

A RECENT DEVELOPMENT IN NISAN SYSTEM

Asano, Choichiro

Research Institute of Fundamental Information Science, Kyushu University | Department of Information Systems, Interdisciplinary Graduate School of Engineering Science, Kyushu University

Watanabe, Michiko

Research Institute of Fundamental Information Science, Kyushu University | Department of Information Systems, Interdisciplinary Graduate School of Engineering Science, Kyushu University

Hosokawa, Koji

Department of Information Systems, Interdisciplinary Graduate School of Engineering Science, Kyushu University

Kurihara, Koji

Department of Information Systems, Interdisciplinary Graduate School of Engineering Science, Kyushu University

<https://doi.org/10.5109/13338>

出版情報 : Bulletin of informatics and cybernetics. 20 (1/2), pp.79-96, 1982-03. Research Association of Statistical Sciences

バージョン :

権利関係 :



A RECENT DEVELOPMENT IN NISAN SYSTEM

By

Chooichiro ASANO*, Michiko WATANABE***, Koji HOSOKAWA****

and

Koji KURIHARA**

(Received November 26, 1981)

Abstract

NISAN system has been developed by a group of Japanese statisticians, since 1978. The package is widely applicable in statistical data analyses, confirmatory and exploratory, with newly developed methods. The system is useful not only for the beginners in data analysis, but also for the senior statisticians with analytical insight, i.e. choosing adequate procedure of statistical analysis and model-building.

In the pretest paper, some features in NISAN system recently developed are presented, especially concentrating on both functions of new monitor and knowledge-commands. These two functions are mainly for statistical researchers. The former is easy to cooperate and to accumulate further commands developed with respect to data-handling and analytical methods by NISAN research members, working in the respective fields of statistics and locating in several parts of Japan. The latter is to investigate and to evaluate properties of statistical methodologies by simulations in some varieties of circumstances given, and jointly may be to recognize the mathematical model previously unspecified for data.

1. Introduction

NISAN system is a New Interactive Statistical Analysis program package, which consists mainly of the following five modules: NISAN-MONITOR module for control, STAT module for statistical methods in confirmatory and exploratory analysis of data, DATA module for data bases, HELP module for various inquiries with three levels, and DOC module for documentation. The conversational terminals are cathode ray tubes, teletypes or graphic terminals, where all the outputs in NISAN system are designed on at most eighty characters per a line.

In this way, NISAN system has been developed concretely not only on ordinarily important commands for methods of statistical data analysis, but also on some commands for statistical data-handling, e. g. stratification, graphic representation, quantification and

* Research Institute of Fundamental Information Science, Kyushu University 33, Fukuoka 812.

** Department of Information Systems, Interdisciplinary Graduate School of Engineering Science, Kyushu University 39, Fukuoka 812.

so on. Thus a relatively simple users' guide was made on this February as a trial, and the booklet obtained good evaluation for statisticians without regard to their experience of computer work. Therefore more commands of analytical methods and statistical data-handling are now reinforced.

On the other hand, two emphasized features, that is, a new NISAN monitor with effective functions and some knowledge-commands for statistical inference are recently studied and developed with the ordinary method-commands.

2. New NISAN-monitor

Generally speaking, the functions of monitor are very important for efficiency of the whole system of statistical program package. Regarding the program construction of NISAN system, the main functions of NISAN-monitor are as follows: interpretation of input commands, allocation of memories, allocation of files, control of several peripheral devices, terminals and programs.

In NISAN system, the monitor is fully written by the standard FORTRAN IV language, although only LOGON and BREAK functions depend on TSS. This assures every university computer of the portability of NISAN system, that is to say, NISAN system is independent from the kinds of computer machines, being available FORTRAN language.

The monitor interprets naturally regular commands and operands. In them, some conventions are particularly provided for users' convenience. For example,

variables $X1:X5$ means all variables between $X1$ and $X5$ in NISAN file,

variables $X(1:5)$ means $X1, X2, X3, X4$ and $X5$,

values $1:5$ means values between 1 and 5, i.e. $[1, 5]$,

and so on. Also several conventions, depending on the command, are given as shown in the following example;

levels $L((X1\ 1\ (2\ 7\ 8)\ (3\ 4))$ means making three levels of variable $X1$, where the respective levels are composed of a value 1, a set of values 2, 7, 8 and a set of values 3, 4, as elements.

For the convenience of successive replication of analysis, changing variables is considered simultaneously by arranging related variables, as shown in the following example,

$SCAT(X1\ X2:X5)$ means four scatter diagrams of $X2, X3, X4$ and $X5$ on $X1$, respectively.

Regarding operands, many varieties are prepared to obtain several features of outputs corresponding to users' requests. In case of omission for operands, either a prompt inquiry is presented or a standard value involved is given, depending on the semantics of the operands. In the monitor, the interactive functions are provided.

In view of statistical analysis, it makes clear in NISAN system for us to use statistical terminologies, defined by mathematical statistics, as the names of commands and operands. For example,

POP means POPulation(s).

TESTIM means statistical TESTs and ESTIMation.

REF means RElative EFiciency.

DEFTP means DEfinition of Test-Process.

The NISAN standard file is applied as a self-descriptive systematized file composed of both parts of directory and data, since there exist big varieties on types and formats in every file of users. The logical structure of data part is a set of record, being called cases. The file has parent file name, and each case consists of a set of items referred by variable names. Group names of variables are also assigned as key names, and in case of continuous type of variables, three data ranges are given as normal, warning and something unusual.

It is one of the most emphasized functions of NISAN monitor for the research members to construct easily new commands with operands on the basis of their FORTRAN programs. The arrangement is fill several definitions for commands and operands in a printed form, and to set initial values for arguments. Thus the setup is completed by a certain simple operations on TSS. As the manual, a thin booklet Commands Construction Manual for NISAN Research Members [3] is prepared to be easily to understand for statisticians, even if they are not so familiar with computer operation.

