CHAPTER VII: REVIEW

By J. Beaman

In the following, one will find no explicit comments on either TN 15 or TN 42. As indicated in the Introduction, TN 15 is provided only as an example of the application of methodology. TN 42 was a combination of general statements about a particular system by Mike Goodchild. : it does not pursue the matter of how a system should be developed, rather it presents some alternatives. For these reasons there is really no basis for constructive critique or other review. However, the reader who wants some value judgments about developing geographic information processing systems will find these elsewhere in this Volume, where there are some quite pointed statements about the value of using geographic information processing systems. The reader may wish to refer to TN 27 and the review remarks on it.

From a global and negative perspective, TN 24 does not really fill a large gap in CORDS analysis. The note offers too little accuracy assessment. This is true in terms of the scope of activities covered and in terms of depth of analysis to give the reader any kind of good feeling for the general accuracy of the CORDS National Survey data and the problems with these data. This is not a critique of the author but rather of CORDS and a critique that applies to many (actually most) other studies of peoples' participation in activities. Fantastic amounts of money are spent every year collecting information about which there exist major doubts either about the reliability of the information gathered or its validity.

When it comes to the matter of reliability and validity, one touches on one of the strong points of TN 24. There are many researchers who are thrust into a research role even though their training may be in planning or other fields. These people, as well as some researchers who should know better, either show no concern with accuracy of information that is collected or, particularly in the case of statisticians, carry out some kind of accuracy analysis on the basis of a split sample, which gives estimates of reliability but which does not confront the basic issue -the accuracy of the information collected. One example serves to show the kinds of concerns that should have been dealt with respect to all activities on which participation data were collected in CORDS. When one uses the 1972 CORDS data on numbers of trips to particular sites to estimate actual 1972 use, one gets two to four times as much use as there actually was. The dramatic way in which this illustrates the deficiencies with the CORDS National Survey data removes any need for further comment on TN 24. The example shows both the lack of reliability and lack of validity of CORDS data. The magnitude of errors (related possibly to poor questionnaire designs, possibly to poor interviewing or to something else) has resulted in peoples' responses to some questions having almost no relation to reality.

To a researcher with a theoretical or mathematical orientation TN 38 has to be a rather fascinating technical note, although the reader who looks for elegance in such treatment is probably rather disappointed in that a number of issues seem to be left hanging, for example:

(1) Why should one use the linear exponential distance function in such an analysis?

(2) Why not start by introducing all the price terms that are ultimately going to be considered so that the results to be presented can be derived in a A-B-C fashion?

(3) What are problems with the formulation that are alluded to, other than the condition that suggests that an increase of price per visit of 10 cents results in the same change in use whether the current charge is $10.00 or 10 cents?

(4) What problems with other methods of estimating consumer surplus, etc. does this method overcomes and are there reasons to suggest that this approach gives better results than such methods or does it merely hide problems in a different mathematical way than some they are

hidden by different approaches?

Still other matters can be cited as showing areas that the paper did not explore and which would have been desirable to explore, but in all fairness the objective of the paper was to deal with one problem. A single paper cannot deal with all matters related to a new methodology nor can the author of the paper be expected to enumerate all the issues that arise. So this paper is in some ways problematic in the way in which earlier papers on estimating demand function have been.

The points raised above are presented because this reviewer feels that the method proposed in TN 38 should be used intelligently. Also, he feels that there are many areas of research opened up by TN 38 which cry for further exploration and that these areas should not be ignored because of disputes among experts about the merits of the consumer surplus concept or because some economists say certain approaches to certain problems should not be used. In fact, in a political situation, choices must be made based on the approach that the opposition will use. Therefore, in resource allocation disputes, high resources values corrected for time bias, etc. should be used and defended rather than losing some dispute with resource developers because researchers chose to be "purists" and the “opposition” produces results to advocate for their position.

The material in TN 8 reflects a shift in emphasis in developing the "loading curve" methodology that took place from 1971 until 1975 (when the note was put in roughly its final form). Since the author of this review is one of the co-authors of the note he knows that, originally, the paper was to be about decomposing all use of a park into two functions, a uniform function and a peaked function. After the actual estimation exercise began it was realized that such curves should be developed for individual origin-destination flows. So, even though the methodology is described as if results had been obtained for origin destination (city-park) pairs estimation results are presented for the total use of certain parks for camping.

