".

We could keep "Z001" and leave the source code blank (by default) if no source is identified. We should also determine whether the current list of source codes is incomplete.

4.2.4 Source Code "D"

Source "D" investigations account for over half of all radiocom investigations. This source code includes several other types of sources, and the following breakdown is suggested in order to provide greater accuracy during the interpretation of investigations:

- Technical discrepancies (currently D01 to D07)
- Operational discrepancies (currently D08 and D09).
- Legal discrepancies (ex. un-licensed station, certificate).

Furthermore, code D24 represents an unidentified source. This code should be treated similarly to current code Z009.

[1]For a detailed list of source codes, consult document [RIM 3.1.3, 1986, Appendix C]. A general list is provided on the next page.

4.2.5 Recommendations, Source Codes

For reference purposes, the following is the current list of source codes:

C	:	Commercial, industrial, or domestic equipment
D	:	Radiocommunications System
I	:	Scientific, industrial, and medical devices
L	:	Electric power
R	:	Interference caused by a receiver
S	:	Signal problem
T	:	Internal combustion engine
W	:	Radiation from cable distribution systems
X	:	Immunity/blocking problem
Y	:	Equipment with no radiation
Z	:	Other

Codes L, S, W, X, Y and C are used frequently, but codes I, R, and T are very rarely used, as can be seen in Appendix 4. These codes, as well as code D, are the main data used for survey analysis. Thus, it is very important that these letters provide as much information as possible on the general situation.

In order to answer the problems raised in the preceding sections, without completely overthrowing the current coding method, the following changes are proposed:

(1) Given that the analysis concentrates primarily on the first letter of the source codes, and not on the complete codes, we recommend splitting source code D into three components: technical, operational, and legal discrepancies. **Thus, it is very important that the numeric part be recorded during the surveys [IPC 3.21.01, 1991].** This recommendation will be particularly useful for "radiocom" investigations, but should also apply to general public investigations, in order to avoid confusion.

(2) Remove source codes D24 and Z009, which are used when no source is identified. In order to simplify things, we will leave the "source" field blank when no source is identified. Thus, code Z will only be used for unjustified complaints.

(3) Remove code D23 (intermodulation), which is actually a physical explanation of a phenomenon.

(4) Given the very low number of sources I and T, we recommend combining these codes with code C, in order to put together all sources attributable to **equipment**.

(5) Given the very low number of source R, we recommend putting together code R with code Y, in order to have a single source code for all sources attributable to **receivers** or **antennas**.

(6) Add a character to the end of each source code for "radiocom" investigations, in order to indicate the "responsible party", according to the model described in document [IPC 3.21.01, Section III.E]. In radiocom investigations, the license holder or supplier may be identified as responsible. The same model may be used for general public investigations, identifying the user or the equipment (i.e. the

supplier) as responsible. Given the necessary supplementary space, we could easily reduce the number of available "source" codes to 2 or 3 (currently 6), which would be quite sufficient.

Thus, the source codes would be:

C	:	Commercial, industrial, or domestic equipment
D	:	Radiocommunications System - technical
E	:	Radiocommunications System - operational
F	:	Radiocommunications System - legal
L	:	Electric power
S	:	Signal problem
W	:	Radiation from cable distribution systems
X	:	Immunity/blocking problem
Y	:	Receivers or antennas
Z	:	Unjustified
Blank	:	Unidentified

4.2.6 Action Codes

This section is used to show what has been done in response to the complaint. The current codes are:

D - Use discontinued
F - Problem repaired or measures taken
N - No economical solution possible
W - Survey cancelled
Z - Survey refusal

In order to make coding more precise, the following list is proposed:

D - Use discontinued
F - Problem repaired or measures taken
I - Effect modified to the satisfaction of the person making the complaint.
N - No economical solution possible
S - Problem solved at source
W - Survey cancelled (problem stopped during survey)

4.2.7 Currently Produced Reports

The main purpose of the investigations is to answer user complaints, but the main purpose of **coding** is to be able to produce meaningful reports to aid in the decision making process. In the 1989-90 DOSP-C Annual Report, the following data and graphs were presented for general public and radiocommunication investigations.

- Total number of complaints and change from previous years.
- Person-years necessary and change from previous years.
- Number of complaints by source.
- Number of complaints by region.
- Number by discrepancy for code D in radiocommunication investigations.

In Section 4.3, we will suggest other methods of using the survey data, in order to draw useful conclusions.

4.2.8 Summary of Recommendations

The different coding stages suggested are as follows:

(1) Receipt of the complaint, general information.

(2) Problem Symptom Code.

(3) Problem Source Code(s).

(4) Action Code.

Thus, the following is the suggested format for Section C of Form 16-902, for **radiocom and general public investigation only**. This format is based on the current format. The modifications made are based on all the recommendations made in Section 4.

Investigation report form

Investigation Report Results	Section C	Control 9999 Names
Visit No.:		
Symptom:		
Sources:		
City: Svc: Intensity:	Solution:	
Company or Manuf:		
Model:		
C.O.S.		
Time T: 0.0	D: 0.0	
Investigation carried out by	Approved by	
Inspector:	Supervisor:	
Date:	Date:	

4.3 Suggested Reports and Analyses

This section will describe the reports and analyses that could be done from year to year in order to draw useful information from the survey data. Thus, this section is extremely important, as data collection is useless if the data cannot be interpreted. The joint use of sampling and survey data will be discussed in Section 5, which deals with the possible relationships between these two activities.

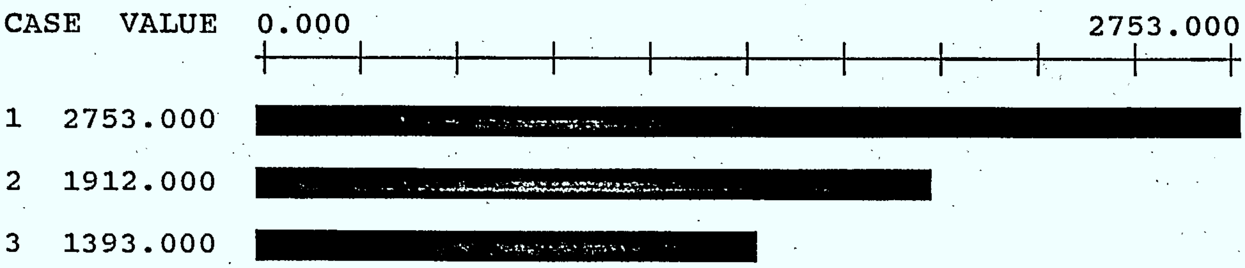
4.3.1 Evolution by Source Code

One of the descriptive methods that can be used to evaluate the situation would be to compare the **absolute** number of investigations by SOURCE from year to year, to detect any particular trends. We used the 1987 to 1990 data and the Mystat software for the following example. "CASES" 1 to 3 correspond to fiscal years 1987-88, 1988-89, and 1989-90 respectively.

Example (Mystat Program), source L:

```
PLOT OF          L
NUMBER OF CASES =     3
MEAN OF SERIES =     2019.333
STANDARD DEVIATION OF SERIES =     560.381

SEQUENCE PLOT OF SERIES
```



4.3.2 Compilations

The annual DOSP-C report describes the **relative** volume of each <u>source</u> code in the form of a pie chart. This approach should be kept, and the same "overall view" could be applied to <u>symptom</u> and <u>action</u> codes. These simple compilations do not allow precise conclusions to be drawn, but they provide a good "overall view" of the situation.

4.3.3 Analysis of Related Costs

Objective:

To describe a method of calculating the costs generated by certain sources, while making it possible to specify the conditions (filters) that should apply to the type of survey to be considered. The possible conditions are:

1) Type of survey (public, radiocom, or both).
2) Specific symptom code.
3) Specific responsibility code.
4) Type of service involved.
5) Specific action code.

We must then obtain, for a given source and an investigations conforming to the filter, the amount of time necessary for the investigations when there is a specific source, and for those when there is no specific source. We can then evaluate the costs involved, both for a specific source (eg: D083) and for a general source (eg. D). **<u>Thus, it is essential to have access to the real amount of time spent on each survey in order to be able to obtain a useful estimate of the costs generated by a specific source.</u>**

Methodology:

The following stages are only applied to investigations that meet conditions 1) to 5) above, if such conditions have been specified. These conditions will be found in the tables in the next sections.

1- Estimate the cost per unit of time and per unit of distance (T and D), which we will call C_t and C_d.

2- Specify the source code corresponding to the desired information, which we will call S.

3- For all investigations where source S is reported to be the sole source of the problem, add the time periods T and D, which we will call T_1 and D_1.

4- For all investigations where source S is reported to be the source of the problem (but not necessarily the only one), add the time periods T and D, which we will call T_2 and D_2.

5- Then, calculate $B_i=(T_1C_t+D_1C_d)$, the lower limit of the cost associated with the source, and $B_s=(T_2C_t+D_2C_d)$.

6- Finally, in order to simplify the reports produced, the average of the upper and lower limits will be used; let **Cost = $(B_i + B_s)/2$**. The margin of error will be very small, since the majority of sources are specific and thus $\mathbf{B_1 \approx B_s}$.

A good estimate of the costs generated by source S is given by the interval: $[B_i, B_s]$, where B_i is the most optimistic scenario and B_s is the most pessimistic scenario. Furthermore, the average of these two limits is a reasonable estimate, and will be used in the reports.

Example:

Number of investigations where S is the only source:	500
Total time (T_1) of these investigations:	1500 hours
Total time (D_1) of these investigations:	250 hours.
Total number of investigations where S is one source:	525
Total time (T_2) of these investigations:	1700 hours
Total time (D_2) of these investigations:	300 hours.

Cost of working time T_i:	$75/hour
Cost of travel time D_i:	$75/hour.