In view of further development of NISAN system, this function of the monitor is very useful, since the research members of the organization of NISAN system are widely located at several universities and institutions in Japan, and the respective members, working in various statistical fields, are able to integrate easily their authorized programs of analytical methodologies to NISAN system.

3. Analytical method-commands

It is natural in statistical program packages to have many method-commands to analyze actual data in data file. In NISAN system, analytical programs in STAT module are originally developed and are integrated under TSS operation. Although the whole method-commands involved are still not so many and such commands are soon to be added, a tentative users' guide of 43 pages was prepared as a trial in this February. Since the manner of writing and form was simple and customary, the booklet had a good reputation to apply easily those commands and operands. In order to show some illustrations, several commands are quoted from the booklet in the following manner.

The contents of the respective commands are composed of (0) Name of command, (1) Input form, (2) Functions, (3) Explanation of operands, (4) Pagination of illustrations of output.

4. Knowledge-commands

4.1. The principle

NISAN system is widely available for both statistical situations, confirmatory analysis and exploratory analysis, and is not only for senior statisticians to obtain analytical wisdom on mathematical model-building for the essence of urgent phenomenon, but also for the beginners of statistical data analysis to learn analytical methods and properties and to train their experience of data analysis. For these purposes in NISAN system, separating from analytical method-commands, so-called knowledge-commands have been developed to investigate or to learn about methodologies of statistical in-

6. Command MEAN2

Significance test between two sample mean vectors in multivariate normal population with unknown variance-covariance.

```
MEAN2  [FI (file name)]*
        [VAR (variable name list)]*
        POP «identification variable
            name value 1 value 2»
        [ALPHA (significance level)]
```

8. Command HIST

Output of histograms for respective variables.

```
HIST    [FI (file name)]*
        [VAR (variable name list)]*
        [MINIMUM (minimum value)]
        [MAXIMUM (maximum value)]
        [SIZE (number of intervals)]
        [DENSITY (frequencies for a
            height scale unit)]
        [CUMULATE] [NUMERAL]
```

12. Command SCAT

Scatter diagrams of the second and the latter variables on the first variable respectively.

```
SCAT    [FI (file name)]*
        [VAR (variable name list)]*
        [MINIMUM (minimum value)]
        [MAXIMUM (maximum value)]
        [WIDTH (number of cells on
            horizontal axis)]
        [HEIGHT (number of cells on
            vertical axis)]
        [GROUP «variable name value
            1 value 2... value n»]
        [MEAN] [FRAME]
```

11. Command STREG

Stepwise forward or backward regression analysis with a constant term.

```
STREG   [FI (file name)]*
        [VAR (variable name list)]*
        [FORWARD ]
        [BACKWARD]
        [FVALUE (F value)]
        [TOLERANCE (tolerance)]
        [FO (file name)]
        [NOPRINT] [LIMIT]
        [RANGE]
```

16. Command LEVEL

Level-making of each variable before cross-table analysis, analysis of variance, and so on.

```
LEVEL   FI (file name)
        L «variable name level 1 level
            2... level n...»
```

17. Command CROSS

Making two-way cross tables.

```
CROSS   [FI (file name)]*
        [VAR (variable name list)]*
        [PRINT ([EXPECTED]
            [ROW] [COLUMN] [TOTAL]
            [NO])]]
        [ASSOC (association indices
            list)]
        [LIMIT] [RANGE] [MISS]
        [AUTO]
        [GROUP (variable name list)]
```

20. Command MNT

Multivariate normality test, including
1. skewness 2. kurtosis 3. Geary's test
4. Shapiro-Wilk 5. Shapiro-Francia 6.
D'Agostino-Pearson

```
MNT     [FI (file name)]*
        [VAR (variable name list)]*
```

18. Command CLUST

Cluster analysis based on distance or similarity.

```
CLUST   [FI (file name)]*
        [VAR (variable name list)]*
        [MAT (category of data)]
        [SOL (analytical method)]
        [CL (number of clusters)]
        [GRAPH] [RESULTS] [ALL]
        [CYCLE] [INTER]
```

Note

- (1) [] means that operand and item involved can be omitted.
- (2) []* means that operand and item involved are not necessary when they are fixed previously by LOCK command.
- (3) { } means that only one operand must be selected from the plural operands shown.
- (4) Underline of operand means the shortened form.

ference, including the single method and the serial process of methods for inference.

In such simple situations, facing the statistical data analysis, one likes to obtain statistical knowledge about the properties of applied methods on the accuracies and precisions for estimates, the distributions or conditional distributions of test statistics, the powers of tests, and furthermore the relative efficiencies or the asymptotic relative efficiencies between two arbitrary tests in case of selecting statistical tests for several alternative hypotheses in varieties of circumstances to be considered. On these accounts, commands DIST, POWER, REF and AREF are available.

From the viewpoint of the process of statistical inference, generally speaking, in order to specify an adequate mathematical model for observations or to investigate the feature pattern of data, statisticians apply more or less their systematic processes, composed of several methods, to their data basing upon their logics of inference.

As a simple example of inference process, called a kind of pooling methodology by Bancroft et al., let us consider an estimation problem for two observation means with a common but unknown variance. Then it may be natural that when the distinction between two observation means is not significant by t -test, two sets of observations may be pooled and the average is given as the best estimates of both population means, because of acceptance of the null hypothesis that two population means are equal. On the other hand, if the distinction is significant, the respective observation means may be given as the respective best estimates of population means. More details are discussed in section 4.3. Ordinary processes of statistical inference may be composed of more tests, and the mathematical formulations are extremely complicated with multiple integrations on complex domain.

Even if the inference process is multiple-stage, NISAN system is able to give very quickly numerical evaluation of the properties with enough precision about inference on the basis of simulating the inference process by NISAN random numbers or by random number generation implemented and arbitrarily transformed as in Table B, keeping away from awfully complicated mathematical formulations and these numerical computations.