As research on park use has continued, it has become clear that to capture the real "structure" associated with visiting parks, it is not enough to develop uniform functions for different origin-destination flows: They should be developed for different classes of visitors. Different functions apply to day-users, long stay campers short stay local campers and non-local users who obviously had other trip purposes, etc. (e.g. see proposals for modelling in TN 40). But it was also clear that the CORDS Park User Survey data were so poorly documented (see the CORDS Data Documentation Volume) that the data set could not be used to estimate the coefficient for the kind of disaggregated loading curves to which reference was just made.

This obviously biased reviewer believes that this note gives researchers insight into questions that should have been asked in CORDS surveys regarding types of park visitors. In other cases it points up technical problems regarding the need to build models for different types of users based on properly collected and weighted data. Also this reviewer believes that the use of the loading curve decomposition method does have great potential in aiding researchers to make effective use of the rather limited information that should be collected in most surveys.

In an organization which is supposed to be efficient in collecting information and utilizing it, it is important that there not be "overkill" in data collection, so if using such an analysis approach adds to the efficiency of use of information this is good. But, as to the implications in the article that the effects of weather can be derived, the merit of this has not been established. A practical question is: What information about the effects of weather on park use do researchers usually need to generate for managers or for planners? Obviously, if one has loading curves for a park and these are broken down by different origins and one has an idea how weather affects weekend, week-day etc. visitors from near and far, then one can take historic weather data and see what effects weather is expected to have in the future. Unfortunately a park does not staff for the exceptional season and analyses may be of more academic interest than of practical interest. The idea of using such information in park design may have some merit, but in reality the kind of weather information that researchers should make available to planners for master planning is down to earth description of the weather in terms that relate to the facilities that they are considering. If weather is often bad so that some of the kinds of users who are predicted to come to the park would not want to be outside, the researchers may ask the planners: Is there provision for an indoor interpretation centre or are there some alternative activities indoors that relate to the objectives and design of the park and such that weather will not be a problem in encouraging use of the park?

The previous points make it clear that much work needs to be done before the "loading curve" methodology can either be endorsed for some purposes or generally rejected as of little use except for some academic research projects.

In reviewing TN 21, one could present a commentary on what happened in CORDS and Parks Canada Park User Surveys from 1968 to 1974 but it would be very similar to the "hindsight" history presented for TN 8. This is because what is presented in the note is some information about a methodology that evolved. However, introductory comments to this chapter and overview comments in Chapter XI cover some relevant background details. So some specific innovations in survey weighting and data collection are noted both to bring them to the readers' attention and to be able to contrast their relevance to general needs for data collection analysis methodology. This review sees methods of data collection other than conventional surveys as very important in meeting many of Parks Canada's needs effectively. Regarding data weighting, a procedure was employed that filled the time spent when interviewers were on lunch and or on other breaks by interviews. One has seen that such gaps were "filled" by data that were collected near the time at which a break was taken. This should help to ensure that the interviews as nearly as possible "represented" the data that would have been collected in the period in which no interviewing took place. In many cases all that is done in a survey is that information is collected for a certain day on the total number of vehicles entering a park, and the questionnaires for that day are given a weight that establishes what part of total traffic (how many vehicles out of the total traffic) a certain questionnaire should represent. This can be problematic when visitor attributes are correlated with data collection times.

This brings up the point that the weighting for a particular interview is much more disaggregated under the system presented than under usual weighting procedures. The fact that information collected in each half hour is related to entering traffic during that half hour (and that this is done on the basis of the breakdown of that traffic by origin, etc.) reflects a legitimate concern with getting as much information out of a given amount of labour input as possible (as well as a concern with accuracy). By having what is described in the article as a "floating sampling rate", it is possible to have field workers work hard when traffic is heavy (by making contact with people in as many of the vehicle as possible that pass through a gate) and at the same time work as hard as possible when traffic is slow by getting interview information from people in all of the vehicles that pass through a gate. One should note that the survey strategy used by Parks Canada from 1971 - 73, which involves getting what is called "entry record information" from every vehicle, is important because many of management's questions can be answered by using only entry record information, most of which is gained by simply observing a vehicle and its occupants. These figures can be obtained very accurately because many vehicles can be stopped when only one or two questions are asked (e.g. do they plan to stop in the park). If the parties in the vehicles do not hand back their questionnaire (and often 50% or more do not) this entry record information can still be used in improving the weighting of those questionnaires that are returned (when date-time of entry, first-repeat entry, party size, state of car plate, and some other information are in entry records).