Therefore,

B_i = $(1500*75 + 250*75) = **$131,250**
B_s = $(1700*75 + 300*75) = **$150,000**

Thus, the probable cost of source S is in the interval [$131,250 , $150,000], and the average of upper and lower limits gives: **$140,625**; this is the value that will be used in the cost tables, which are described in the next paragraph.

4.3.4 Comparison by Source Code

A table is probably the clearest means of representing relationships between two sets of data. The SCORE software currently produces reports indicating the number of complaints by source and service involved, as well as by source and action code. This type of report is available for both "general public" and "radiocom" investigations. On the other hand, it will be important to allow the production of such a report at all levels (national, regional, and district). In SCORE, the regional and national reports are only a compilation of district reports, which is not very practical.

Furthermore, we recommend taking sources as a whole; which is to say dropping the distinction between primary and secondary sources. In order to analyze what happens at all levels of the investigation, the source codes should be analyzed in all the following ways:

a) Sources vs. Symptoms - Volume
b) Sources vs. Symptoms - Cost
c) Sources vs. Services involved - Volume
d) Sources vs. Services involved - Cost
e) Sources vs. Responsibility codes - Volume
f) Sources vs. Responsibility codes - Cost
g) Sources vs. Action codes - Volume
h) Sources vs. Action codes - Cost

It should be possible to produce these compilations for **specific** source codes (eg. D083) or for **classes** of source codes (eg. D). Please note that the "Volume" represent a number of investigations; the "Cost" calculation method was explained in paragraph 4.3.3. Thus, it is possible to produce 8 reports (four with volume, 4 with cost) for general public OR radiocom investigations, and for specific sources OR classes of sources. The report format should always be similar. The number of reports at this stage may seem excessive; but it is important to keep in mind that this is an **annual** compilation aiming to provide information on many subjects at various levels. Furthermore, all these compilations may be computerized, which means less loss of time in producing the reports, and thus more time for analyzing them. Finally, experience will show if some of these reports turn out not be useful. If this is the case, it would then be desirable to abandon them.

Each row on the various tables represents a SOURCE (class or specific), and the columns represent the type of compilation (service, symptom, responsibility, or action). The next two pages contain 4 examples of such reports; two cost reports and two reports per volume of survey. The two reports of each type show the sources by class or in detail. These four examples contain the data for services, but only the names of the columns will change for the other possibilities, to present data by symptom, responsibility, or action code.

Examples:

NOTE: The numerical values on the tables are not drawn from real experience, as it is not possible to obtain these values with the current survey coding.

Compiled by source (details) and Service

							National Rapport Cost Estimate
General Public Investigation (HYPOTHETICAL FIGURES)						1990-91	
Compiled by source (details) et Service							
			Services affected				
Sources	AM	FM	TV	HEAD	MATV	ETC ...	TOTAL
C002	$ 500	$ 200	$ 0	$ 1 100	$ 12 500	$ 45 000	$ 59,300
C003	$ 100	$ 500	$ 0	$ 600	$ 32 500	$ 25 000	$ 58,700
C004	$ 30 000	$ 530	$ 0	$ 0	$ 0	$ 26 000	$ 56,530
C006	$ 0	$ 0	$ 1 600	$ 1 600	$ 16 500	$ 5 200	$ 24,900
C008	$ 20	$ 600	$ 200	$ 200	$ 22 500	$ 0	$ 23,520
C011	$ 0	$ 800	$ 100	$ 100	$ 5 200	$ 3 000	$ 9,200
C111	$ 0	$ 0	$ 2 100	$ 2 100	$ 0	$ 95 000	$ 99,200
D083	$ 0	$ 500	$ 100	$ 100	$ 30 000	$ 25 000	$ 55,700
ETC..	$ 0	$ 0	$ 0	$ 0	$ 95 000	$ 0	$ 95,000
TOTAL	$ 30,620	$ 3,130	$ 4,100	$ 5,800	$ 214,200	$ 224,200	$ 482,050

Compiled by source and service

National Report							Number of Complaints
General Public Investigation (HYPOTHETICAL FIGURES)							1990-91
Compiled by source and service							
			Service affected				
Sources	AM	FM	TV	STAC	HEAD	ETC ...	TOTAL
C002	0	10	0	0	300	4	314
C003	100	0	0	0	500	0	600
C004	0	0	0	0	0	0	0
C006	0	0	0	0	110	0	110
C008	20	0	0	0	0	0	20
C011	0	40	0	0	504	10	554
C111	0	0	90	0	300	10	400
D083	0	500	0	0	520	45	1,065
ETC ...	0	0	0	0	0	0	0
TOTAL	120	550	90	0	2,234	69	3,063

Compiled by Source (class) et Service

National Rapport							Cost Estimate
General Public Investigation (HYPOTHETICAL FIGURES)					1990-91		
Compiled by source (class) et Service							
Service Involved							
Sources	AM	FM	TV	HEAD	MATV	ETC ...	TOTAL
C	$ 500	$ 200	$ 0	$ 1 100	$ 12 500	$ 45 000	$ 59,300
D	$ 100	$ 500	$ 0	$ 600	$ 32 500	$ 25 000	$ 58,700
E	$ 30 000	$ 530	$ 0	$ 0	$ 0	$ 26 000	$ 56,530
F	$ 0	$ 0	$ 1 600	$ 1 600	$ 16 500	$ 5 200	$ 24,900
L	$ 20	$ 600	$ 200	$ 200	$ 22 500	$ 0	$ 23,520
S	$ 0	$ 800	$ 100	$ 100	$ 5 200	$ 3 000	$ 9,200
W	$ 0	$ 0	$ 2 100	$ 2 100	$ 0	$ 95 000	$ 99,200
X	$ 0	$ 500	$ 100	$ 100	$ 3 000	$ 25 000	$ 28,700
ETC..	$ 0	$ 0	$ 0	$ 0	$ 95 000	$ 0	$ 95,000
TOTAL	$ 30,620	$ 3,130	$ 4,100	$ 5,800	$ 187,200	$ 224,200	$ 455,050

Compiled by source (class) and service

NATIONAL REPORT						NUMBER OF COMPLAINTS	
General Public Investigations							1990-91
Compiled by source (class) and service							
Services affected							
Sources	AM	FM	TV	HEAD	MATV	ETC ...	TOTAL
C	0	10	0	0	300	4	314
D	100	0	0	0	500	0	600
E	0	0	0	0	0	0	0
F	0	0	0	0	110	0	110
L	20	0	0	0	0	0	20
S	0	40	0	0	504	10	554
W	0	0	90	0	300	10	400
X	0	500	0	0	520	45	1,065
ETC ...	0	0	0	0	0	0	0
TOTAL	120	550	90	0	2,234	69	3,063

4.3.5 Response Time to Complaints

A compilation of the % of complaints processed in 30, 60, 90 and 120 days by region and district should be presented in tabular form, along with the national average. The following is an example of such a presentation:

Hypothetical data - partial table

Area	30 days	60 days	90 days	120 days
District x	60%	80%	90%	94%
District y	51%	71%	87%	89%
Région z	55%	74%	88%	91%
*** CANADA ***	**63%**	**79%**	**91%**	**95%**

4.3.6 Detailed Source Codes

A more precise compilation of some source codes may seem useful. This type of analysis was already done in the 1989-90 DOSP-C Annual Report for source code D, in the form of a graph showing the number of source D investigation attributable to each discrepancy. The same type of detailed compilation by source code could be done in other cases, for example:

1) By category (eg. industrial, domestic), Code C
2) By type of apparatus for code C
3) By type of discrepancy for sources D, E, and F.
4) By type of service for codes D, E, F, and X.
5) By type of company for code L.
6) By type of system for code W.

These compilations could be presented in a table, in a bar graph, or in a pie chart.

4.3.7 Summary

The analyses described above could be carried out at all levels, for a district, for a region, or for the country as a whole. Everything that has been described in this section represents a large amount of information and many tables. On the other hand, all this information would only be compiled once a year, and it is much more useful to have too much information than not enough, especially when that information is **available**!

Some of the analyses described will probably be abandoned or modified if the conclusions are not useful, and new types of analyses and data presentations could be developed, based on several years experience. What is important is to leave room for imagination and changes, as it is impossible to develop a perfect data analysis method that provides precise answers to ALL the questions that may arise.

Recommendation

In order to allow the managers to use the results of the analyses described here, we recommend publishing all these results annually, at least at the regional and national levels. Thus, it would be useful to develop a software that would make it possible to obtain compilations, tables, and graphs described in this section efficiently.

The most important thing is to obtain **useful** reports and to make these results **visible**, since this is the only way to justify survey coding, and to have results that can be helpful to manage the spectrum.

5 Investigations/Sampling Relationships

Until now, we have considered the investigations and sampling activities of the Department of Communications in detail but separately. However, these two activities, which are carried out independently, may nevertheless have some connections. We will call these "external" relationships, as opposed to the "internal" relationships described for investigation. This study will be done in two parts:

1 - Use the historical data (1984-90) to carry out a statistical study on the correlation between the data. We will do the following analyses:

1 - Determination of alert thresholds for sampling discrepancies 1 to 15.
2 - Forecasting annual complaint volume on the basis of samples.
3 - Study of the impact of technical and operational discrepancies (1 to 9).

2 - Use the recommendations in Sections 2 to 4 concerning sampling and investigations coding. Obviously we cannot do a statistical study of the data, but suggestions will be formulated on any "eventual" relationships and how they may be used. The "Mystat" program will be used to obtain various data such as coefficients of correlation, parameters of curves generated by the least squares method, graphs, etc.

5.1 Analysis of Historical Data

5.1.1 Validity of Data

Before describing the statistical analyses themselves, some comments should be made on the data used. These data are for 1984 to 1990, from both investigation and sampling. One initial remark that should be made is the impact of the introduction of supplier codes in sampling after 1989-90. This impact is due to the fact that, at that time, a distinction was made between a negative test and a test that was not carried out. The following table shows the immediate impact of this improvement in the coding. We can see, in fact, that discrepancy percentages were systematically underestimated until 1989, due to coding limitations.