Such inference processes are also applied frequently in case of nonparametric inference, since the structure of observations is too complicated to know what kinds of nonparametric tests are suitable to the data feature, which may be observed from a composite population. Since every test itself has a specific aspect without a consistence in the test-statistic, and actually several nonparametric tests are tried. Then some tests indicate significant but another indicate nonsignificant. Therefore, in cases of exploratory analysis or in the situations of incompletely specified model of data, it may be dissatisfied for us to apply merely a statistical test. It may be needed ordinarily to take a test process.

Thus at any rate, the performances and properties of such inference processes have to be theoretically evaluated and, if any, the optimality improvement should be considered as statistician's knowledge. For instance, the overall power of a test-process, the relative efficiency between arbitrary two test-processes, the accuracy and precision of estimate due to preliminary test-process should be studied.

These knowledges may be also able to prevent misapplication of statistical methods

and of statistical program packages, and jointly to recognize the mathematical model of data.

For these purposes, knowledge-commands DEFTP, CONDIST, TPATTN, POWER and TESTIM are prepared in NISAN system. In the following section, more details of the commands are introduced.

4.2. Formation of each command

The manner of writing and form is, of course, similar with the method-commands but a little complicated in users' guide. In order to show concisely their functions, operands and varieties, we quote the expression of these commands in its entirety, as below.

(i) Command DIST

DIST	<u>TEST</u> (test name)
	<u>POP</u> (number of populations)
	<u>SIZE</u> (sample 1 sample 2...)
	[<u>TYPE</u> (type of R.N.)]
	[SIM (size of simulation)]
	[<u>PARA</u> (parameters of R.N.)]
	{UP LW BO}
	[PERC (percent point)]
	[PERT (probability level)]
	[<u>HIST</u>]

[Function]

To obtain the distribution of a test statistic in a certain environment given.

[Operands]

TEST ...A test name shown in Table A.
 POP ...The number of populations.
 SIZE ...The respective sizes of samples obtained from the populations.
 TYPE ...A name of type of random numbers shown in Table B. In case of omission of key-in, NOR is applied.
 HIST ...A histogram of test statistic is shown by the simulation. In case of omission of PERC- and RERT-commands, the histogram of test statistic is shown, even if HIST command is omitted.
 SIM ...The size of simulation in a cycle. In case of omission, the size is five thousands.
 PARA ...Parameters of the random numbers defined by TYPE.

UP ...Output of the upper probability level, or the upper percentile of test statistic.
 LW ...Output of the lower probability level, or the lower percentile of test statistic.
 BO ...Output of either sides of the probability level, or of the both percentiles.
 PERC ...Output of the probability level for the percentile given.
 PERT ...Output of the percentile for the probability level given.

(ii) Command POWER

POWER	<u>TEST</u> (test name)
	<u>POP</u> (number of populations)
	<u>SIZE</u> (sample 1 sample 2...)
	[<u>TYPE</u> (type of R.N.)]
	[<u>ALPHA</u> (probability level)]
	[SIM (size of simulation)]
	[<u>PARA</u> (parameters of R.N.)]
	[SIG (pattern of sig.)]

[Function]

To obtain the power of test given.

[Operands]

TEST ...The name of a test shown in Table A, or the name of a test process defined by DEFTP command.
 SIG ...Only if a test process is applied in TEST operand, SIG operand is needed. The manner is the same as DEFTP command.

The other operands are quite the same as

those in DIST command.

(iii) **Command REF**

REF	<u>TEST</u> (test name 1 test name 2)
	<u>POP</u> (number of populations)
	<u>SIZE</u> (sample 1 sample 2...)
	[<u>TYPE</u> (type of R.N.)]
	[<u>ALPHA</u> (probability level)]
	[<u>SIM</u> (size of simulation)]
	[<u>PARA</u> (parameters of R.N.)]

[Function]

To obtain the relative efficiency of test 2 to test 1 in a certain environment given.

[Operands]

TEST ...The name of tests shown in Table A, or the name of test process(es) defined by DEFTP command.

The other operands are quite the same as those in DIST command.

(iv) **Command AREF**

AREF	<u>TEST</u> (test name 1 test name 2)
	<u>POP</u> (number of populations)
	<u>SIZE</u> (sample 1 and 2...)
	[<u>TYPE</u> (type of R.N.)]
	[<u>ALPHA</u> (probability level)]
	[<u>SIM</u> (size of simulation)]

[Function]

To obtain the asymptotic relative efficiency of test 2 to test 1 in a certain environment given, where test 1 is to be the best test in the Eeden's sense.

[Operands]

Operands involved are quite the same as those in REF command.

(v) **Command DEFTP**

DEFTP	<u>NAME</u> (test process name)
	<u>TEST</u> (test 1...test $n+1$)
	<u>SIG</u> (pattern of sig.)
	[<u>ALPHA</u> (lev. 1...lev. n)]

[Function]

To define a test process (test-series), where individual tests are in Table A.

[Operands]

NAME ...The name of a test process to be defined.

TEST ...The names of tests applied in a test process.

ALPHA ...Significance levels for individual tests, excepting a level for the final test in a test process. The final significance level depends on the succeeding command. In case of omission, all levels must be given at the succeeding command.

SIG ...Significance pattern defined by a test process, where 1, 0 and * mean significance, non-significance and ignorance, respectively. Examples are shown as below.

SIG(111) ...significance in view of a test process, when all individual tests are significant.

SIG(11*) ...significance in view of a test process, where the first and the second tests are significant. The third test does not work anything.

SIG(1*0) ...significance in view of a test process, when the first test is significant, the third test is non-significant.

SIG(1) ...significance in view of a test process, specially when anyone of individual test is significant.

(vi) **Command TPATTN**

TPATTN	<u>TEST</u> (list of test name)
	<u>POP</u> (number of populations)
	<u>SIZE</u> (sample 1 sample 2...)
	[<u>TYPE</u> (type of R.N.)]
	[<u>ALPHA</u> (probability level)]
	[<u>SIM</u> (size of simulation)]
	[<u>PARA</u> (parameters of R.N.)]