The preceding comments serve to acknowledge that much of what is recorded in the TN involves potentially useful modifications to usual survey procedures. Survey “enhancements” either serve to improve accuracy/reliability statistically and/or increase the cost efficiency with which a given level of accuracy/reliability is attained. However, in 1973 development of computer programs to process the data collected, the rather time consuming procedures of defining appropriate weights for different survey days for which estimates had to be made, etc. involve much more work than occurred when a survey was processed by simply specifying a collection of day-by-day or “stint” weights to be associated with all data collected at certain times. In 1973, depending on what accuracy/reliability of results was needed, there could be justification for employing a very simple survey strategy. The amount of time that survey staff will be used (related to the total data to be collected) and the improvement in accuracy attained by using more sophisticated surveying procedures that increase processing and editing costs, and use excessive manpower, etc. could make a "simplistic" but well thought out survey procedure cost effective. Still, no comparative accuracy figures, procedures for determining what accuracy will be achieved one way or another, etc. were created. Post 2000, technology has changed radically with computer use cost for surveys being nil and automated data capture and accurate processing being facilitated. Therefore, criteria based on 30 year old technology would not be relevant in the second millennium.

Possibly the most important thing to realize regarding Technical Note 21 is that it presents details on a survey procedure which, although carried out from 1971 to 1973, was not the best procedure for collecting the information for which there was really a need. One can now see that person power was used in an inefficient way in counting the vehicles: traffic counters that cost less than one field "surveyor's" salary for the summer could have been used. One may also note that traffic counters do not take afternoons off, nor do they take coffee breaks or lunch breaks: they usually work seven days a week for 24 hours a day with a minimum of care.

It is not unusual to find that there are money and person power available to carry out a survey that are quite out of proportion to the resources available to decide what the objectives of the surveys are and to take care in operationalizing these into a data collection analysis, report preparation effort. For example, if, as has been the case in the past, the primary use of survey information collected at National Parks is to develop good use estimates for those parks, it has been recognized that there is no need to stop vehicles at park gates. In 1975 a new method of getting data for use estimates was being tested in which entry to parks is recorded by traffic counters at all park gates. "Survey teams" simply obtain license plate numbers and other data (such as the number of occupants of a vehicle) for certain blocks of time. Similar information collected at park facilities (campgrounds, etc.) allows the movement of a vehicle within a park to be traced and (for example) allows length of stay of visitors in a park to be monitored in a truly accurate way. It is surprising how much information can be generated in this way.

Now a reasonable question is: What is the cost of non-traditional survey related procedures and are there operational problems involved in employing them that may mean that, in many cases, a survey strategy where entering or exiting parties are actually interviewed may produce acceptable results at a lower cost? One should note that having an objective record of what people did in a park is often superior to results obtained by survey because people often do not recall what they did or know exactly what something they did is called, so that they often cannot answer a questionnaire adequately. One example is that when the standard park users survey technique reported on TN 21 was "improved" and employed in a survey in Prince Edward Island National Park, it was found that so few people realized where they had been in the park that use estimates for the Park were 50 percent higher than the known actual use. (An error of over 250,000 visitors was made in estimating the number of out of province visitors to the park.) This kind of error simply cannot occur with a study in which license plate information is recorded. In this latter type of study one is not dependent on a person recognizing that certain locations are park boundaries.

Thus although TN 21 has presented some useful innovations in survey weighting, it has also indicated that refining weighting to gain slightly more efficiency in estimates of some use figures may miss the real point of what needs to be done. It is impossible to stress sufficiently the importance of careful definition of information generation objectives. Collecting information that is more accurate than is needed, that is more extensive than is needed, etc. are park user survey inefficiencies that really need to be eliminated. However, eliminating them is far from simple and it is only as researchers proceed to build up a catalogue of the best current procedures for information collection to meet specific objectives that there can be widespread use of appropriate research tools.

