METHOD AND INTERPRETATION OF SURVEYS ON NOISE ANNOYANCE

Frank Krammer
Noise Pollution Control Section
Environment Ontario

ABSTRACT

Major practical considerations during the design and analysis of sociological noise surveys are briefly discussed, particularly as they relate to choice and interpretation of statistical tests. The discussion concentrates on the appropriate use of data and scale transformations which may not only assist in the interpretation of results but also clarify seeming discrepancies both within a study as well as between apparently conflicting results reported in the literature. Applicable scale transformations are supported by the specification and discussion of theoretically based quantitative functions which may be used to predict human response from noise-level measures of loudness.

The interest of the Noise Pollution Control Section of Environment Ontario in the results of sociological noise surveys is primarily to validate the use of a noise descriptor as an indicator of individual reaction (relative to the Model By-Law) or as a predictor of community impact (relative to land use guidelines and approval criteria). In addition, the choice and effectiveness of noise control measures may be better evaluated with an understanding of differences in sensitivity of people resulting from observable differences in demographic situations. Social surveys are used to answer such questions.

SURVEY PROCEDURES

The statistical analysis and interpretation of a survey are influenced by each step taken during the survey:

1. PROBLEM DEFINITION

The type of analysis will depend on the hypothesized relationships between noise impact or disturbance and noise or other predictors, and the level of qualitative or quantitative information contained in the obtained data.

2. QUESTIONNAIRE DESIGN AND CODING

Questions are carefully worded and placed in the appropriate context to elicit responses that are meaningfully related to the underlying disturbance or impact, so that statistical analyses will have the desired meaning. Questions are therefore very specific to minimize misinterpretation and simplified to assure they are within the capability of the individual to answer.

For example, in a pilot study, people who were asked: Would you say this neighbourhood is quiet, noisy or neither?" sometimes rated their neighbourhood on a busy-quiet, or active-boring scale. The revised question: "Is it generally quiet or noisy in this neighbourhood, or is it neither?" avoided some of the confusion but was often a reportedly difficult decision to make. Many answered that it was sometimes noisy and sometimes quiet and that integration over different situations was difficult.

Allowable answers are coded to represent the underlying scale at a level of measurement as near as possible to that assumed by the statistical test. The fundamental assumption of statistics is that events may be assigned numbers which are values of a "random variable" which in turn is assumed to have specific statistical properties. The number codes are the values of the random variable assumed for the test.

For example, the coding for the previous question on neighbourhood noise would be:

1. quiet
2. neither
3. noisy

reflecting an increased degree of perceived noisiness on a rank-ordered scale. In other words, the value of the code increases with an increase in the factor measured. The number of points chosen for the scale will depend on the respondent's ability to make the judgement, the level of subsequent analysis and the anticipated range of responses. In general, a somewhat finer scale than that required will be chosen since the range of responses is hot known a priori. A detailed scale can later be transformed to a cruder scale but a crude scale cannot be changed to a detailed one, e.g. 1-4 agreeable
5 neutral
6, 7 somewhat disturbing
8, 9 highly disturbing (Hemingway and Krammer, 1977).
Having assured that the question is meaningful and that the codes represent an underlying scale, we would also like to assure that different people use a similar criterion for making the judgement. For example,the

respondent may be asked to make the judgement of noisiness with respect to other neighbourhoods in the same city or town. Alternatively, one of the scale points (e.g. ratings) may be more precisely defined, so that other judgements can be made in relation to it.

3. SAMPLING OR CHOICE OF SITES

The choice of sites and of respondents within these sites will determine the generality of the statistical results, as well as the amount of unwanted variability in the data due to the presence of extraneous effects. As a rule of thumb, there must be differences in factors that will subsequently be examined in the statistical analysis. The range of these should be sufficient to include the range over which the hypothesized relationship is to be confirmed. Factors that are not relevant to the problem under investigation, on the other hand, are either kept constant as much as possible or are included as a random component in the statistical model by the choice of a larger (also more costly) sample.

Different statistical analyses assume different methods of sampling. For example, correlation analysis assumes that both variables or factors are random variables, or in other words that the investigator exerts no control over the value that these measures assume. For regression analysis alternatively, values of the independent variable or the predictor of human reaction, is assumed to be predetermined. Minor violations of these assumptions are routine and commonplace, but major violations can be serious.

4. COST EFFECTIVENESS

Sampling is usually the greatest factor influencing cost. Travel costs may be reduced by restricting the physical area to be surveyed. Also, the number of questions asked may be reduced to only those which will relate to the problems of interest.

5. INTERVIEWER TRAINING

Interviewers are trained to administer the questionnaire consistently, interpreting questions as intended.

6. PILOT STUDY

A pilot study points out problems in administration and permits final clarification, or final necessary ...