[Function]

To obtain frequencies for the respective significance and non-significance patterns due to several tests involved in a test process.

[Operands]

All operands are quite the same as those

of POWER command.

(vii) **Command TESTIM**

TESTIM TEST (test name)
POP (number of populations)
SIZE (sample 1 sample 2...)
TYPE (type of R.N.)
ALPHA (probability level)
SIM (size of simulation)
PARA (parameters of R.N.)
SIG (pattern of sig.)

[Function]

To obtain the mean square error and value of a kind of risk function of estimating the mean of the first population after a preliminary test or a test process given.

[Operands]

TEST ...The name of a test or a test process given. The name of tests involved in the process are shown in Table A, like those in POWER command.

ALPHA ...A probability of significance level. In case of omission, $\alpha = 0.05$.

The other operands are quite the same as those of POWER command.

(viii) **Command CONDIST**

CONDIST TEST (test process name)
POP (number of populations)
SIZE (sample 1 sample 2...)
TYPE (type of R.N.)
ALPHA (list of pro. level)
PAL (probability level)
SIM (size of simulation)
PARA (parameters of R.N.)
HIST

[Function]

To obtain the conditional distribution of test statistic used in the final test at a test-process (test series) defined.

[Operands]

TEST ...The name of a test-process defined by DEFTP command.

ALPHA ...Individual significance levels when they are not given by DEFTP command. See DEFTP command.

PAL ...A significance level desired throughout the test process. Then a significant value of test statistic used in the final test at a test process is shown.

The other operands are the same as those used in POWER command.

Table A. Simple Statistical Tests Involved at Present*

One-Sample Tests	Two-Sample Tests	Multi-Sample Tests
<u>W</u> ilcoxon one sample test	<u>W</u> ilcoxon rank sum test	<u>K</u> Ruskal-wallis test
<u>C</u> HI-square test	<u>A</u> nsari-Bradley test	<u>C</u> HI-square test
<u>K</u> OLmogorov-Smirnov test	<u>L</u> EPage test	<u>F</u> RIEdman test
	<u>C</u> HI-square test	<u>M</u> EDian test
	<u>S</u> IGN test	<u>L</u> OGrank test
	<u>K</u> OLmogorov-Smirnov test	
	<u>D</u> -test (Grove's)	
	<u>G</u> eneralized <u>W</u> ilcoxon test	
	<u>C</u> OX-Mantel test	
	<u>M</u> ANtel-Haenzel test	
	<u>T</u> -test	
	<u>W</u> ELch-aspin test	
	<u>F</u> -test	

(* Tests are still increased.)

Table B. Types of Random Numbers and their Input-Forms

Types (notations)	Input-Forms		
	1st Parameter	2nd Parameter	3rd Parameter
<u>RE</u> ctangular [a, b]	a [0]	b [1]	—
<u>BI</u> nomial B(n, p)	n [10]	p [0.5]	—
<u>PO</u> isson mean a	a [5]	—	—
<u>NO</u> rmal N(a, b)	a [0]	b [1]	—
<u>EX</u> ponential $be^{-b(x-a)}$	a [0]	b [1]	—
<u>WE</u> ibull $(c/b)Z^{c-1}e^{-Z^c}$ $Z \equiv (X-a)/b$	a [0]	b [1]	c [3]

[]: Value in case of omission for key-in.

4.3. A simple illustration of the correspondence of command TESTIM with the mathematical formulation

(i) PROBLEM: The inference of population mean after one-sided t -test (Asano [6]).

Let $O_{n_1}: (x_{11}, x_{12}, \dots, x_{1n_1})$ be a random sample of n_1 from a normal population $N(\mu_1, \sigma^2)$ and let $O_{n_2}: (x_{21}, x_{22}, \dots, x_{2n_2})$ be another random sample of n_2 drawn from another normal population $N(\mu_2, \sigma^2)$. These two normal populations are known to have the same population variance σ^2 , whose value is however unknown. Then, our rule of the statistical procedure is formulated in the following way.

(1) Let \bar{x}_i and s_i^2 be the sample mean and the sample variance defined by O_{n_i} ($i=1, 2$).

(2) Let the statistic t be defined by

$$t = (\bar{x}_1 - \bar{x}_2) / \left\{ \left(\frac{1}{n_1} + \frac{1}{n_2} \right)^{1/2} \left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} \right)^{1/2} \right\}.$$

(3) Let us define the statistic \bar{x} in the following way:

- (a) $\bar{x} = \bar{x}_2$, if $t \leq \lambda$,
- (b) $\bar{x} = (n_1 \bar{x}_1 + n_2 \bar{x}_2) / (n_1 + n_2)$, if $\lambda < t < A$,
- (c) $\bar{x} = \bar{x}_1$, if $t \geq A$,

where the switching constants A and λ are prescribed constants.

THEOREM 4.1. The distribution function of \bar{x} is given by

$$(4.1) \quad \Pr. \{ \bar{x} < u \} = \frac{1}{\sqrt{2\pi}} \int_{T \leq \lambda} \int e^{-p^2/2} \frac{V^{\frac{n_1+n_2}{2}-2} e^{-V}}{\Gamma\left(\frac{n_1+n_2-2}{2}\right)} \Phi\left(\frac{u-\mu_2}{\sigma} \sqrt{n_1+n_2} - \sqrt{\frac{n_1}{n_2}} p\right) dp dV$$

$$+ \frac{1}{\sqrt{2\pi}} \Phi\left(\frac{(n_1+n_2)u - (n_1\mu_1 + n_2\mu_2)}{\sigma \sqrt{n_1+n_2}}\right) \int_{\lambda < T < A} \int e^{-\mu^2/2} \frac{V^{\frac{n_1+n_2}{2}-2} e^{-V}}{\Gamma\left(\frac{n_1+n_2-2}{2}\right)} dp dV$$

$$+ \frac{1}{\sqrt{2\pi}} \int_{T \geq A} \int e^{-p^2/2} \frac{V^{\frac{n_1+n_2}{2}-2} e^{-V}}{\Gamma\left(\frac{n_1+n_2-2}{2}\right)} \Phi\left(\frac{u-\mu_1}{\sigma} \sqrt{n_1+n_2} + \sqrt{\frac{n_2}{n_1}} p\right) dp dV,$$

where

$$T \equiv \left(\sqrt{\frac{n_1 n_2}{n_1+n_2}} \frac{\mu_2 - \mu_1}{\sigma} + p \right) / \sqrt{\frac{2}{n_1+n_2-2} V}.$$