The key need today in recreation research is careful definition of objectives of research and careful operationalization of these objectives in as efficient a way as possible by a thorough description of exactly how such and such a table or factor analysis result will be presented to a planner or manager to answer questions. The researcher with practical concerns working for a parks organization or at a university cannot be expected to be totally aware of all of the statistical "tools" that are best to achieve certain ends for the variety of projects that arise. But this will be less of a problem when specialized researchers have produced (and updated on a regular basis) a procedures manual. Such a manual must not just be a survey procedure manual but must have statements of objectives related to an information need. Based on this, a discussion should present how a specific information gathering procedure is justified by explaining exactly how the data obtained are to be processed to produce the tables, etc. that are to be used in achieving the project objectives by doing (a)..., (b)..., etc. It is this kind of complete project analysis guide that is desperately needed. This approach to project improvement is not described in TN 21.

There is commentary elsewhere in this volume on TN 10, specifically in relation to Technical Notes 32 and 37. (See also various commentaries in Chapters VI, VII, IX and XI.) Actually, Technical Notes 32 and 37, as indicated in Chapter VI where these notes appear, deal with many of the matters that might be taken up in review comments here so the reader is referred to these notes for details on the possible applications of methodologies described in TN 10 and a critique of the methodologies.

The one point is not sufficiently stressed elsewhere. This concerns determining natural "classes of people" as opposed to mathematically defined homogeneous groupings. The concern is with a scheme that allows individuals to be defined as members of a class to which they belong even if some of their characteristics are at odds with some characteristics of other individuals in the same class. For example, who would think, seeing a tamarack without its needles, that although most people call it an evergreen, it is in fact a deciduous conifer. That its needles fall off in the winter is not the important factor to consider in classing this tree as a conifer. Similarly, in the context of the outdoor recreation demands study data, it is not surprising that in the middle 60's very few people on the Canadian prairies skied (Nordic) during the winter. There were no ski hills! So in defining activity packages it is critical that a person not be excluded from a certain activity package which includes people with participation in other similar activities simply because at a given point in time they live in a region in which they do not have facilities for certain activities. Such people may even substitute snowmobiling for skiing so that their activity package appears to be more disparate with some people with whom they should "naturally" be grouped than if they simply did not ski.

The clustering computer program used in deriving the clusters presented in TN 10 was not set up in such a way that supply or other trade-off kind of considerations could be taken into account. It is possible to argue that by using a special computer-program to prepare data for analysis, one could structure information so that skiing would be adjusted for supply and could be e.g. treated so there could be a trade-off with snowmobiling in depending on supply. Assuming a program can be created, using it and deriving clusters, would allow one to argue that more "natural" classes of people were derived. In CORDS one used the "monothetic-divisive" clustering program of TN 10 or an agglomerative program (Romsa 1973; appendix TN 10). With the divisive approach, a person can be excluded from a clustering because of the "score” on one variable. Simply going to the kind of agglomerative programs described in the Appendix to TN 10 just reduces the risk of a person ending up in a cluster when there might be one with better overall compatibility. Regardless of which kind of clustering program is used, it is essential that much thought go into considering how availability of supply affecting participation relates to clusters determined. If supply is available for X and a person does not participate in an activity, then an important question is: is it the availability of supply for an alternative activity, Y, which has resulted in the lack of participation? Has substitution taken place or is this person legitimately different from others who participate in X because of preferences for activities? If the supply configuration confronting two people is similar and one participates in one activity and another person in another activity, substitution is probably not an issue. Still, even faced with common supply, one person may participate in one activity and another only because of background, what friends do, etc. But, when supply varies an assessment is much more difficult to make. Given that TN 29 has shown something of our limited capability to know how much supply a person perceives to be available, one may see the impossibility of making much progress in studying substitutability or improving clustering until there is some progress in measuring people response to supply in making decisions (this is taken up in some detail from a theoretical perspective in TN 33).

Obviously, much more theoretical writing on the problem of forming clusters in a behaviourally meaningful way must be done in parallel with discussion of how this theory can be operationalized in terms of using various clustering techniques. TN 10 has only demonstrated the feasibility of two techniques for extracting information about clusters of people and clusters of activities. Furthermore, in the perspective of this reviewer, the procedure of factor analysis actually has little application because it is not tied closely to behavioural considerations except in the context of an aggregate theory which should be of little relevance in recreation planning. So in the note one has really only seen demonstrated one analysis procedure which may eventually have a great deal of utility but now has limited utility (see TN 32).