Table of Discrepancies

DISCREPANCY	AVREAGE 1984-89 (%)	1989-90 (%)	DIFFERENTIAL
1	3.0	3.8 *	+27%
2	1.2	3	+150%
3	14.6	24	+64%
4	5.0	7	+40%
5	5.8	12	+107%
6	8.2	14	+71%
7	2.6	4	+54%
8	1.2	8	+567%
9	16.2	32	+98%
10	0.2	1	+400%
11	7.0	13	+86%
12	1.2	3	+150%
13	11.4	24	+111%
14	4.4	1	-77%
15	0.2	1	+400%

* Indicates that the value obtained by SCORE was not kept due to a bias in the calculations used. The other values were obtained by SCORE, and have been verified. The results have a margin of error and have been rounded, but this does not change the general trend.

Impact:

Since this phenomenon is now recognized, we will put greater emphasis on more recent data without ignoring the others. Thus, there was an underestimation in the proposed alert thresholds, but, as we did not have enough data using the supplier codes, better a lower limit than nothing at all!

5.1.2 Alert Thresholds

At this time, there is a list of alert thresholds for each of the 15 defined discrepancies. These thresholds have been determined subjectively, without a detailed analysis of the historical data. Thus, the purpose of this section is to revise these thresholds as objectively as possible, using data from previous years. Three different methods were used to obtain the alert thresholds for the discrepancy rates. The following data were analyzed:

1 - Volume of complaints vs. discrepancy rates.
2 - Volume of "radiocom" complaints vs. discrepancy rates.
3 - Volume of source D or Z complaints vs. discrepancy rates.

In all three cases, the general method used for each of the 15 discrepancies was as follows:

1) The available sampling and investigation data to be compared were entered for each discrepancy.

2) The correlation matrix of the data entered was obtained, in order to be able to evaluate the degree of linear relationship between the data. Since it is possible to have a non-linear relationship, this matrix was used mainly to eliminate data pairs with a negative correlation. In fact, if a relationship exists between the discrepancy rate and the number of complaints, it is not logical to expect that this ratio would be inversely proportional, which indicates a high negative correlation.

3) The discrepancy graph was drawn as a function of the number of complaints for all cases where a relationship was suspected. Only cases where a conclusion could be reached were retained.

4) On the basis of all of these analyses, determine an alert threshold, if possible. To determine this threshold, one of the two following methods was used for each graph:

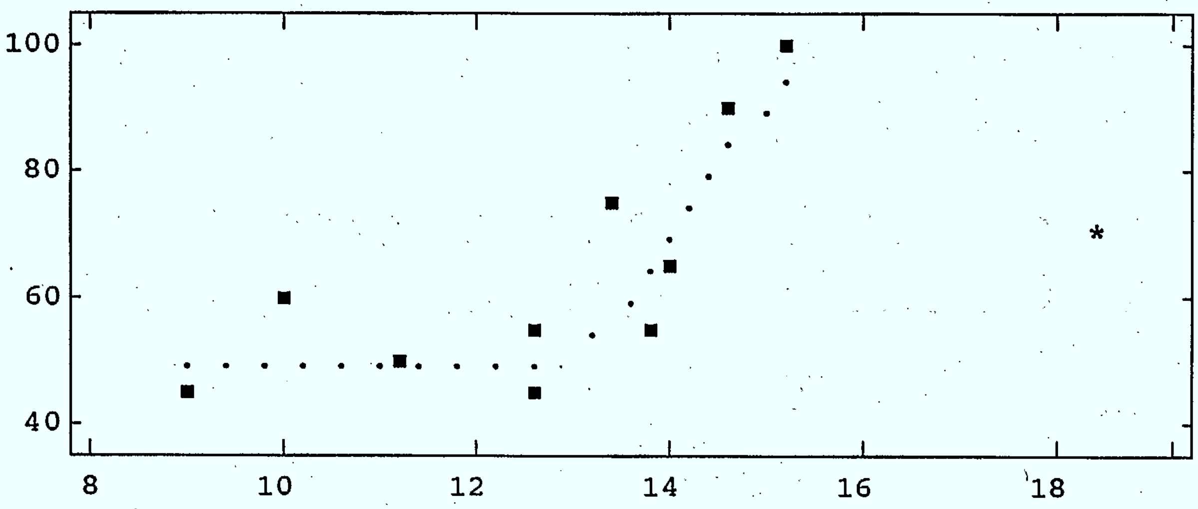
(i) If the approximation curve of the number of complaints as a function of the percentage of discrepancies tends to increase relatively quickly after a given value (for the percentage of discrepancies), this value can be used as the alert threshold.

(ii) If such a point cannot be determined, an attempt is made to determine whether the extreme points (high number of complaints) are all located beyond a given value for the percentage of discrepancies; if this is the case, this value can be used as an alert threshold.

Only the study done with source "D" complaints and, to a lesser extent, source "Z" complaints, showed a sufficiently strong correlation to suggest alert thresholds. Thus, these are the results presented. It is important to note that a statistical study cannot be more precise than the data used. As investigations were not always done in the same manner over the course of the period under consideration in the different regions, the following results contain an "intrinsic margin of error". The statistical details and the data are not included here, due to the volume of data that this would represent, and because most of the data are stored electronically. However, examples of the graphs obtained follow:

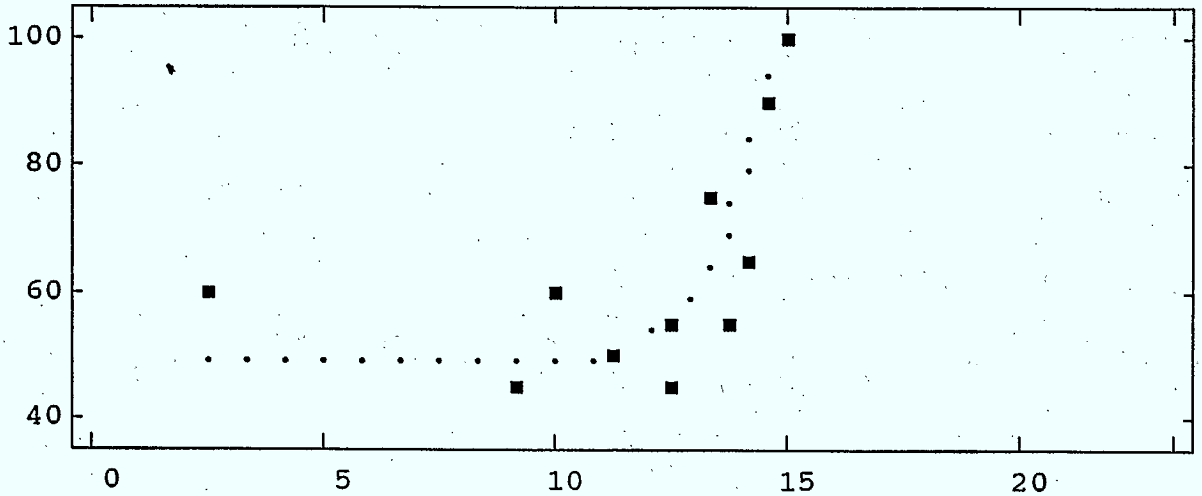
Examples:

Ratio #1 - complaints/population



Discrepancy #3(%)

Ratio #2 - complaints/population



Discrepancy # 3 (%)

These two graphs represent a comparison between "**number of complaints per 10,000 population**" and a "**percentage discrepancy**". The comparison is done with discrepancy #3, and the complaints for sources D and Z. The analysis performed is purely statistical, and is only based on the **data**. In this case, the curves tend to increase significantly around 12%. Other data and graphs have been analyzed, and the 12% threshold seems to be acceptable in most cases, and thus this threshold is recommended for discrepancy #3.

Recommendations

The following table is a compilation of all the results obtained, as well as the current alert thresholds, and those recommended based on the analyses performed.

Alert Thresholds

	Current threshold	Analyze stat.	**Recommendation**
Disc.#1	5% - 5%	4%	**4%**
Disc.#2	5% - 10%	?	**5% ***
Disc.#3	5% - 10%	12%	**12%**
Disc.#4	5% - 5%	5%	**5%**
Disc.#5	5% - 10%	8%	**8%**
Disc.#6	5% - 10%	10%	**10%**
Disc.#7	5% - 5%	10%	**5% ****
Disc.#8	10% - 20%	?	**10%**
Disc.#9	25% - 25%	20%	**20%**
Disc.#10	not verified	?	**? ***
Disc.#11	10% - 10%	8%	**8%**
Disc.#12	10% - 10%	?	**10% ***
Disc.#13	25% - 25%	20%	**20%**
Disc.#14	not verified	20%	**20%**
Disc.#15	15% - 15%	?	**15% ***

* : Indicates that the (minimum) current threshold is recommended, as the analyses did not make it possible to obtain another value.

** : This threshold was set at 5% due to the major impact of this problem on the radio spectrum, as can be seen on the table on page 53.

? : No conclusion possible on the basis of the available data.

Current Thresholds: These are the rates currently in use. Please note that two rates are set, one for high antenna park density regions, and the second for other areas. Our data did not allow us to make this distinction, which is why there is only one recommended threshold for each discrepancy.

Recommendation:

These recommendations should not be considered to be perfect alert thresholds, but rather an attempt to set these thresholds objectively, based on historical data.

Recommendations:

1) It should be remembered that the alert threshold results were obtained with data which did not differentiate between verified discrepancies with no problems and unverified discrepancies. In a few years, using new data **that make this differentiation** and as a result of the experience obtained with the thresholds recommended here, it would be worthwhile to revise these values as objectively as possible.

2) Note that alert thresholds were underestimated, for the reasons explained in paragraph 5.1.2.

5.1.3 Forecasts

One of the questions that arises when analyzing samples is at what point the results of these samples can serve to "forecast" the volume of future complaints. Several different methods were tried in order to attempt to obtain a mathematical model allowing to "forecast" the coming year's complaints.