THEOREM 4.2. The mean value $E\{\bar{x}\}$ and the mean square error $M.S.E.\{\bar{x}\}$ of estimate \bar{x} are given by

$$(4.2) \quad E\{\bar{x}\} = \mu_1 - \frac{(\mu_1 - \mu_2)}{n_1 + n_2} \{n_1 Pr.\{D_0\} + n_2(1 - Pr.\{D_2\})\}$$

$$+ \frac{\sigma}{\sqrt{n_1 + n_2}} \cdot \left\{ \sqrt{\frac{n_1}{n_2}} I_{D_0}(p) - \sqrt{\frac{n_2}{n_1}} I_{D_2}(p) \right\},$$

$$(4.3) \quad M.S.E.\{\bar{x}\} = \frac{\sigma^2}{n_1 + n_2} + \frac{\sigma^2}{n_1 + n_2} \left\{ \frac{n_2}{n_1} I_{D_2}(p^2) + \frac{n_1}{n_2} I_{D_0}(p^2) \right\} + \{1 - Pr.\{D_2\}\} \\ \cdot \left(\frac{n_2(\mu_1 - \mu_2)}{n_1 + n_2} \right)^2 + Pr.\{D_0\} \cdot \left\{ \mu_1 + \frac{n_2\mu_1 + n_1\mu_2}{n_1 + n_2} \right\} \left\{ \mu_2 + \frac{n_1\mu_1 + n_2\mu_2}{n_1 + n_2} \right\},$$

where

$$I_{D_i}(p^j) = \iint_{D_i} p^j f(p, V) dp dV, \quad (j=1, 2),$$

$$f(p, V) = \frac{1}{\sqrt{2\pi}} e^{-p^2/2} \frac{V^{\frac{n_1+n_2}{2}-2} e^{-V}}{\Gamma\left(\frac{n_1+n_2-2}{2}\right)},$$

$$D_i : \begin{cases} T \leq \lambda, & (i=0), \\ T \geq A, & (i=2), \end{cases}$$

$$Pr.\{D_i\} = \iint_{D_i} f(p, V) dp dV.$$

(ii) **SOLUTION:** For a practical situation, when the estimate of μ_1 is given by a result of preliminary test of significance between two population means, i.e. $\bar{x} = \bar{x}_1$ for the significance and $\bar{x} = (n_1 \bar{x}_1 + n_2 \bar{x}_2) / (n_1 + n_2)$ for the nonsignificance, where $\lambda = -\infty$ in the previous theorems, the properties of such a procedure are evaluated in the following way.

```

NISAN
TESTIM TEST(T) POP(2) SIZE(5 15) TYPE(NORMAL) -
ALPHA(5.0) SIM(5000) PARA((0.0 0.0 1.0 1.0) -
(0.0 -0.05 1.0 1.0) (0.0 -0.1 1.0 1.0) -
(0.0 -0.50 1.0 1.0) (0.0 -1.0 1.0 1.0) -
(0.0 -2.00 1.0 1.0) (0.0 -5.0 1.0 1.0)) AVE UP

```

TESTIMATION FOR MEAN1

TEST NAME	STUDENT T
NUMBER OF POPULATIONS	2
TYPE OF RANDOM NUMBER	NORMAL
SIGNIFICANCE LEVEL	0.050
NUMBER OF SIMULATION	5000

NO. 1	SAMPLE1	5	SAMPLE2	15
1)	PARA.(1)	0.00	PARA.(1)	0.00
	PARA.(2)	1.00	PARA.(2)	1.00
	MEAN			0.039
	VARIANCE			0.075
	M.S.E.			0.077
	RELATIVE RISK			0.399
2)	PARA.(1)	0.00	PARA.(1)	-0.05
	PARA.(2)	1.00	PARA.(2)	1.00
	MEAN			0.007
	VARIANCE			0.089
	M.S.E.			0.089
	RELATIVE RISK			0.438
3)	PARA.(1)	0.00	PARA.(1)	-0.10
	PARA.(2)	1.00	PARA.(2)	1.00
	MEAN			-0.015
	VARIANCE			0.096
	M.S.E.			0.097
	RELATIVE RISK			0.476
4)	PARA.(1)	0.00	PARA.(1)	-0.50
	PARA.(2)	1.00	PARA.(2)	1.00
	MEAN			-0.157
	VARIANCE			0.202
	M.S.E.			0.227
	RELATIVE RISK			1.110
5)	PARA.(1)	0.00	PARA.(1)	-1.00
	PARA.(2)	1.00	PARA.(2)	1.00
	MEAN			-0.171
	VARIANCE			0.327
	M.S.E.			0.356
	RELATIVE RISK			1.761
6)	PARA.(1)	0.00	PARA.(1)	-2.00
	PARA.(2)	1.00	PARA.(2)	1.00
	MEAN			-0.011
	VARIANCE			0.226
	M.S.E.			0.226
	RELATIVE RISK			1.140
7)	PARA.(1)	0.00	PARA.(1)	-5.00
	PARA.(2)	1.00	PARA.(2)	1.00
	MEAN			-0.002
	VARIANCE			0.193
	M.S.E.			0.193
	RELATIVE RISK			1.000