7. REVISIONS OF QUESTIONNAIRE

8. DATA COLLECTION

Noise data are usually collected after the survey to avoid respondent biases that may result from previous knowledge of the study.

9. DATA CODING

Data is coded on the questionnaire or the information transferred onto a coding sheet which permits easier keypunching. Care is taken that codes define a variable meaningfully as mentioned earlier. (see Table I)

10. KEYPUNCHING

The codes are transferred to a specific set of card columns for each variable. This set is called a "field". Each questionnaire makes up a "record", one for each respondent. Each record is constructed the same way and will have only different codes for the value of each field.

RECORDS (CARDS)	RESPONDENT	SITE		NOISE RATING	VOLUNTEERED	L_{EQ}		S.D.	
	1	1	1	2	2	6	1	4	2
	2	1	1	3	2	6	1	4	2
	3	1	1	1	1	6	1	4	2
	4	1	1	3	1	6	1	4	2
	5	1	1	1	1	6	1	4	2
	6	1	1	1	2	6	1	4	2
	7	1	1	5	1	6	1	4	2
					FIELDS				

Table I. Sample of a coding sheet, showing the format in which information is handled by the computer.

11. COMPUTER FILE STRUCTURE

The cards are read into computer storage records, each record being a sample. The use of the SPSS package program will be assumed here. The names of the variables are written in the order in which the fields appear on the record, on a "variable list" card. A "format" card specifies the number of columns that are to be read for each variable. (see Table I)

12. DATA TRANSFORMATION

(a) Definition

This is probably the most useful method used by statisticians. The variables defined in the questionnaire are by no means the only variables that can be analysed from the data. For example, a new variable could be truncated to provide fewer scale points. Or, a variable could be defined as a combination of the variables obtained directly from the coded answers as in a count of the number of ways disturbed, the occurrence of any noise disturbance, and the combination of disturbance reports (as obtained from factor analysis, for example) that would most closely approximate the underlying human reaction that is of interest. Also, a functional transformation can be made. For example, a logarithmic transformation on the dependent variable will make an exponential function linear with the transformed variable or a squared transformation will make a parabolic function linear, allowing standard statistical tests to be used.