Given that sampling, measures the compliance of radio stations, we concentrated specifically on **radiocom investigations** and **source codes D or Z investigations**. The latter analysis turned out to be the most accurate, due to the high correlation between the data, and the higher credibility of the results.

Forecasts, "Radiocom" Complaints (all sources)

It was impossible to find a model for this first test. In fact, the graphs comparing the data were too irregular to be able to detect reasonable approximation curves. In fact, it did not seem very useful to present these approximation models here, as they had margins of error of 50% or higher! Thus, we tried to determine whether the results would be better if we limited ourselves to specific types of complaints...

Forecasts, Sources D and Z

Source Z graphs were, again, too irregular to be able to find a valid model. On the other hand, we were able to determine several approximation models based on **source D radiocom investigation activity (D17)** in various ways. An example of the statistical details of the variance analysis and the production of a curve using the least squares method is presented in Appendix 5.

The three best models obtained were as follows:

1. **D17 = 16.3782 + 0.00646 (DISC1_7)2**
2. D17 = 18.4480 + 0.0012 $(\text{DISC1\_15})^2$
3. D17 = 18.7615 + 0.0024 $(\text{DISC1\_9})^2$

* N.B.: Value "D17" is, in fact, the ratio between the number of **source D radiocom complaints** and the **radio population** of this region (or country), expressed in tens of thousands of stations. The "DISC1_x" variables are a compilation of discrepancies 1 to x, inclusive.

It would take too long to present a comparison of these three models here with ALL the data used. A comparison of the most recent data (1989-90) using these three models is shown in Table 6, in order to provide an idea of the accuracy of these models:

Model	Average Deviation from Reality
Model 1	20.4 %
Model 2	30.8 %
Model 3	27.8 %

Furthermore, if we compare certain technical data from these three models (standard deviation of coefficients, variance analysis, p-value, etc.) we still conclude that the first model is the most accurate of the three.

Recommendations

In all cases, the survey data used dated from before 1989, and, at that time, there was no code to differentiate between analyses that were not carried out and analyses that were carried out with no problems. Thus, the data are not ideal and the models obtained can never be more precise than the data used. However, the following are the main conclusions that could be drawn from the preceding analyses:

1. Due to the very low correlation between available data, it does not seem possible to be able to estimate a reasonably accurate overall "forecast" model, even if we limit ourselves to radiocommunication investigations.

2. It also seems that it is only possible to estimate the number of source D radiocom complaints. Thus, this model is very specialized, as it only allows the number of source D complaints to be estimated (which corresponds to sources D, E, and F of the coding suggested in Section 4).

3. We will abandon the "forecast" approach in order to concentrate on a more promising method - "impact studies" of various discrepancies - as this approach will provide us with useful information in a simple way.

5.1.4 Study of the Impacts of Discrepancies 1 to 9

The study done in this section was inspired by, among others, a document entitled "Calculations of Tolerance Levels", published by the Quebec region in April 1984. It is an interpretation of the results of on-site sampling using survey data, an aspect of the "relationship" we wish to establish. The results presented here are based on a large quantity of data on the entire country, including those for 1989-90. Furthermore, we have tried to introduce a "classification" of discrepancies according to their impact, in order to obtain an easily interpretable portrait.

The main advantage of the method that we will use is its simplicity, which means that it can be applied quickly and easily year after year, thus providing an up to date picture of the condition of the spectrum.

Before presenting the "coefficients of impact" table, an example of how these numbers were determined will be discussed. These coefficients represent **the probable number of complaints per 1000 discrepancies.**

Example:

The 1989-90 population considered during the surveys was about 87,424. The results cannot be exact as the number of licenses varies constantly over a year, but the figure is precise enough, as we are interested in trends, not exact numbers. We obtained the following results for investigation and sampling:

Example of impact calculation[1]

Discrepancy	Rate	Stations with Discrepancy	Complaints	Impact
1	3.8 %	3322	17	5.1
2	3 %	2623	30	11.4
3	24 %	20982	35	1.7
4	7 %	6120	258	42.2
5	12 %	10491	123	11.7
6	14 %	12239	11	0.9
7	4 %	3497	507	145.0
8	8 %	6994	445	63.6
9	32 %	27976	29	1.0

For discrepancy #1 (for example), the number of stations with complaints is a calculation of the **probability**, obtained with the population and the discrepancy rate. The number of complaints (here, 17) is the total number of "radiocom" and "general public" complaints one of whose sources was identified as discrepancy #1. The coefficient of impact is obtained using the calculation: (Complaints/Stations with discrepancies)*1000. **Thus, a result is obtained that indicates that a discrepancy #1 will generate a complaint in 5.1 cases out of 1,000, on AVERAGE.**

[1]See Appendix 3, on page A-6, for the list of discrepancy codes.

The same calculations were done for three fiscal years. The results are summarized in the table on the next page. On the right side of the table, we introduce a distribution into five coefficient of impact categories, in order to simplify the interpretation of the results.

Comparison of Coefficients of Impact[2]

	Coefficients of impact			Level of impact *		
Discrepa.	89/90	87/88	86/87	89/90	87/88	86/87
7	145	436	495	2	2	2
8	64	4088**	2433**	3	1	1
4	42	47	59	3	3	3
5	12	26	23	3	3	3
2	11	81	183	3	3	2
1	5	8	52	4	4	3
3	2	4	6	5	5	4
9	1	2	2	5	5	5
6	1	1	4	5	5	5

*This scale is read as follows:

1 - Enormous impact, over 500 problems per 1000 stations
2 - Major impact, between 100 and 500 problems per 1000 stations
3 - Average impact, between 10 and 100 problems per 1000 stations
4 - Low impact, between 5 and 10 problems per 1000 stations
5 - Negligible impact, under 5 problems per 1000 stations

This classification is obviously subjective and could be modified. It aims to allow a **simple, quick, and homogenous** interpretation of the results from year to year.

** These results suggest a problem in the sampling data collection. In fact, results over 1000 indicate a major underestimation of discrepancy rates. It seems, however, that the introduction of source codes in 1989-90 has greatly improved this situation.

Note:

On the basis of this study, we can already see the trend evolving from year to year. Thus discrepancies[1] #2, 4, 5, 7, and 8 seem to affect the "health" of the spectrum, while discrepancies[1] #1, 3, 6, and 9 have little effect, especially the last two.

---

[2]See Appendix 3 on page A-6 for the meaning of the discrepancy codes.

This result does not reflect the same order of importance as the alert threshold table [section 5.1.2]. Thus, it would be desirable to use the results included here to set the alert thresholds. Such a correction has already been made for discrepancy #7.

5.2 Use of Recommendations

The purpose of this last section is to attempt to summarize the recommendations made in this document, in order to arrive at the production of useful and significant reports using the data from **sampling** and **investigation.** The results (compilations) for the investigations alone have already been discussed in Section 4.3

Given that many of the recommendations in this report have only been used for a few months, and that others have not yet been tried, it is not possible to carry out **statistical** analyses of the data. Instead, we will attempt to discuss the **methods** of analysis, without being able to use real data as examples. Thus, attention must be paid to the techniques that we will use instead of the numbers, as the latter will not mean anything.

5.2.1 Use of Surveys

The results of on-site surveys lend themselves particularly well to graphic presentation and this at all levels where the quantity of data is sufficient, depending upon the sampling plan carried out. Thus, the model is very simple, as we have already seen how to calculate the discrepancy percentages and the associated errors. Furthermore, alert thresholds were set in Section 5.1.2. The following pages show an example of the presentation of the results, based on the current SCORE format. If possible, it would be advantageous to improve the graphic format.

Note that the table in question shows discrepancies 1 to 15, as described in Appendix 3. Any other irregularities should be presented in the same way, but alert thresholds should be set. On the other hand, if multiple choice questions are introduced, the results should be presented separately, as described in Section 3.1.7.

It is not sufficient to do annual compilations, but a follow-up might turn out to be an essential technique for discovering a problem before it reaches unacceptable proportions. Thus, tables from previous years should always be consulted when a new report is produced. In terms of alert thresholds, those in Section 5.1.2 may be modified in several years, when more valid data are available.

LAND-FIXED SAMPLING

DISCREPANCY REPORT
BY DISTRICT

FINANCIAL YEAR:	1989/90
PERIOD:	APRIL-MARCH
DISTRICT NO.:	5672
CONFIDENCE LEVEL:	95%
POPULATION SIZE:	3473
EXPECTED SAMPLE:	480
N. INSPECTIONS PERFORMED:	277

NO. OF DISC.	NO. DISC. OBSERVED	NO. TESTS PERFORMED	PROBABILITY INTERVALS (%)	STATIONS INVOLVED
1	11	261	4 ± 2	146 ± 81
2	11	122	9 ± 5	313 ± 173
3	81	259	31 ± 5	1086 ± 188
4	14	263	5 ± 3	184 ± 90
5	65	260	25 ± 5	868 ± 175
6	28	256	11 ± 4	379 ± 127
7	14	200	7 ± 3	243 ± 119
8	6	210	3 ± 2	99 ± 75
9	82	239	34 ± 6	1191 ± 201
10	5	260	2 ± 2	66 ± 55
11	58	261	22 ± 5	771 ± 168
12	4	262	2 ± 1	53 ± 49
13	82	258	32 ± 5	1103 ± 189
14	1	112	1 ± 2	31 ± 59
15	5	251	2 ± 2	69 ± 57

Note:

Taking the source coding introduced in document [IPC 3.21.01], it would be interesting to add some columns to this graph indicating the percentage of observed discrepancies attributable to each potential source.

In addition, it would be preferable to add a decimal to the definition of the probability intervals (%), as several values are quite small.