NISAN

4.4. The other illustrations of outputs

(i) COMMAND DIST (jointly checking the precision)

```

NISAN
DIST TEST(T) POP(2) SIZE(20 20) TYPE(NORMAL) -
SIM(5000) PARA(0.0 0.0 1.0 1.0) PERT(1.0 5.0 -
10.0 20.0) BO HIST

TEST NAME                      STUDENT T
NUMBER OF POPULATIONS          2
TYPE OF RANDOM NUMBER          NORMAL
NUMBER OF SIMULATION           5000

NO. 1  SAMPLE1      20      SAMPLE2      20
      1) PARA.(1)    0.00    PARA.(1)    0.00
      PARA.(2)    1.00    PARA.(2)    1.00
                                     Theoretical
NO.    SIG. LEVEL(TWO-TAILED)    PERCENT POINT    values
1      0.010                     2.626           2.712
2      0.050                     2.005           2.024
3      0.100                     1.682           1.656
4      0.200                     1.310           1.304

THE DISTRIBUTION AND PROBABILITIES OF T -TEST STATISTICS

VALUES OF      55 110 165 220 275 330 385 440 495 550 605 660
TEST-STAT FREQ. +-----+-----+-----+-----+-----+-----+
INT. 1      0 I
INT. 2      0 I
INT. 3      2 I*
INT. 4      0 I
INT. 5      1 I*
INT. 6      1 I*
INT. 7      5 I*
INT. 8     14 I**
INT. 9     33 I***
INT.10     69 I*****
INT.11    115 I*****
INT.12    192 I*****
INT.13    270 I*****
INT.14    360 I*****
INT.15    473 I*****
INT.16    548 I*****
INT.17    616 I*****
INT.18    581 I*****
INT.19    504 I*****
INT.20    433 I*****
INT.21    287 I*****
INT.22    211 I*****
INT.23    132 I*****
INT.24     76 I*****
INT.25     40 I****
INT.26     21 I**
INT.27      8 I*
INT.28      5 I*
INT.29      3 I*
INT.30      0 I
+-----+-----+-----+-----+-----+-----+
WHERE
INT. 1 = (-5.00 -4.70), CUM.FREQ. 0 CUM.PROB. 0.0 PERCENT
INT. 2 = (-4.70+ -4.40), CUM.FREQ. 0 CUM.PROB. 0.0 PERCENT
INT. 3 = (-4.40+ -4.10), CUM.FREQ. 2 CUM.PROB. 0.04 PERCENT
INT. 4 = (-4.10+ -3.80), CUM.FREQ. 2 CUM.PROB. 0.04 PERCENT
INT. 5 = (-3.80+ -3.50), CUM.FREQ. 3 CUM.PROB. 0.06 PERCENT
INT. 6 = (-3.50+ -3.20), CUM.FREQ. 4 CUM.PROB. 0.08 PERCENT
INT. 7 = (-3.20+ -2.90), CUM.FREQ. 9 CUM.PROB. 0.18 PERCENT
INT. 8 = (-2.90+ -2.60), CUM.FREQ. 23 CUM.PROB. 0.46 PERCENT
INT. 9 = (-2.60+ -2.30), CUM.FREQ. 56 CUM.PROB. 1.12 PERCENT
INT.10 = (-2.30+ -2.00), CUM.FREQ. 125 CUM.PROB. 2.50 PERCENT
INT.11 = (-2.00+ -1.70), CUM.FREQ. 240 CUM.PROB. 4.80 PERCENT
INT.12 = (-1.70+ -1.40), CUM.FREQ. 432 CUM.PROB. 8.64 PERCENT
INT.13 = (-1.40+ -1.10), CUM.FREQ. 702 CUM.PROB. 14.04 PERCENT
INT.14 = (-1.10+ -0.80), CUM.FREQ. 1062 CUM.PROB. 21.24 PERCENT
INT.15 = (-0.80+ -0.50), CUM.FREQ. 1535 CUM.PROB. 30.70 PERCENT
INT.16 = (-0.50+ -0.20), CUM.FREQ. 2083 CUM.PROB. 41.66 PERCENT
INT.17 = (-0.20+ 0.10), CUM.FREQ. 2699 CUM.PROB. 53.98 PERCENT
INT.18 = (0.10+ 0.40), CUM.FREQ. 3280 CUM.PROB. 65.60 PERCENT
INT.19 = (0.40+ 0.70), CUM.FREQ. 3784 CUM.PROB. 75.68 PERCENT
INT.20 = (0.70+ 1.00), CUM.FREQ. 4217 CUM.PROB. 84.34 PERCENT
INT.21 = (1.00+ 1.30), CUM.FREQ. 4504 CUM.PROB. 90.08 PERCENT
INT.22 = (1.30+ 1.60), CUM.FREQ. 4715 CUM.PROB. 94.30 PERCENT
INT.23 = (1.60+ 1.90), CUM.FREQ. 4847 CUM.PROB. 96.94 PERCENT
INT.24 = (1.90+ 2.20), CUM.FREQ. 4923 CUM.PROB. 98.46 PERCENT
INT.25 = (2.20+ 2.50), CUM.FREQ. 4963 CUM.PROB. 99.26 PERCENT
INT.26 = (2.50+ 2.80), CUM.FREQ. 4984 CUM.PROB. 99.68 PERCENT
INT.27 = (2.80+ 3.10), CUM.FREQ. 4992 CUM.PROB. 99.84 PERCENT
INT.28 = (3.10+ 3.40), CUM.FREQ. 4997 CUM.PROB. 99.94 PERCENT
INT.29 = (3.40+ 3.70), CUM.FREQ. 5000 CUM.PROB. 100.00 PERCENT
INT.30 = (3.70+ 4.00), CUM.FREQ. 5000 CUM.PROB. 100.00 PERCENT

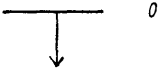
```

NISAN

(iii) COMMAND POWER (for a process of testing)

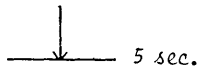
```
NISAN
POWER TEST(W T) POP(2) SIZE(20 20) TYPE(NORMAL) -
ALPHA(5.0 5.0) SIM(5000) PARA((0.0 0.0 1.0 1.0) -
(0.0 0.2 1.0 1.0) (0.0 0.4 1.0 1.0) -
(0.0 0.6 1.0 1.0) (0.0 0.8 1.0 1.0) -
(0.0 1.0 1.0 1.0)) SIG(1)
```