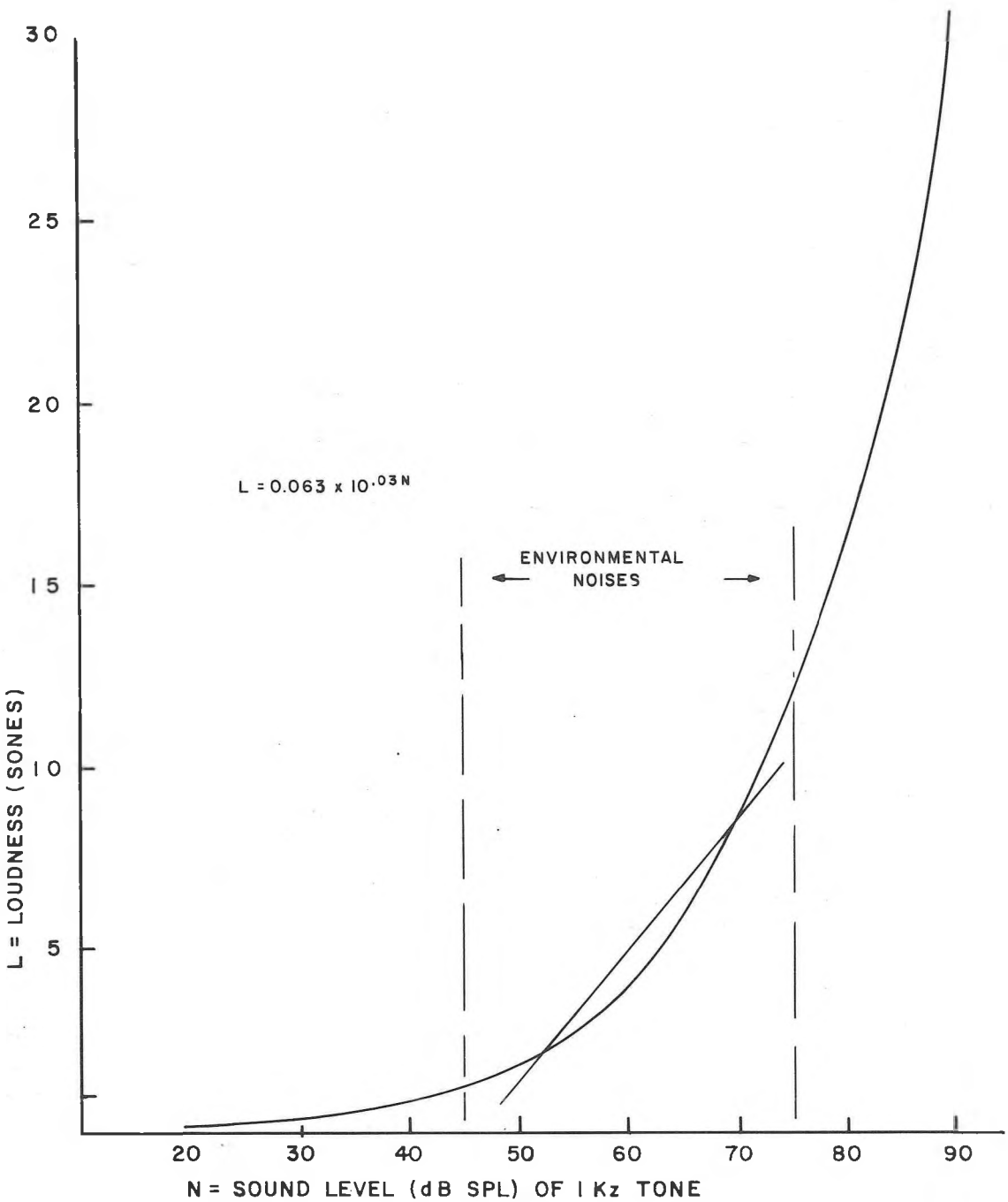

Fig. 1. Antilogarithmic form of the power function of loudness with sound intensity. The graph shows that, within the range of environmental noise levels, a linear approximation may be adequate.

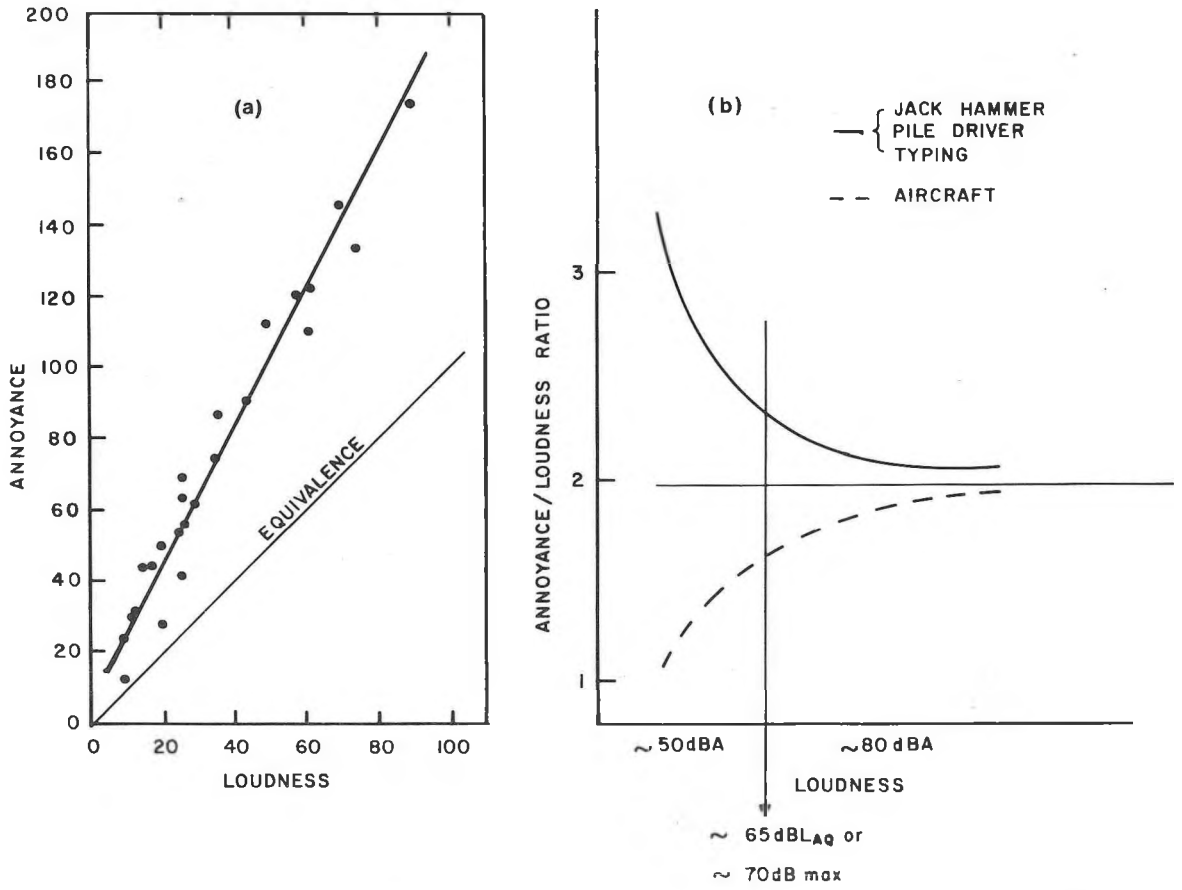

Fig. 2. Somewhat schematized relationship between annoyance and loudness (from Berglund et al ,1976). The perceived annoyance is always greater than the perceived magnitude of the loudness.