04 MARCH 91

ON-SITE SAMPLING

DISCREPANCY REPORT
BY DISTRICT

FINANCIAL YEAR: 1989/90
PERIOD: APRIL-MARCH
DISTRICT NO.: 5672

CONFIDENCE LEVEL: 95%
POPULATION SIZE: 3473
EXPECTED SAMPLE: 480
N. INSPECTIONS PERFORMED: 277

DISCREPANCIES PROBABILITY INTERVALS

```
     0   10   20   30   40   50   60   70   80   90  100
    -+----+----+----+----+----+----+----+----+----+----+
1    | *]*
2    |  *]****
3    |     ]    ******
4    | **]*
5    |    ]    ******
6    |    *]**
7    |  *]**
8    |****  ]
9    |        ]  *******
10   |***                                     ]
11   |    ]    ******
12   |**   ]
13   |        ]  *******
14   |**       ]
15   |***     ]
    -+----+----+----+----+----+----+----+----+----+----+
```

***: Confidence intervals
] : Proposed alert thresholds

5.2.2 Impact Studies

One method of using sampling data AND investigation for analytical purposes would be to redo the study done in Section 5.1.4 with the 1984-90 data on the impact of discrepancies on the "health" of the spectrum. We will not repeat the detailed explanation of the method used, but Section 5.1.4 (page 52) describes the method and the results. Nevertheless, it would be interesting to take into account the new recommendations concerning the following points:

1) Use the new source codes for discrepancies, i.e., D, E, and F, and associate each discrepancy with a corresponding number from the sample surveys.

2) Use the new discrepancy codes (1 to 20), as described in the most recent document [IPC 3.21.01, 1991, Appendix 3].

3) Ensure that the <u>same</u> discrepancies are compared between investigation and sample surveys, for example comparing discrepancy #1 with source "D01*", where * represents the type of radiocommunication station, the list of which is available in document [IPC 3.1.4, 1986, Appendix C]. Thus, it would be important to use the same types of stations for the investigation as those inspected by the sample surveys, in order to compare data which refers to the same thing.

5.2.3 Notes on the Source Codes

If we look at the general public and radiocom source codes, we see that only codes:

1) D, E, and F (previously D)
2) X (immunity/blockage)

involve discrepancies related to telecommunications systems. As a result, all future statistical studies aiming to:

1 - Set alert thresholds (as in Section 5.1.2)
2 - Establish a mathematical relationship model (as in Section 5.1.3)
3 - Relate surveys and investigations in whatever manner (as, for example, in the impact study in Section 5.1.4.)

should use the source codes described above, as no correlation can exist between the other source codes.

5.2.4 Cost Analysis

One of the most useful data for the evaluation of the impact of a problem is knowing what it costs. In Section 3.1.6, we discussed a method for evaluating the number of illegal mobile units associated with a licensed base, which gave us an initial calculable "loss". Subsequently, Sections 4.3.3 and 4.3.4 showed us a method of calculating and presenting the costs caused by COMPLAINTS, as a function of source codes and other parameters.

Using the method suggested in Section 4.3.4 for the estimation of costs associated with a given source, we can then present the discrepancy percentages as a function of expenses caused by that discrepancy, in terms of survey time required as a result of a complaint. As data on **time** are not yet available for all investigation, it is not possible to present a real example here. Thus, the following table is only an example of the presentation of the results.

Example:

Discrepancy	Percentage	Source	Cost	Cost per 1%
1	(7.1 ± 1.2) %	D01*	$ 100 000	**$ 14 065**
2	(21.0 ± 3.2) %	D02*	$ 50 000	**$ 2 381**
etc ...	....	...	...	...

The "cost per 1%" column shows how much could be saved if the number of stations with this discrepancy dropped by 1%. This is obviously an **approximation**, which assumes a linear relationship between the number of stations with discrepancies and the number of complaints having that discrepancy as a source.

In practice, the relation is probably not linear, as can be seen on the graphs in Section 5.1.2. Nevertheless, taking into account the fact that such a curve is always approximate, because it varies from year to year, and discrepancy to discrepancy, it is preferable to use the AVERAGE costs by percent, which still gives a good idea of the situation.

APPENDICES

Appendix # 1 - References

[Arkin]: Hubert Arkin, Handbook of Sampling for Auditing and Accounting, McGraw Hill, New York, 1963.

[IPC 3.21.01]: Field Sampling, Internal Procedures Circular of the Department of Communications

[IPC 3.25.01]: Reception Sampling, Internal Procedures Circular of the Department of Communications.

[FS]: François Théberge, Functional Specifications - Ground Station Sampling, DC DOSP-C Section, February 1991

[Kish]: Leslie Kish, Survey Sampling, Wiley, New York, 1965.

[Radio Regulations]: International Telecommunication Union, Radio Regulations, volume 1, 1982.

[RIM 3.1.3]: Radiocommunication Investigation, Department of Communications, second edition, 1986, Broadcasting Regulations Branch

[RIM 3.1.4]: General Public Investigation, Department of Communications, second edition, 1986, Broadcasting Regulations Branch.

[Schaeffer]: Schaeffer, Mendelhall, Ott, Elementary Survey Sampling, Duxbury Press, 1979.

[Théberge]: François Théberge, Technical Report, 1st Edition, DC DOSP-C Section, August 1990

Appendix # 2 - Data Tables

Table 1. 1990/91 Sampling (sample sizes)

Universe/District	Population (see note page 63)	Plan C	Plan B	Plan A
Pacific				
112001	1389	187	111	11
112002	716	166	109	18
District 12	1696	-	220	29
113001	432	144	93	12
113002	1389	187	127	15
District 13	1621	-	220	27
District 14	Now with D.O. 16	176	176	25
District 15	1819	196	196	28
District 16	2855	197	197	31
District 17	932	182	182	31
118001	490	150	110	13
118002	460	147	110	15
District 18	869	-	220	28
District 25	787	161	161	21
TOTAL	10,579	912	1,572	220
Centre				
222001	1286	185	101	10
222002	7451	210	119	20
District 22	7983	-	220	30
223001	47,630	180	103	15
223002	4921	207	117	16
District 23	5409	-	220	31
District 24	2436	200	200	32
District 26	2203	176	176	24
232001	735	167	119	20
232002	3994	205	220	22
District 32	4966	-	92	42
233001	480	149	128	16
233002	3646	204	220	14
District 33	4251	-	87	30
234001	346	133	133	15
234002	4412	206	220	16
District 34	5009	-	-	31
TOTAL	32,257	376	995	220

Table 1. ... continued

Universe/District	Population (see note page 63)	Plan C	Plan B	Plan A
Ontario				
442001	2365	198	112	12
442002	1154	182	108	19
District 42	2795	-	220	31
443001	920	175	106	14
443002	920	175	114	16
District 43	1493	-	220	30
444001	698	165	104	13
444002	2230	197	116	21
District 44	2509	-	220	34
445001	450	146	95	12
445002	1722	192	125	18
District 45	1914	-	220	30
446001	501	151	96	17
446002	3994	205	124	22
District 46	4791	-	220	39
District 48	1172	182	182	26
449001	797	170	105	16
449002	1115	181	115	14
District 449	1912	-	220	30
TOTAL	16,586	182	1,502	220
Québec				
549001	320	129	117	19
549002	237	113	103	22
District 549	557	-	220	41
552001	2365	198	110	20
552002	1722	192	110	24
District 52	2925	-	220	44
553002	1998	195	57	15
553301	1239	184	55	7
553302	2230	198	67	12
553303	382	138	41	13
District 53	1906	-	220	47
554001	220	109	37	6
554002	1574	190	67	16
554402	536	154	59	12
554403	574	157	57	14
District 54	2256	-	220	48
555001	698	165	54	8
555002	2365	198	66	15
555901	423	143	59	11
555902	314	128	41	6
District 55	3074	-	220	40
TOTAL	10,718	0	1,100	220

Table 1...continued

Universe/District	Population *	Plan C	Plan B	Plan A
Atlantic				
663001	536	154	45	11
663002	2230	197	63	19
663301	561	156	52	12
663302	1239	184	60	15
District 63	2839	-	220	57
665001	166	94	67	18
665002	2230	197	153	40
District 65	2222	-	220	58
666001	291	124	58	20
666002	1446	188	99	30
666601	270	120	63	18
District 66	1721	-	220	68
District 69	381	145	145	37
TOTAL	7163	-	-	220
CANADA	777303	-	-	1100

The six digit codes in the left-hand column represent universes that are part of the next district on the list. A district that is not preceded by such a list of universes only contains one universe.

*Since exact data were not available at this level, the UNIVERSE population is an approximation obtained using the quantity of samples obtained (1990-91) and Arkin's formula, isolating the population "N" in the formula.

The population of districts (and also regions and the country) is that on 30 March 1991, and not necessarily equal to the sum of the universes for two reasons:

1 - There was a difference between the data
2 - The breakdown of universes was not necessarily mutually exclusive, thus a station might be included in the populations of two universes.

These are "commercial private" stations, LESS exclusions (National Defence, RCMP, municipalities).

Table 2 - Data available on SCORE for 1989/90

Region	Pacific	Centre	Ontario	Quebec	Canada
Population	1521	3530	367	7990	13408
Number of tests performed	538	492	1	755	1786
Number of discrepancy #1	16	11	1	39	67
Value of "p" (SCORE)	3 ± 1	2 ± 1	100	5 ± 2	7 ± 1

Table 3. Quebec Region, 89-90

Universe	Population	Test carried out	Discrepancies found
555201	1811	5	0
555202	384	15	4
555203	220	19	6
555204	190	20	1
555205	252	0	0
555206	309	82	28
555207	307	98	43
555301	953	45	38
555303	336	46	32
555401	670	18	15
555402	701	2	1
555403	274	7	4
555404	252	0	0
555501	1020	65	11
555506	39	0	0
555507	38	0	0
555902	119	64	31
555905	115	22	8
TOTAL	7,990	508	222

Table 4. Montreal District 89-90

UNIVERSE	POPULATION	TEST CARRIED OUT	DISCREPANCIES FOUND
555201	1811	6	3
555201	384	16	3
555203	220	20	6
555204	190	27	3
555205	252	0	0
555206	309	86	15
555207	307	101	35
TOTAL	3,473	256	65

Appendix # 3 - Discrepancy Codes

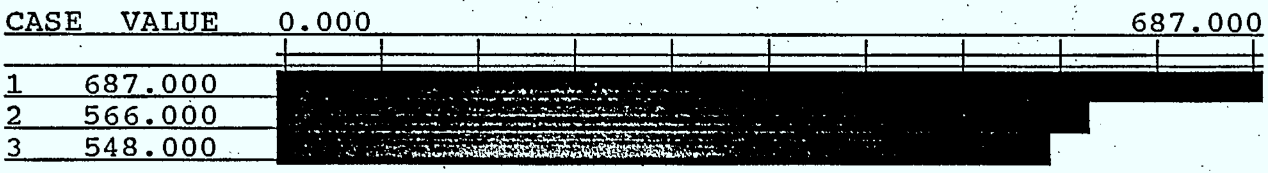
These are the 1991 codes. Consult the most recent [IPC 3.21.01] for the modifications.