(required
time)



POWER OF TEST STATISTICS

TEST NAME(1)				WILCOXON RANK SUM	
TEST NAME(2)				STUDENT T	
NUMBER OF POPULATINS				2	
TYPE OF RANDOM NUMBER				NORMAL	
SIGNIFICANCE LEVEL(1)				0.050	
SIGNIFICANCE LEVEL(2)				0.050	
NUMBER OF SIMULATION				5000	
TEST(1)				(SIG.)	
TEST(2)				(SIG.)	
NO. 1	SAMPLE1	20	SAMPLE2	20	
1)	PARA.(1)	0.00	PARA.(1)	0.00	
	PARA.(2)	1.00	PARA.(2)	1.00	
	POWER				0.062
2)	PARA.(1)	0.00	PARA.(1)	0.20	
	PARA.(2)	1.00	PARA.(2)	1.00	
	POWER				0.104
3)	PARA.(1)	0.00	PARA.(1)	0.40	
	PARA.(2)	1.00	PARA.(2)	1.00	
	POWER				0.255
4)	PARA.(1)	0.00	PARA.(1)	0.60	
	PARA.(2)	1.00	PARA.(2)	1.00	
	POWER				0.482
5)	PARA.(1)	0.00	PARA.(1)	0.80	
	PARA.(2)	1.00	PARA.(2)	1.00	
	POWER				0.719
6)	PARA.(1)	0.00	PARA.(1)	1.00	
	PARA.(2)	1.00	PARA.(2)	1.00	
	POWER				0.884



NISAN

(iv) COMMAND TPATTN

NISAN
 TPATTN TEST(W ANS T) POP(2) SIZE(20 20) TYPE(NORMAL) -
 ALPHA(5.0 5.0 5.0) SIM(3000) PARA(0.0 0.0 1.0 1.0)

SIGNIFICANCE TABLE OF TEST

TEST NAME1	WILCOXON RANK SUM
TEST NAME2	ANSARI BRADLEY
TEST NAME3	STUDENT T
NUMBER OF POPULATIONS	2
TYPE OF RANDOM NUMBER	NORMAL
SIGNIFICANCE LEVEL1	0.050
SIGNIFICANCE LEVEL2	0.050
SIGNIFICANCE LEVEL3	0.050
NUMBER OF SIMULATION	3000

NO. 1	SAMPLE1	20	SAMPLE2	20
1)	PARA.(1)	0.00	PARA.(1)	0.00
	PARA.(2)	1.00	PARA.(2)	1.00

FREQ.	WIL	ANS	T
2701	0	0	0
34	0	0	1
128	0	1	0
1	0	1	1
22	1	0	0
112	1	0	1
1	1	1	0
1	1	1	1
3000			

WHERE 0:NON-SIG., 1:SIG.

NISAN

(v) COMMAND REF

NISAN
 REF TEST(T W) POP(2) SIZE(20 20) TYPE(NORMAL) -
 ALPHA(5.0) SIM(5000) PARA((0.0 0.2 1.0 1.0) -
 (0.0 0.4 1.0 1.0) (0.0 0.6 1.0 1.0) -
 (0.0 0.8 1.0 1.0) (0.0 1.0 1.0 1.0))

TEST1 NAME	STUDENT T
TEST2 NAME	WILCOXON RANK SUM
NUMBER OF POPULATIONS	2
TYPE OF RANDOM NUMBER	NORMAL
SIGNIFICANCE LEVEL	0.050
NUMBER OF SIMULATION	5000

RELATIVE EFFICIENCY OF TEST2 TO TEST1

NO. 1	SAMPLE1	20	SAMPLE2	20
1)	PARA.(1)	0.00	PARA.(1)	0.20
	PARA.(2)	1.00	PARA.(2)	1.00
	EFFICIENCY			0.860
2)	PARA.(1)	0.00	PARA.(1)	0.40
	PARA.(2)	1.00	PARA.(2)	1.00
	EFFICIENCY			0.956
3)	PARA.(1)	0.00	PARA.(1)	0.60
	PARA.(2)	1.00	PARA.(2)	1.00
	EFFICIENCY			0.972
4)	PARA.(1)	0.00	PARA.(1)	0.80
	PARA.(2)	1.00	PARA.(2)	1.00
	EFFICIENCY			0.935
5)	PARA.(1)	0.00	PARA.(1)	1.00
	PARA.(2)	1.00	PARA.(2)	1.00
	EFFICIENCY			0.956

NISAN

(vi) COMMANDS DEFTP AND TESTIM

```

NISAN
DEFTP NAME(WILT) TEST(W T) SIG(1)
NISAN
TESTIM TEST(WILT) POP(2) SIZE(5 15) TYPE(NORMAL) -
ALPHA(5.0 5.0) SIM(5000) PARA((0.0 0.0 1.0 1.0) -
(0.0 -0.1 1.0 1.0) (0.0 -0.5 1.0 1.0) -
(0.0 -1.0 1.0 1.0) (0.0 -2.0 1.0 1.0)) AVE UP

TESTIMATION FOR MEAN1

TEST NAME                                WILT
TEST(1)                                WILCOXON RANK SUM
TEST(2)                                STUDENT T
SIGNIFICANCE PATTERN
TEST(1)                                (SIG.)
TEST(2)                                (SIG.)
NUMBER OF POPULATIONS                    2
TYPE OF RANDOM NUMBER                  NORMAL
SIGNIFICANCE LEVEL(1)                  0.050
SIGNIFICANCE LEVEL(2)                  0.050
NUMBER OF SIMULATION                    5000

NO. 1  SAMPLE1      5      SAMPLE2      15
1)  PARA.(1)      0.00    PARA.(1)      0.00
    PARA.(2)      1.00    PARA.(2)      1.00

    MEAN                                0.020
    VARIANCE                            0.094
    M.S.E.                              0.094
    RELATIVE RISK                       0.489

2)  PARA.(1)      0.00    PARA.(1)     -0.10
    PARA.(2)      1.00    PARA.(2)      1.00

    MEAN                                -0.030
    VARIANCE                            0.109
    M.S.E.                              0.109
    RELATIVE RISK                       0.536

3)  PARA.(1)      0.00    PARA.(1)     -0.50
    PARA.(2)      1.00    PARA.(2)      1.00

    MEAN                                -0.172
    VARIANCE                            0.197
    M.S.E.                              0.227
    RELATIVE RISK                       1.118

4)  PARA.(1)      0.00    PARA.(1)     -1.00
    PARA.(2)      1.00    PARA.(2)      1.00

    MEAN                                -0.151
    VARIANCE                            0.339
    M.S.E.                              0.362
    RELATIVE RISK                       1.769

5)  PARA.(1)      0.00    PARA.(1)     -2.00
    PARA.(2)      1.00    PARA.(2)      1.00

    MEAN                                -0.020
    VARIANCE                            0.226
    M.S.E.                              0.227
    RELATIVE RISK                       1.121

NISAN