(b) Application - Theoretical Nonlinearities

An important underlying relationship is that between noise level and annoyance. The relationship between sound intensity and loudness is accepted as being a power function and therefore roughly an exponential function with the logarithmic sound level scale (see (Fig. 1). From work by Berglund et al (1976) annoyance appears to be linearly related to loudness (see Fig. 2 (a)). Therefore, annoyance may also be basically an exponential function of sound level.

The underlying relationship between sound level and annoyance should then be roughly exponential. In the Berglund et al (1976) study, the sources examined were about twice as annoying as they were loud at levels over about 70 dBA. At lower noise levels, the levels of annoyance relative to loudness appeared to depend on the source of the noise (see Fig. 2 (b)). Preliminary results by Hall, Breston and Taylor (1977) also suggest a possible underlying relationship of reduced house prices with increased noise levels, significant differences in house prices being observed only at sites with noise levels greater than about 70 dBA L_{eq}.

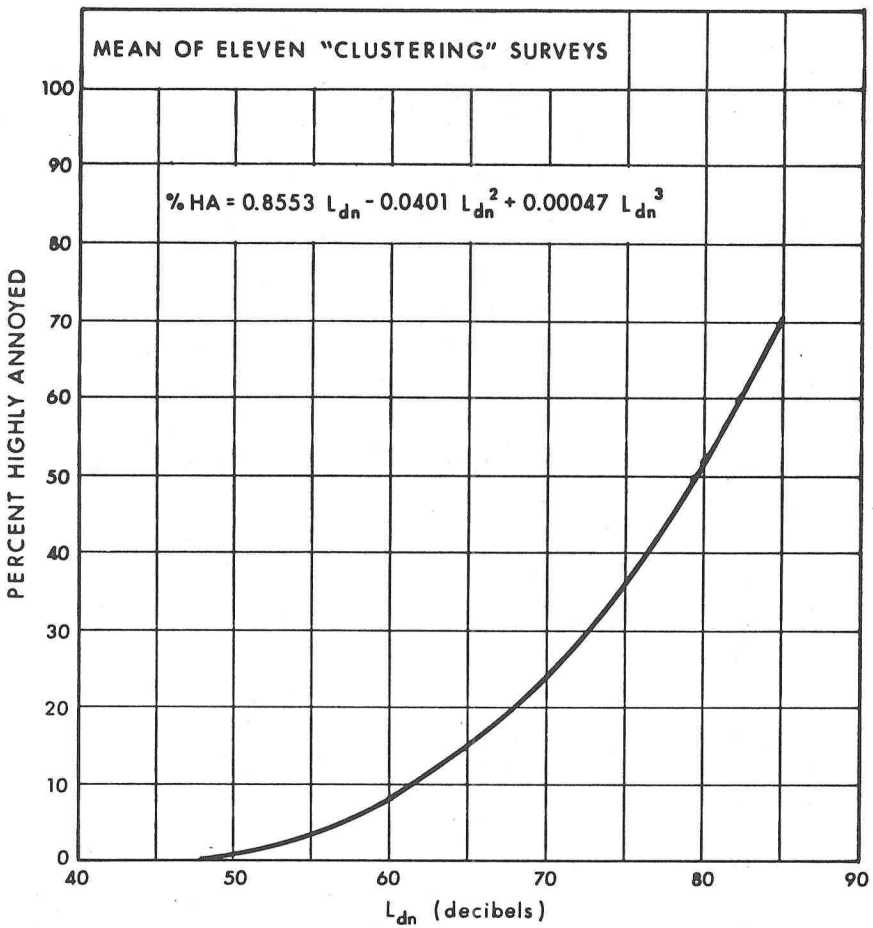

Fig. 3a. Schultz's synthesis of survey results. The mean of data from eleven surveys, shown here, has been proposed as the best currently available estimate of public annoyance due to noise of all kinds. (from Schultz, 1978)