1. Off-frequency operation.
2. Over/under modulation or deviation, excessive bandwidth.
3. Power or ERP in excess of that authorized.
4. Unauthorized operation with respect to frequency.
5. Unauthorized operation with respect to location as specified on license.
6. Antenna characteristics or radiation pattern other than as authorized.
7. Excessive spurious or harmonic radiation.
8. Incorrect operating procedures; superfluous, unauthorized communications
9. Improper or non-identification
10. Unsafe installation
11. Associated unlicensed station.
12. Antenna structure not in accordance with approved height, lighting, painting, or other marking requirements
13. Station installation does not comply with regulations with respect to documents, logs, equipment maintenance, or spare parts
14. Operating personnel inadequately certified
15. Non-approved equipment

Appendix # 4 - Complaints by Source

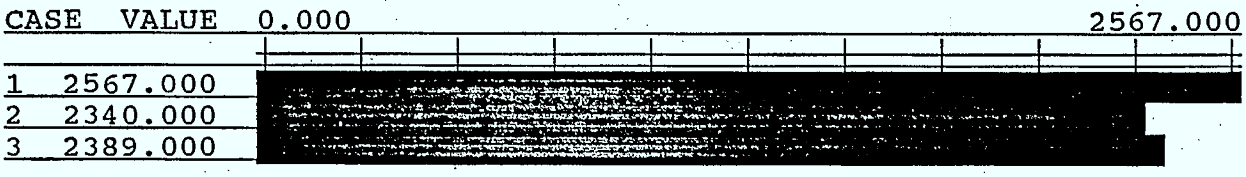
PLOT OF C

NUMBER OF CASES = 3
MEAN OF SERIES = 600.333
STANDARD DEVIATION OF SERIES = 61.722
SEQUENCE PLOT OF SERIES :



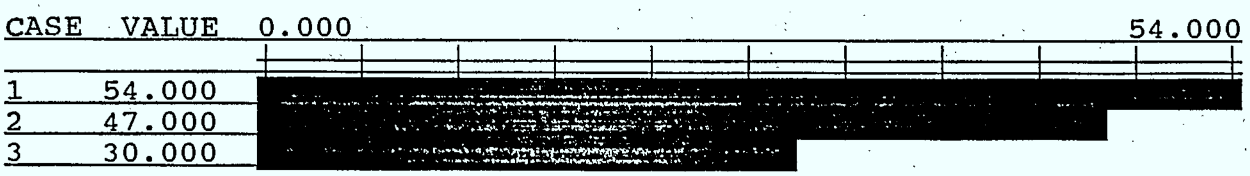
PLOT OF D

NUMBER OF CASES = 3
MEAN OF SERIES = 2432.000
STANDARD DEVIATION OF SERIES = 97.533
SEQUENCE PLOT OF SERIES :



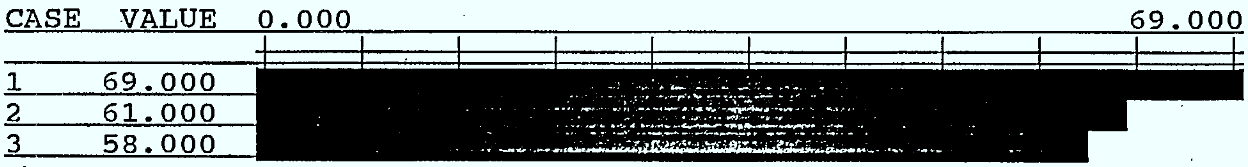
PLOT OF I

NUMBER OF CASES = 3
MEAN OF SERIES = 43.667
STANDARD DEVIATION OF SERIES = 10.077
SEQUENCE PLOT OF SERIES :

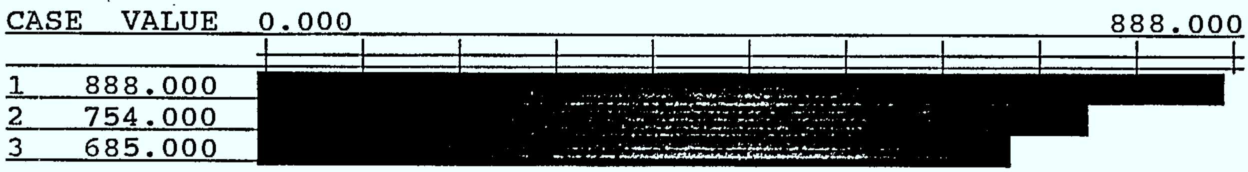


PLOT OF L
NUMBER OF CASES = 3
MEAN OF SERIES = 2019.333
STANDARD DEVIATION OF SERIES = 560.381
SEQUENCE PLOT OF SERIES :

PLOT OF R
NUMBER OF CASES = 3
MEAN OF SERIES = 62.667
STANDARD DEVIATION OF SERIES = 4.643
SEQUENCE PLOT OF SERIES :



PLOT OF S
NUMBER OF CASES = 3
MEAN OF SERIES = 775.667
STANDARD DEVIATION OF SERIES = 84.279
SEQUENCE PLOT OF SERIES :



PLOT OF T

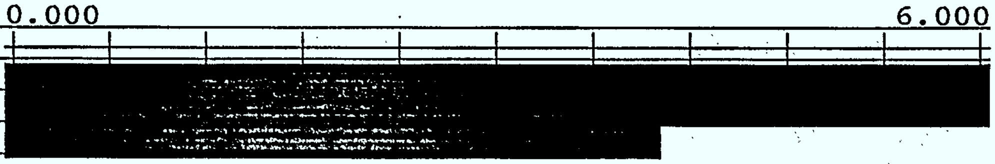
NUMBER OF CASES = 3

MEAN OF SERIES = 5.333

STANDARD DEVIATION OF SERIES = 0.943

SEQUENCE PLOT OF SERIES :

CASE	VALUE
1	6.000
2	6.000
3	4.000



PLOT OF W

NUMBER OF CASES = 3

MEAN OF SERIES = 773.667

STANDARD DEVIATION OF SERIES = 74.415

SEQUENCE PLOT OF SERIES :

CASE	VALUE
1	874.000
2	751.000
3	696.000

PLOT OF X

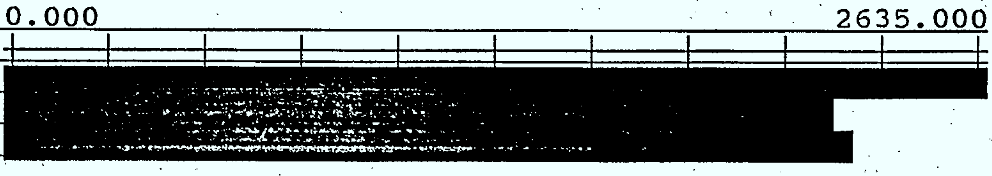
NUMBER OF CASES = 3

MEAN OF SERIES = 2392.333

STANDARD DEVIATION OF SERIES = 173.694

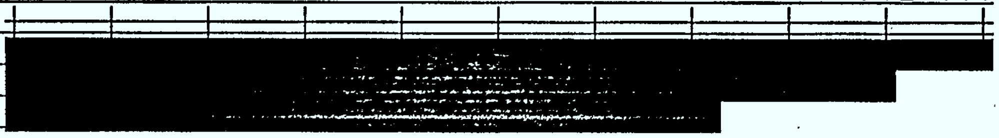
SEQUENCE PLOT OF SERIES :

CASE	VALUE
1	2635.000
2	2238.000
3	2304.000



PLOT OF Y
NUMBER OF CASES = 3
MEAN OF SERIES = 846.667
STANDARD DEVIATION OF SERIES = 107.351
SEQUENCE PLOT OF SERIES :

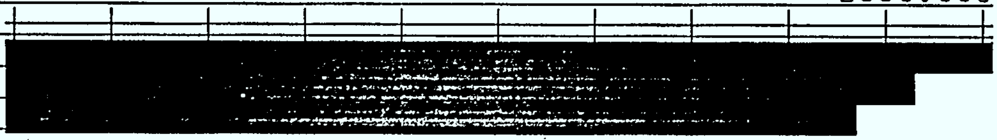
CASE VALUE 0.000 961.000



CASE	VALUE
1	961.000
2	876.000
3	703.000

PLOT OF Z
NUMBER OF CASES = 3
MEAN OF SERIES = 2686.667
STANDARD DEVIATION OF SERIES = 154.439
SEQUENCE PLOT OF SERIES :

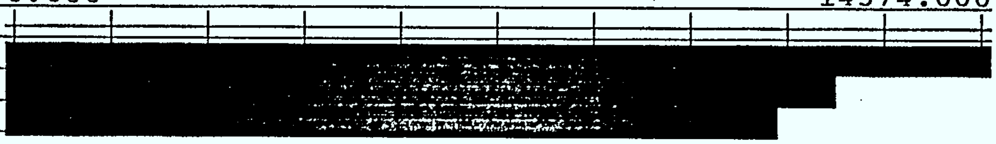
CASE VALUE 0.000 2880.000



CASE	VALUE
1	2880.000
2	2678.000
3	2502.000

PLOT OF TOTAL
NUMBER OF CASES = 3
MEAN OF SERIES = 12638.333
STANDARD DEVIATION OF SERIES = 1283.128
SEQUENCE PLOT OF SERIES :

CASE VALUE 0.000 14374.000



CASE	VALUE
1	14374.000
2	12229.000
3	11312.000

Appendix # 5 - Example - Mystat Program

The following are the detailed technical data obtained using the "Mystat" program for the model retained for the estimation of the number of source D radiocom complaints, noted as "D17". These data are included here simply for reference purposes, to provide an example of how to use the "Mystat" program.