TN 19 is a much simpler note on which to comment than TN 21 or TN 10. It has simple objectives that were attained and gives a relatively straightforward way to obtain more efficient estimates for the parameters of gravity models than are obtained by the ordinary least squares procedures usually employed by researchers (e.g. as used in TN B, TN 4 and TN 18). Also it was possible to define an absolute test of model adequacy. The importance of this finding and the practical application (and implications) of it are adequately stressed and explained in the note.

There is one point which is not stressed, on which some readers may find it useful to have some comments. If a model is not structurally adequate for the data to which it is applied, the use of GLS estimation does not necessarily result in improved estimation efficiency. Yet this should not be taken as a reason to avoid using GLS estimation. At least if GLS estimates are used, valid tests on model adequacy can be made to assess if a given model is adequate. Certainly nothing is lost by using GLS estimation. And what is gained, as is stressed in the note, is that one can examine the “standardized” residuals to see which differences between observed flows and their predicted flows must change to improve a model.

TN 36 cannot be faulted on the basis of the achievement of the basic objective that was set for the research reported. The author proceeded in a rather elegant way to show the value of R^2 that would be expected when models like those presented in TN 12 were developed. However, some readers may wish that the author had used a rather more easily understood derivation that does not involve using the calculus. Dealing with the problem using discrete distributions and "simple" probability theory is possible and quite easy. It is probably also true that the reader who has a strong enough methodological orientation to read the article has the calculus necessary. Still, in the view of the reviewer, the average reader has a better chance of getting a clear idea of what has been proved, why and how if the discussion is translated into one using summations and "discrete" probabilities.

The substantive grounds on which this article can be criticized obviously do not relate to the main objective of the article. Rather, when the author proceeds to introduce a test for the structural adequacies of a model he does not consider that the ordinary least squares estimation approach on which his formulas are based gives residuals which should not be used in the test proposed. Basically, the author (as indicated in TN 12 and a footnote in TN 20) is working with a model in which the assumption is that one is calculating probabilities of people's participation. These probabilities (the dependent variable) when estimated define a heteroscedastic dependent variable distribution, so the author's test for structural adequacy must be made taking this into account. His test equation should be replaced by:

$$\chi^2_{(N\text{-}df)} = \begin{cases} \Sigma\,(O_k-P_k)^2/P_k(1-P_k) \text{ for all } k \text{ WHERE } 1-d > P_k > d > 0 \\ \Sigma\,(O_k-P_k)/d(1-d) \text{ for } P_k > 1-d \text{ or } P_k \end{cases}$$

WHERE $\chi^2_{(N\text{-}df)}$ is approximately a χ^2 distribution with N-df degrees of freedom

O_k is an observed value of 1 for participated or 0 for did not;

P_k is the predicted “probability”;

N-df is the number of observations minus the degrees of freedom “used” by by regress (number of parameters including general mean) and

d is a constant such as 0.01 or 0.001 such that if P_k is within d of being outside the 0-1 interval it is treated as having the value 1-d or d based on being near 1 or 0.

The reason for using this formula is indicated in the note below. If one estimates the model parameters, considering the heteroscedasticity of the dependent variables, then the formula would be used after the results of estimating the model a first time have been used to specify that the variance in associated with a respondent having a probability P_k is $P_k(1-P_k)$. For P_k outside d to 1-d, variance would be taken to be d(1-d). A second GLS estimation cycle would get new probabilities (P_k) and this could continue. However, what may be established is that $\chi^2_{(N\text{-df})}$ remains significant because the model simply does not fit the data well.

SUPPLEMENTARY NOTE

To give the reader a further feel for how CORDS research has progressed, the following note has been added to this chapter review, rather than changing the material on the "chi-squared test of model adequacy" just presented. While the author of this review was working on revisions to what had been thought to be the final version of TN 6 and reviewing the Smith and Cicchetti article (see the Appendix), the practical import of the following conclusion by Smith and Cicchetti became clear:

> *Most survey research using dichotomous dependent variables (0-1 responses) for the estimation of linear probability functions utilizes fairly large data sets (in excess of one hundred observations) for statistical estimation. As a consequence, "probit-like" estimators impose substantial computational costs* (in 1970-1975). *Therefore, OLS or GLS procedures are frequently chosen. While our results cannot discern the loss in efficiency if an OLS estimator is utilized, they do indicate that there is little or no gain from Aitken/Goldberger estimation (McGillivray 1970) for such large data sets. Moreover, traditional hypothesis testing with Student's t-test for the coefficients of such models can probably be expected to yield reasonably powerful tests with large samples.*