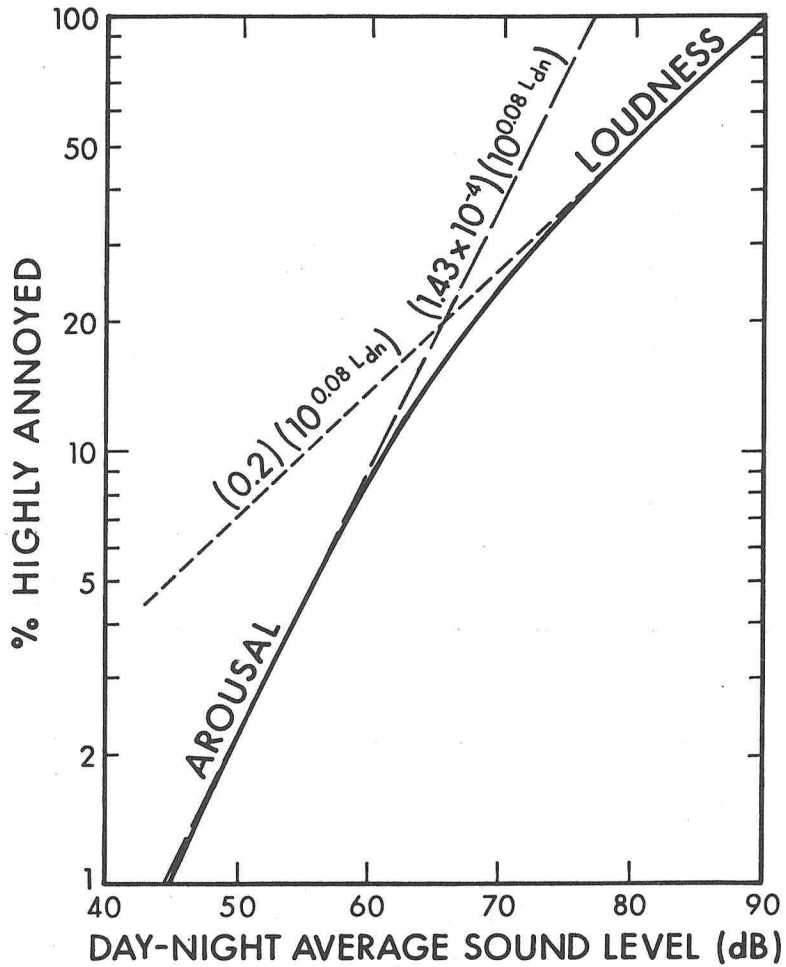

Fig. 3b. Schultz's power function approximation to the cubic equation for relating annoyance to day-night average sound level. (from Schultz, 1978)

If this relationship is indeed exponential, then the relationship between sound level and the logarithm of annoyance should be approximately linear. In other words, a constant difference in noise level should be related to a proportional increase in annoyance. For example a 10 dB increase is about twice as loud and possibly roughly twice as annoying on average.

(c) Application - Theoretical Interpretation and Extension of Survey Data

In addition, some evidence of a nonlinear relationship between sound level and percent highly annoyed for grouped data is reported by Schultz (1978b), as reproduced here in Fig. 3. Despite the initial resemblance of the curve to an exponential relationship, however, this nonlinearity is not expected to and does not follow the same relationship as degree of annoyance with sound level. This underlying relationship is most reasonably assumed to follow the cumulative normal distribution, as originally proposed by Fechner (1860), who suggested that the conversion of response frequency data to normal deviates should provide a straight line when plotted against the physical parameter in his psychophysical experiments.

Another way of regarding the synthesized sociological survey data of Schultz (1978b) is to consider the best estimates of percentage highly annoyed at various noise levels as sample estimates of the underlying probability of high annoyance at these noise levels. The true probabilities should follow a statistical distribution. The most reasonable first assumption is that it is the cumulative normal. The normal sigmoid curve may be changed to a straight line by applying a probit transformation to an ordinate in percentages or probabilities (see Fig. 4(a) and (b)). The probit scale is simply a scale in which each unit is a normal equivalent deviate (N.E.D.), or a standard deviation of the applicable normal distribution. In Fig. 5, Schultz's (1978b) synthesized curve is plotted with a percentage scale linear in probits. From Fig. 5 it is obvious that the relationship is effectively linear. The relevant curve describing Schultz's (1978b) synthesis is therefore the cumulative normal curve with mean about 79 dB and standard deviation about 13 dB.

A more precise specification of these parameters may not be advisable without direct access to Schultz's synthesized data. In other words, the synthesized model is purposely descriptive whereas the proposed model is desired to be theoretically valid. Specifically, the least squares criterion of the descriptive regression fit will tend to assign less weight to low (or high) values of the grouped (percentage) observations than to intermediate values. As put by Finney (1971, p.180), "grouped data tend to underestimate the slope of the line, which should be drawn so as apparently to err slightly on the side of steepness."