PEARSON CORRELATION MATRIX

	D17	DISC1_7	DISC1_9	DISC1_15
D17	1.00000			
DISC1_7	**0.53708**	1.00000		
DISC1_9	0.44391	0.91644	1.00000	
DISC1_15	0.46438	0.86868	0.94889	1.00000

* This matrix shows us that source D17 investigation seem to have a high correlation with the sum of discrepancies 1 to 7. This datum will be used to produce our model, whose statistical details are as follows:

DEP VAR: D17 N:18 MULTIPLE R: .580
SQUARED MULTIPLE R: .336 ADJUSTED SQUARED MULTIPLE R: .294
STANDARD ERROR OF ESTIMATE: 9.16623

VARIABLE	COEFFICIENT	STD ERROR	P(2 TAIL)
CONSTANT	**16.37820**	**4.20349**	0.00128
DISC1_7*			
DISC1_7	**0.00646**	**0.00227**	0.01171

ANALYSIS OF VARIANCE

SOURCE	SUM-OF-SQUARES	DF	MEAN-SQUARE	F-RATION	P
REGRESSION	679.83001	1	679.83001	8.09130	**0.01171**
RESIDUAL	1344.31733	16	84.01983		

* Thus, the model "$0.00646(DISC\_7)^2 + 16.3782$" is obtained as a prediction of the relative volume of complaints.

Appendix # 6 - Utility Programs

(A) User's Guide

The "STAT" program performs most of the statistical calculations described in this report, in addition to calculations related to license conditions activity. The program is written in GWBASIC, and is available in French (STAT.VF) and English (STAT.VA) versions.

To use these programs, start GWBASIC and load the program with the command LOAD"STAT.VA for the English version.

Subsequently, menus will appear, and an option on the main menu provides a short description of the available options. The six available options are as follows:

(1) Plans 0, A, and B, surveys

Calculates the confidence interval for the percentage of discrepancies for an anomaly. The data used may be national (Plan 0), regional (Plans 0 and A), or by district (Plans 0, A, and B). Allows the output of the results on disk C or on screen.

(2) Quality Control, by District

Allows the data from a district to be compared with the national average. See Section 3.1.4 of this report for details.

(3) Illegal Mobile Units - Samples

Calculates the number of illegal mobile units (confidence interval) for a given territory based on sample results. The calculation may be done by district, by region if one stops at Plan A, or nationally if only Plan 0 has been completed.

(4) Compliance with License Conditions

This option allows the number of illegal stations to be determined, based on license condition activity results for a region or nationally.

(5) Generation of Random Numbers

Allows x random numbers between 1 and y to be drawn. The user has the choice of x and y. Output is to drive C (modifiable).

(6) Multiple Choice, Surveys

Allows a confidence interval to be obtained for a given response in a multiple choice framework. This calculation may be done at the national, regional (Plans A and 0), or district (Plans 0, A, and B) levels.

Finally, the following is the list of output files used, as well as the options that use them, and the lines of the program where they can be changed.

Option	File	Lines
(1)	C:PLANS.DAT	1830 and 2090
(4)	C:RESPECT.DAT	5870 and 6190
(5)	C:HASARD.DAT	7015 and 7020
(6)	C:MULTIPLE.DAT	8880 and 9170