When the estimate of variance in predictions of number of male hunters in Quebec were computed and it was noted that with regression coefficients based on 1291 observations, both unweighted and weighted regression results gave equally accurate predictions. A question had to be answered! Why go to the trouble and expense of carrying out weighted regressions? In reality, expense was the big concern because each weighting iteration required that data on every observation be weighted, a covariance matrix computed, etc. until finally after a matrix inversion, which was costly for large matricies, new regression coefficients were produced (hundreds of dollars per hour were paid for computation that took minutes and in 2006 take seconds on a PC). For some CORDS research, getting weighted regression results based on obtaining initial estimates and then doing a GLS estimate, involved reading data on 2,000 people from magnetic tape about 100 times and inverting a 40-by-40 matrix the same number of times.

Why incur the expense? At first, the obvious answer appeared to be so that tests of model structural adequacy, such as the one described just before this supplement, could be performed. However, some thought made it clear that if parameter estimates are not affected much by weighting when sample sizes of concern are 600 to 25,000 (which are what are being used in CORDS analyses), one could use the U and B(.)'s computed in the unweighted regression to calculate:

$$\chi^2_{(N\text{-df})} = \begin{cases} \Sigma\,(O_k\text{-}P_k\,)^2/P_k(\,1\text{-}P_k\,) \text{ for all k WHERE } 1\text{-d}> P_k>d>0 \\ \Sigma\,(O_k\text{-}P_k\,)/d(\,1\text{-d}\,) \text{ for } P_k>1\text{-d or } P_k \end{cases}$$

When observations are only accessed for carrying out a regression and computing $\chi^2_{(N\text{-df})}$, costs are a fraction of what occurs when one repeats computations multiple times.

Actually, when the chi-squared values were being generated as suggested above, other useful statistics on the a model could be generated. When there are P(i)'s that are less than 0 or greater than 1 and thus do not make sense as probabilities of participating, counting such cases based on how negative they are and even writing out data on them for further analysis can be very valuable. Often, seeing what characteristics these "outliers" have is important in identifying problems with a model.

In closing one should note that the Appendix to TN 29 shows why the chi-squared test of model adequacy for hunting carried out in TN 6 did not show that the model was inadequate. Only with the over 25,000 observations available in the data used to compute supply or with similar large data sets does the "supply factor" sum of squares become large enough compared to the TSS that it would be detected by concluding that structural error was arising for some reason. But as reported in TN 29 (and in TN 6 and TN 20), if supply factors or other "effects" that should be considered are ignored, predictions made using a model may be seriously in error EVEN if estimates of the accuracy of the predictions show only a few percent error is likely to occur. Accuracy estimates, based in the method presented in TN 6, are appropriate IF a model is structurally sound. As stressed elsewhere in this Volume, the kinds of chi-squared test introduced at the end of this review are weak tests. If one rejects a model as unsound, one can be confident but it is accepted as sound one cannot be confident that estimates made using the model will be as accurate as desired, because undetected structural problems with the model may lead to substantial biases in many estimates.

It was only while these review comments were being prepared that it was recognized that the use of the kind of iterative variance estimation (heteroscedasticity correction) just described gives a test of model adequacy similar to that developed in TN 19. When GLS (weighted) regression was used in TN 29 to explain peoples' participation in activities after the second estimation cycle one finds that:

Residual Sum of Squares = 31759

Degrees of Freedom = 25219

Because the regression has been carried out by normalizing every dependent variable observation by its variance, if the model is structurally sound, the residual sum of squares is a sum of squares of approximately zero-one variables with 25219 degrees of freedom. But one can test the value of RSS to see if it probably came from a chi-squared distribution with 25219 degrees of freedom by considering that because RSS = χ^2_{25219}, d = $(2(RSS))^{1/2} - (2(25219)-1)^{1/2}$ is approximately an observation of a normal zero-one variable (Stewart 1943, pp. 294-295). But d congruent to 18, which means that RSS is too large by so much that it would have the same probability of occurring as an observation of 19 or greater in sampling from a normal zero-one variable. This will occur less than one time in 10**50 if a rough underestimate is correct. Obviously there is room for improving the TN 29 model.