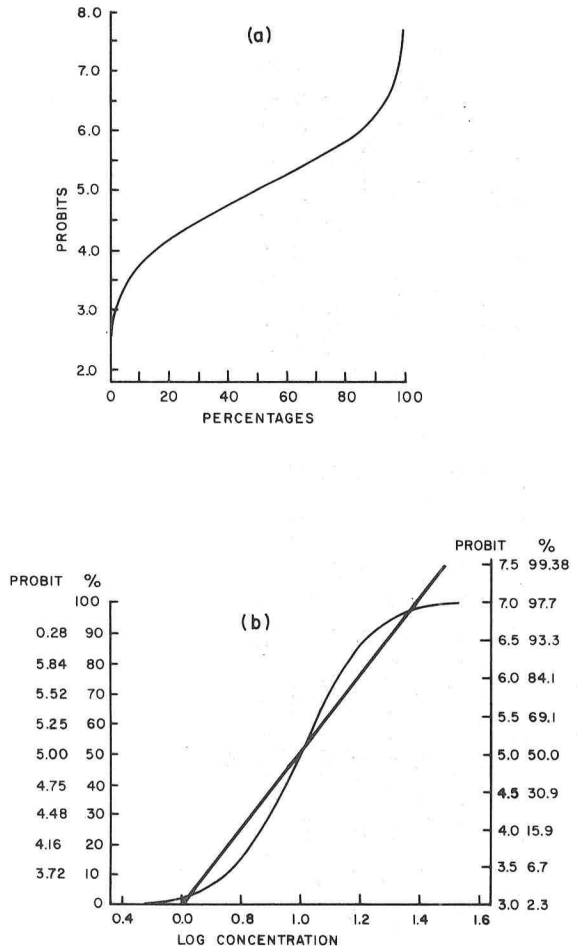

Fig. 4. When the probit transformation shown in (a) is applied, the normal sigmoid curve is transformed to a straight line as shown in (b) when ordinates are on a scale linear in probits instead of percentages. (from Finney,1971,pp.23-24)

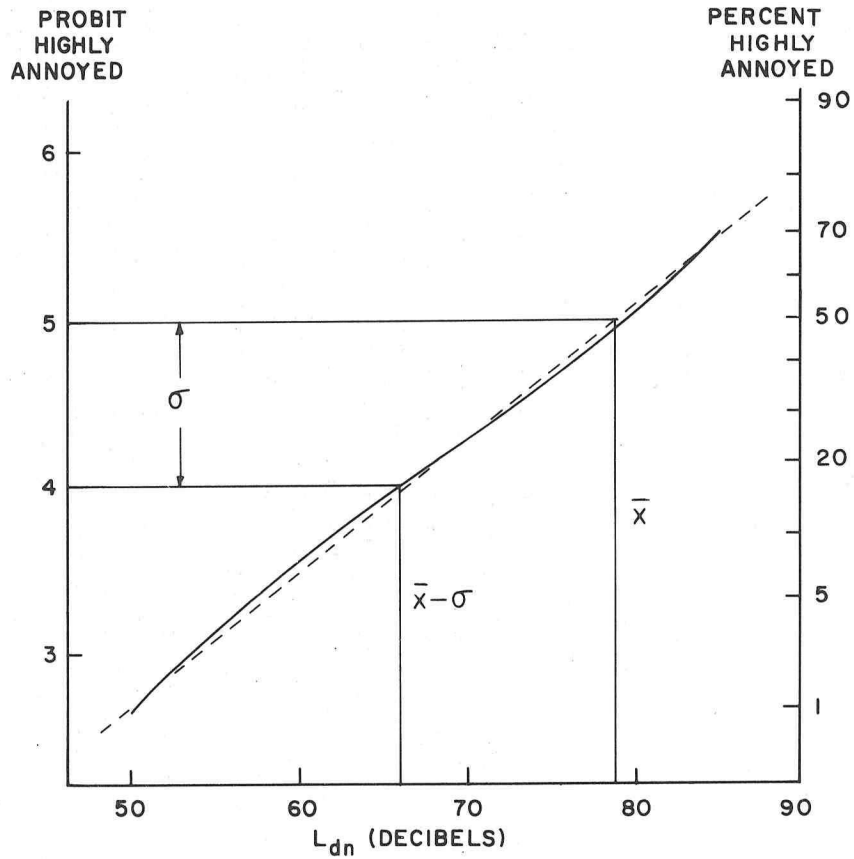

Fig. 5. Schultz's (1978) synthesized survey results plotted with a probit transformation of the percentage scale of annoyance.

The proposed theoretical sigmoid function has both theoretical and practical advantages over Schultz's descriptive model. For example, it is more reasonable to extend a theoretical curve to include lower and higher noise levels (as recognized by Schultz, 1978a). Also, apparent discrepancies in results may be clarified. Referring to Fig. 4(b), we note that responses of between 20 and 80 percent are effectively described by a linear model. The bias of percentage estimates toward decreased slope is unlikely to be of practical significance unless the line is extrapolated beyond this sample space of the range of predictor (independent) variables. On the other hand, if response probabilities (or percentages) are frequently lower than 20 and higher than 80 percent use of a probit transformation may be effective. For example, low response probability is expected for "highly disturbed" ratings or "low-noise" sites (e.g. arterial traffic) and high response probability, for "somewhat disturbed" ratings or "high-noise" sites (e.g. freeway traffic). In such situations, the researcher commonly obtains anomalous results as a lack of significance on statistical analyses.