(B) Program Listing

```
10      Program for the statistical calculation of spectrum activities. This program supports simple and stratified (weighted)
        calculations. Type "run" to start.
36      There may be some code and comment repetitions, since this program is a combination of several routines.
50      By: François Théberge, DOSP-C, March 1991
140     PRINT "*                   Department of Communications        *"
150     PRINT "*              Spectrum Control          *"
170     PRINT "*        STAT - Statistical Calculations   *"
180     PRINT "*                Program                                 *"
190     PRINT "*            (1) Plans 0, A, and B, surveys          *"
200     PRINT "*  (2) Quality Control by District        *"
210     PRINT "*  (3) Illegal Mobile Units, surveys      *"
220     PRINT "*            (4) Compliance With License Conditions  *"
225     PRINT "*  (5) Random Number Generation           *"
228     PRINT "*  (6) Multiple Choice Surveys            *"
230     PRINT "*  (9) Help                          *"
240     PRINT "*  (0) End of Session                 *"
1010    Program to calculate confidence intervals on the basis of survey data. This program supports simple and stratified
        (weighted) calculation. Type "run" to start.
1050    By: François Théberge
1100    Basin data input
1120    A$="(1) Plan 0 (zero) only - no stratification"
1130    B$="(2) Plans 0 and A - stratification by region"
1140    C$="(3) Plans 0, A, and B - stratification by district"
1143    PRINT "** Land Fixed Station Surveys **"
1180    PRINT "(4) Return to Main Menu"
1220    INPUT "Discrepancy Considered (Number) :",Y
1230    PRINT "Confidence Threshold (enter 90, 95 or 99; default=95)";
1310    Plan 0 - no stratification
1330    INPUT "Total Population ", POP
1340    INPUT "Discrepancies Observed ",ANO
1350    INPUT "Verifications Carried Out ",TEST
1450    Surveys with stratification
1470    PRINT "number of ";STRATA$
1520    PRINT "population, ";STRATA$;" # ";NO;
1540    PRINT "discrepancies observed, ";STRATA$;" # ";NO;
1560    PRINT "verifications carried out (at least 2), ";STRATA$;" # ";NO;
1620    *** WEIGHTED CALCULATIONS***
1800    Menu + screen output
1820    PRINT "(1) Output to Screen (default)"
1830    PRINT "(2) Add to file C:PLANS.DAT:
1863    PRINT "Survey Results"
1910    PRINT "Discrepancy Considered   :",Y
1920    PRINT "Confidence Threshold (%)   :",Threshold
1920    PRINT "Confidence Intervals:   ";
1990    PRINT "Probable number of stations involved :  ";
2030    INPUT "Press <ENTER> to continue",XYZ
2070    File output
2103    PRINT#1,"Survey Results"
2150    PRINT "Discrepancy Considered   :",Y
2160    PRINT "Confidence Threshold (%)   :",Threshold
2170    PRINT "Confidence Intervals:   ";
2230    PRINT "Probable number of stations involved :  ";
2280    INPUT "Press <ENTER> to continue",XYZ
2320    Error message
2350    PRINT " *** Data Error !!! *** "
```

```
3001  Quality control
3020  PRINT "** Quality Control by District **"
3040  INPUT "Discrepancy Considered : ",ANO
3050  INPUT "District Considered : ",DISTRICT$
3060  PRINT "NATIONAL verification rate (%) for the discrepancy:",ANO$
3090  PRINT "NATIONAL discrepancy percentage, discrepancy:",ANO$
3100  INPUT "Percentage :",A
3110  INPUT "Associated Error, in % :",E
3130  PRINT "DISTRICT data"
3150  PRINT "Number of inspections carried out"
3160  PRINT "Number of verifications, discrepancy",ANO$
3190  PRINT "Number of discrepancies",ANO$
3230  *** Calculations
3300  PRINT "*** Quality Control ***"
3320  PRINT "District :";DISTRICT$
3330  PRINT "Discrepancy :";ANO$
3360  PRINT "% of verifications carried out  :";
3400  PRINT "(i) - % ratio of verifications, district/national: ";
3430  PRINT "% of discrepancies for this district  :";
3460  PRINT "(ii) Comparison with national rate: ";
3483  PRINT "See the DOSP-C technical report"
3485  PRINT "for the meaning of (i) and (ii)"
3490  INPUT "Press <ENTER> to continue",XYZ

3500  calculation of (i)
4001  Illegal mobile units
4020  PRINT "** Illegal Mobile Units **"
4033  INPUT "Territory considered (eg: district) : ",TER$
4040  PRINT "Total mobile station population:;
4042  PRINT "with license, in this service : "
4090  PRINT "Survey results : ";
4110  PRINT "Number of LICENSED mobile stations";
4112  PRINT "observed during the survey"
4140  PRINT "TOTAL number of mobile stations";
4142  PRINT "observed during the survey"
4221  PRINT "** Illegal mobile units **"
4225  PRINT "Territory considered : ",TER$
4240  PRINT "Confidence interval (95%), QUANTITY of illegal mobile units : "
4280  PRINT "Press <ENTER> to continue",XYZ
4310  PRINT "SURPLUS...No gain to be had in this case!"
4330  PRINT "Press <ENTER> to continue",XYZ
5010  Program to calculate the confidence interval, compliance with the conditions of the license. This program supports simple
      and stratified (weighted) calculations
5050  By: François Théberge, DOSP-C, February 1991
5100  Basic data input
5120  A$="(1) National Report"
5130  B$="(2) Regional Report"
5160  PRINT "(3) Return to Main Menu"
5200  PRINT "Type of Station (ex: Mobile : ";
5220  PRINT "Cost of a license (in $)  : ";
5240  PRINT "Confidence Threshold (enter 90, 95 or 99; default=95)";

5310  Regional
5332  PRINT "DATA CONCERNING STATIONS : ";TYPE$
5340  PRINT "Total station population";
5342  PRINT "with licenses (in that service),"
5345  INPUT "for the region considered : ",N
```

```
5350    PRINT "Number of LICENSED stations
5552    PRINT "observed during the survey"
5360    National
5480    Number of regions
5554    PRINT "Region # ",NO;
5570    PRINT "TOTAL number of stations";
5572    PRINT "observed during the survey" ";
5574    INPUT "Region # ",NO;
5630    Weighted calculations
5840    Menu + screen output
5860    PRINT "(1) Output to Screen (default)"
5870    PRINT "(2) Add to file C:RESPECT.DAT"
5925    PRINT " - compliance with license conditions"
5940    PRINT "Type of Station  :",TYPE$
5950    PRINT "Confidence Threshold (%)   :",SEUIL
5960    PRINT "Confidence Intervals:   ";
5965    PRINT " - compliance with license conditions : ";
6030    PRINT "PROBABLE NUMBER OF ILLEGAL STATIONS"
6040    PRINT "Lower Limit : ",Set(N/(P+E)-N)
6050    PRINT "Most probable value : ",Set (N/P-N)
6060    PRINT "Upper Limit : ",Set(N/(P-E)-N)
6080    PRINT "Losses"
6090    PRINT "Lower Limit : ",Set(N/(P+E)-N)*C,"$"
6100    PRINT "Most probable value : ",Set (N/P-N)*C,"$"
6110    PRINT "Upper Limit : ",Set(N/(P-E)-N)*C,"$"
6130    INPUT "Press <ENTER> to continue",XYZ
6225    PRINT#1 " - compliance with license conditions"
6240    PRINT#1 "Type of Station  :",TYPE$
6250    PRINT#1 "Confidence Threshold (%)   :",SEUIL
6260    PRINT#1 "Confidence Intervals:   ";
6265    PRINT#1 " - compliance with license conditions : ";
6320    PRINT#1 "PROBABLE NUMBER OF ILLEGAL STATIONS"
6330    PRINT#1 "Lower Limit : ",Set(N/(P+E)-N)
6340    PRINT#1 "Most probable value : ",Set (N/P-N)
6350    PRINT#1 "Upper Limit : ",Set(N/(P-E)-N)
6370    PRINT#1 "Losses"
6380    PRINT#1 "Lower Limit : ",Set(N/(P+E)-N)*C,"$"
6390    PRINT#1 "Most probable value : ",Set (N/P-N)*C,"$"
6400    PRINT#1 "Upper Limit : ",Set(N/(P-E)-N)*C,"$"
6420    INPUT "Press <ENTER> to continue",XYZ
6490    PRINT " ** Data Error !!! ** "
7015    PRINT "The list will be written to the file C:HASARD"
7030    INPUT "Upper limit for the choice of numbers ",X
7040    INPUT "size of the list ",Y
7100    INPUT "Press <ENTER> to continue",XYZ
8120    A$="(1) Plan 0 (zero) only - no stratification"
8130    B$="(2) Plans 0 and A - stratification by region"
8140    C$="(3) Plans 0, A, and B - stratification by district"
8150    PRINT "** Land Fixed Station Surveys  - Multiple Choice **"
8200    PRINT "(4) Return to Main Menu"
8250    INPUT "Question Number :",Y
8255    INPUT "Answer to analyze (ex: F) : ",REP$
8260    PRINT "Confidence Threshold (enter 90, 95 or 99; default=95)";
8360    INPUT "Total Population ", POP
8365    PRINT "Number of answers ";REP$;" obtained ";
8380    INPUT "TOTAL number of responses with an opinion (eg: F) ",TEST
8383    INPUT "Number of responses with no opinion (K) ",SO
```

```
8386    INPUT "Number of non-applicable responses (L) ",NA
8510    PRINT "number of ";STRATE$;
8560    INPUT "Population, ";STRATE$;" # ";NO
8580    PRINT "Number of responses ";REP$;" obtained ";STRATE$;" # ";NO;
8380    INPUT "TOTAL number of responses with an opinion (eg: F) ";STRATE$;" # ";NO;
8383    INPUT "Number of responses with no opinion (K) ",STRATE$;" # ";NO;
8386    INPUT "Number of non-applicable responses (L) ",STRATE$;" # ";NO;
8870    PRINT "(1) Output to Screen (default)"
8880    PRINT "(2) Add to file C:MULTIPLE.DAT"
8920    PRINT "Survey results - multiple choice"
8980    PRINT "Question considered   : ",Y
8985    PRINT "Answer considered      : ",REP$
8986    PRINT "Number of responses with an opinion :  ",REP$
8383    PRINT "Number of responses with no opinion: ",SO
8386    PRINT "Number of non-applicable responses : ",NA
8990    PRINT "Confidence Threshold (%)   :",SEUIL
9000    PRINT "Confidence Intervals:   ";
9060    PRINT "Probable number with this opinion : ";
9110    INPUT "Press <ENTER> to continue",XYZ
9250    PRINT#1 "Discrepancy considered";Y
9252    PRINT#1 "Answer considered      : ",REP$
9254    PRINT#1 "Number of responses with an opinion :  ",REP$
9256    PRINT#1 "Number of responses with no opinion: ",SO
9258    PRINT#1 "Number of non-applicable responses : ",NA
9260    PRINT "Confidence Threshold (%)  :",SEUIL
9270    PRINT "Confidence Intervals:   ";
9330    PRINT "Probable number with this opinion : ";
9390    INPUT "Press <ENTER> to continue",XYZ
9460    PRINT " ** Data Error !!! ** "
10040   PRINT "(1) Plans 0, A, and B, surveys"
10060   PRINT "Calculates the confidence interval,"
10070   PRINT "for the % of discrepancies, for a particular discrepancy"
10080   PRINT "The data entry may be national (plan 0), regional"
10090   PRINT "(plans 0 and A) or by district (plans 0, A, and B)."
10100   PRINT "(2) Quality Control, by district"
10130   PRINT "Allows the data for a district to be compared with national data"
10140   PRINT "Please refer to Section 3.1.4 of the DOSP-C"
10150   PRINT "technical manual (April 1991) for details"
10170   PRINT "(3) Illegal mobile units - surveys"
10190   PRINT "Calculates the number of illegal mobile units (confidence interval)"
10200   PRINT "for a given district, based on survey results.  The"
10210   PRINT "calculation may also be done for a region if plan A"
10220   PRINT "is used, or for the country if only Plan"
10230   PRINT "0 is used."
10250   INPUT "Press <ENTER> to continue",XYZ
10270   PRINT "(4) Compliance with License Conditions"
10290   PRINT "This option allows the number of illegal stations"
10300   PRINT "and the losses to be determined, on the basis of"
10310   PRINT "the activity respecting compliance with the licence conditions"
10320   PRINT "for a region or on the national level"
10340   PRINT "(5) Random number generation"
10360   PRINT "Allows x random numbers between 1 and y to be drawn."
10370   PRINT "The user may choose x and y.  Output is to Disk"
10380   PRINT "C (modifiable)"
10391   PRINT "(6) Multiple Choice Surveys"
10393   PRINT "Allows a confidence interval for a particular response"
10394   PRINT "to be calculated in a chosen multiple choice"
```

```
10395   PRINT "questionnaire, for the country, by region, or by district"
10400   INPUT "Press <ENTER> to continue",XYZ
10420   PRINT "The following is a list of the output files used, and the"
10420   PRINT "options that use them, and the lines of the program"
10430   PRINT "where they can be modified"
10460   PRINT "Option      File                  Lines"
10470   PRINT "----------------------------------------"
10480   PRINT "(1)         C:PLANS.DAT     1830 and 2090"
10490   PRINT "(4)         C:RESPECT.DAT   5870 and 6190"
10500   PRINT "(5)         C:HASARD.DAT    7015 and 7020"
10505   PRINT "(6)         C:MULTIPLE.DAT 8880 and 9170"
10520   INPUT "Press <ENTER> to continue",XYZ
```

Appendix # 7 - List of Symptoms

The following is an "initial" list of possible symptoms for general public and radiocom surveys. Refer to Section 4.2.2 for details.

General Public Surveys

1. Unidentified Voice or Noise

a) On TV (on or off)
b) Radio (AM or FM)
c) Non-radio equipment (ex: Organ, Stereo, VCR, Phone, Answering Machine, Console)
d) Persistent or intermittent background noise ("Rumble", Static)
e) Rumbling noises

2. TV Image

a) Poor general quality (ex: snow)
b) Bright dots on the screen
c) Phantom images
d) Loss of reception of a station
e) Loss of image
f) Two (or more) images shown simultaneously
g) Bars on the screen (fixed/mobile;vertical, horizontal or diagonal)
h) Reception of cable stations with an antenna
i) Reception of U.S. stations

3. Radio

a) Loss of signal (one or more stations)
b) Intermittent noises
c) More than one station tuned in at the same time
d) Loss of reception at night
e) Reception of U.S. stations

4. Electronic System Affected by a Radio Signal

a) Alarm system
b) TV on and off
c) Garage door
d) etc.

5. Others

Radiocommunication Surveys

1. **Frequency Sharing Problems**

    a) User not complying with frequency "sharing"
    b) Obscene language
    c) New station on the frequency, legal or not
    d) Unidentified voices
    e) Overlap of several conversations
    f) Incorrect use of frequency by operators
    g) Abnormal noise on the frequency
    h) Sharing the same tone

2. **Reception Problems**

    a) Weak or low-quality signal
    b) Reception of U.S. stations
    c) Static (continuous or intermittent)
    d) Paging or telephone signal on the frequency
    e) Distorted reception
    f) Only one side of a conversation heard

3. **Transmission Problems**

    a) Difficulty sending messages (overloaded frequency)
    b) Reduced broadcast area
    c) Continuous wave carrier
    d) Excessive power

4. **Defective Equipment or Improper Use of That Equipment**

5. **Others**[3]

---

[3]Intermodulation problems are not included on this list, as they are physical explanations of certain effects.

INDEX

DATE DUE
DATE DE RETOUR

CARR McLEAN 38-296