Unfortunately, probit transformation can only partially rectify the handling of such extreme data. Two reasons are the bias of grouped estimates previously noted and the fact that probits of 0 and 100 percent are not defined, although ways of dealing with these problems are available. One solution is to perform the analysis directly on individual data as opposed to analysing grouped estimates obtained from this data. The most appropriate analysis of such data, particularly when several predictors are considered simultaneously, appears to be probit or logit analysis, as previously applied to similar questionnaire data (Ugge, 1977; McCafferty, 1978).

13. STATISTICAL ANALYSIS

The analysis is performed on data structured as in Table I. Practically, nonparametric statistical tests are performed on data assumed to be on an interval-scale level (i.e., equal units) or better. For example, the percentage scale is effectively not an interval scale of annoyance or other human reactions at its extremes. Also since this scale is bounded at its extremes, the "homogeneity of variance" assumption of commonly used tests is violated (without probit transformation), the variability of extreme values being restricted. The bias is expected to be toward increased significance and decreased slope of a regression (or least-square fit) line. Nonparametric statistics are more safely used when the researcher lacks confidence in the parametric characteristics of his data.

To the reader unfamiliar with the field of applied statistics, lest the negative tone of the previous discussion be overinterpreted, it must again be stressed that the violation of statistical assumptions is routine, most popular tests being relatively insensitive to minor violations. The caution applies to the case of a researcher unintentionally violating these assumptions and consequently falsely interpreting statistical results.

SAMPLE OF 30 (GROUPED)

INTERCORRELATION OF NOISE DESCRIPTORS

DIFFERENCE $(.6-C_L)$	.95	.9	.8	.7
.05	X	X	X	X
.1	SIG	X	X	X
.15	SIG	SIG	X	X
.2	SIG	SIG	SIG	X
.25	SIG	SIG	SIG	SIG

SAMPLE OF 300 (INDIVIDUAL)

INTERCORRELATION OF NOISE DESCRIPTORS

DIFFERENCE $(.4-C_L)$	.95	.9	.8	.7
.02	X	X	X	X
.05	SIG	SIG	X	X
.075	SIG	SIG	SIG	X
.1	SIG	SIG	SIG	SIG

Table II. Significance of differences between correlation coefficients, based on typical intercorrelations of noise descriptors and sample sizes used in sociological surveys (c_L = correlation coefficient of lower value; significance level: $p \leq 0.05$ for two-tailed test).

14. INTERPRETATION OF RESULTS

This final step of the survey encompasses all previous steps, as already noted. As a simple illustration, a statistical treatment of a controversial topic is illustrated in Table II. The topic is the choice of the "best" noise descriptor. The basis for the choice, statistically, is the significance of differences between correlation coefficients of various noise descriptors with a selected measure of human reaction. Practically, however, one faction reports all differences as being (non-statistically) significant. The opposing faction may cast doubt on all differences. A third faction may consider the problem of no consequence since all noise descriptors are highly intercorrelated in any case and human reactions much less so. Although the final conclusion must inevitably be based on a philosophical premise, the objective basis for the decision is significant by its absence in all reports coming to this author's attention. Therefore it is presented here (Table II).

It is of interest to observe that higher intercorrelations between noise descriptors give greater confidence to observed differences in correlation coefficients with human reaction. Differences as small as 0.15 or less may be judged statistically significant when noise descriptors near freeway sites are compared in the manner described.

If the data is analyzed individually, confidence in the smaller obtained value of the correlation coefficient is greater, differences as little as 0.05 being statistically significant. In addition, the statistical significance of the correlations themselves relative to correlations of grouped data are considerably more significant (from studies: Hemingway and Krammer, 1977; Seshagiri and Krammer, 1976), allowing more confidence to be placed in the existence of a relationship. Please note that the smaller coefficient is the more powerful here and that this comparison is between two different dimensions. For grouped data, the dependent variable is probability (observed percentage) of disturbance at or greater than a defined criterion. For individual data, the relevant measure is degree of disturbance. The justification for their comparison is the close agreement of results obtained by each method (Hemingway and Krammer, 1977; Seshagiri and Krammer, 1976). Also, individual ratings may be truncated at the same criterion used for grouped data. The primary objection to the analysis of the original individual data is that the measure of individual disturbance may not represent an interval-level scale and that quantitative data may therefore be suspect. Nevertheless, emp rically, it is the ratio of the random error (largely due to individual differences) to the scaling error that is more directly relevant to the interpretation of results.

SUMMARY

Major considerations relevant to statistical analyses and their interpretation have been discussed under the following headings:

1. Problem Definition
2. Questionnaire Design and Coding
3. Choice of Sites (Sampling)

4. Consideration of Cost Effectiveness

5. Interviewer Training

6. Pilot Study

7. Revision of Questionnaire

8. Data Collection

9. Data Coding

10. Keypunching

11. Computer File Structure

12. Data Transformation

13. Statistical Analysis

14. Interpretation of Results

Recognition of the considerations discussed was viewed as the major factor determining the success of a survey. The choice of a specific methodology on the other hand, was viewed as an optimization of such considerations as they relate to the specific application. The successful interpretation of results was viewed as a product of the recognition of inevitable weaknesses in methodology and the elucidation of theoretical and philosophical assumptions.

A companion paper will express the assumptions discussed here as a mathematical theory. Theoretical fits and interpretations of social survey data and their application to the synthesis of social survey data and the estimation of noise impact on people will be discussed in further follow-up reports.

ACKNOWLEDGEMENTS

The author is indebted to brief but valuable comments and leads provided by Dr. Ogilvie and Dr. Andrews, Applied Statistics, University of Toronto; Dr. F. Hall, McMaster University; Dr. T. Embleton, NRC; Dr. B. Dunn, University of Calgary; Dr. D. Sweeney, DCIEM; and Mr. Sharma Central Statistical Services, Gov't. of Ontario. Mr. Purchase programmed the statistical test for differences in correlation coefficients. Special thanks to Mr. A. Ugge, Systems Research, M.T.&C. for the loan of his thorough collection of references on modal split analysis and his readiness to discuss probit/logit models. Foremost, this work would not have been possible without the foresight of Mr. J. Manuel, Supervisor, Noise Pollution Control Section, in recognizing the value of an interdisciplinary team approach by involving the author in-depth throughout the Impulse and Railway

Yard Surveys (Seshagiri and Krammer, 1976; Seshagiri, 1977; Krammer and Dixit, 1979) and during the computer analysis and interpretation of results phases of the Metro Toronto Survey (Hemingway and Krammer, 1977) on annoyance to community noise.

REFERENCES

Berglund, B., Berglund, U. and Lindvall, T. (1976) "Scaling loudness, noisiness and annoyance of community noise", J. Acoust. Soc. Am.60(5), Nov., pp 119-1125.

Bradley, J. (1977) "Some comparisons of traffic noise measures", Meeting of the Acoustical Society of America, Pennsylvania State University, May.

Hall, F.L. Breston, B.E. and Taylor, S.M. (1977) "The effect of highway noise on residential property values", Presented at the Canadian Acoustical Association Annual Meeting Ottawa, Ontario, October 27.

Hemingway, J.R. and Krammer, F. (1977) "An investigation of the human response to highway, arterial and railway noise", Presented at Canadian Acoustical Association Annual Meeting, Ottawa, October 25.

Fechner, G.T. (1860) Elemente des Psychophysik. Leipzig: Breitkopf und Hartel.

Finney, D.J. (1971) Probit Analysis. University Press: Cambridge, England.

Krammer, F.G. (1977a) "The application of statistical techniques to sociological noise surveys." Presented at the C.A.A. Annual Meeting, October.

Krammer, F.G. (1977b) "The application of statistical techniques to the interpretation of sociological noise surveys." Ontario Ministry of the Environment Internal Report.

Krammer, F.G. (1978) "Predicting human reaction to community noise," in Proceedings of the Eleventh Annual Meeting, Human Factors Association of Canada, September 7-9, pp. 11.1.-11.8.

Krammer, F.G. and Dixit, A.K. (1979) "Community reaction to railway yard noise: Annoyance and sleep disturbance related to noise level exposure" Ontario Ministry of the Environment I,ternal Report.

Krammer, F.G. (1979a) "Human reaction to train noise". Ontario Ministry of the Environment Internal Report.

Krammer, (1979b) "Method and Interpretation of Surveys on Noise Annoyance" submitted for publication to "Acoustics and Noise Control in Canada" July.

McCafferty, Desmond (1978) "Modelling the effect of non-acoustical variables on individual response to community noise." M.A. Thesis, McMaster University, September.

Robinson, D.W. (1971) "Towards a unified system of noise assessment", Journal of Sound and Vibration, 14(3), pp. 279-298.

Schultz, Theodore, J. (1978b) "Synthesis of social surveys on noise annoyance." J. Acoust. Soc. Am. 64 (2), August.

Schultz, T.J. (1978a) Personal communication, February 1.

Seshagiri, B.V. and Krammer, F.G. (1976) "Community survey of human reaction to impulse noise", Presented at Canadian Acoustical Association Annual Meeting, Vancouver, B.C., October 6-8.

Seshagiri, B.V. (1977) "Community reaction to impulse noise" J. Acoust. Soc. Am. 62, Suppl. No. 1, p. 57. (A).

Ugge, A.J. (1977) "Disaggregate modal split models of downtown trips in Metropolitan Toronto", Ministry of Transportation and Communications, Systems Research and Development Branch. Report No. 3303, April.