```

```
NISAN
CONDIST TEST(WILT) POP(2) SIZE(20 20) TYPE(NORMAL) -
ALPHA(5.0) PAL(10.0) SIM(5000) PARA(0.0 0.0 1.0 1.0) -
HIST
```

TEST NAME			WILT	
TEST(1)			WILCOXON RANK SUM	
TEST(2)			STUDENT T	
SIGNIFICANCE PATTERN				
TEST(1)			(S.I.G.)	
TYPE OF RANDOM NUMBER			NORMAL	
SIGNIFICANCE LEVEL(1)			0.050	
NUMBER OF SIMULATION			5000	
NO. 1	SAMPLE1	20	SAMPLE2	20
1)	PARA.(1)	0.00	PARA.(1)	0.00
	PARA.(2)	1.00	PARA.(2)	1.00

NO.	SIG. LEVEL(TWO-TAILED)	PERCENT POINT
1	0.100	1.733

[illegible]

INT.	1 =	-3.00	-2.80	CUM.FREQ.	0	CUM.PROB.	0.0	PERCENT
INT. 2	=	-2.80+	-2.60	CUM.FREQ.	0	CUM.PROB.	0.0	PERCENT
INT. 3	=	-2.60+	-2.40	CUM.FREQ.	1	CUM.PROB.	0.02	PERCENT
INT. 4	=	-2.40+	-2.20	CUM.FREQ.	8	CUM.PROB.	0.17	PERCENT
INT. 5	=	-2.20+	-2.00	CUM.FREQ.	29	CUM.PROB.	0.61	PERCENT
INT. 6	=	-2.00+	-1.80	CUM.FREQ.	84	CUM.PROB.	1.77	PERCENT
INT. 7	=	-1.80+	-1.60	CUM.FREQ.	159	CUM.PROB.	3.35	PERCENT
INT. 8	=	-1.60+	-1.40	CUM.FREQ.	297	CUM.PROB.	6.26	PERCENT
INT. 9	=	-1.40+	-1.20	CUM.FREQ.	476	CUM.PROB.	10.03	PERCENT
INT.10	=	-1.20+	-1.00	CUM.FREQ.	692	CUM.PROB.	14.58	PERCENT
INT.11	=	-1.00+	-0.80	CUM.FREQ.	949	CUM.PROB.	20.00	PERCENT
INT.12	=	-0.80+	-0.60	CUM.FREQ.	1216	CUM.PROB.	25.63	PERCENT
INT.13	=	-0.60+	-0.40	CUM.FREQ.	1551	CUM.PROB.	32.69	PERCENT
INT.14	=	-0.40+	-0.20	CUM.FREQ.	1930	CUM.PROB.	40.67	PERCENT
INT.15	=	-0.20+	-0.00	CUM.FREQ.	2340	CUM.PROB.	49.32	PERCENT
INT.16	=	-0.00+	0.20	CUM.FREQ.	2754	CUM.PROB.	58.04	PERCENT
INT.17	=	0.20+	0.40	CUM.FREQ.	3136	CUM.PROB.	66.09	PERCENT
INT.18	=	0.40+	0.60	CUM.FREQ.	3489	CUM.PROB.	73.53	PERCENT
INT.19	=	0.60+	0.80	CUM.FREQ.	3795	CUM.PROB.	79.98	PERCENT
INT.20	=	0.80+	1.00	CUM.FREQ.	4053	CUM.PROB.	85.42	PERCENT
INT.21	=	1.00+	1.20	CUM.FREQ.	4271	CUM.PROB.	90.01	PERCENT
INT.22	=	1.20+	1.40	CUM.FREQ.	4430	CUM.PROB.	93.36	PERCENT
INT.23	=	1.40+	1.60	CUM.FREQ.	4555	CUM.PROB.	96.00	PERCENT
INT.24	=	1.60+	1.80	CUM.FREQ.	4637	CUM.PROB.	97.72	PERCENT
INT.25	=	1.80+	2.00	CUM.FREQ.	4711	CUM.PROB.	99.28	PERCENT
INT.26	=	2.00+	2.20	CUM.FREQ.	4738	CUM.PROB.	99.85	PERCENT
INT.27	=	2.20+	2.40	CUM.FREQ.	4745	CUM.PROB.	100.00	PERCENT
INT.28	=	2.40+	2.60	CUM.FREQ.	4745	CUM.PROB.	100.00	PERCENT
INT.29	=	2.60+	2.80	CUM.FREQ.	4745	CUM.PROB.	100.00	PERCENT
INT.30	=	2.80+	3.00	CUM.FREQ.	4745	CUM.PROB.	100.00	PERCENT

5. Conclusion

A new development of NISAN system is presented mainly in view of the functions of NISAN monitor, fully written by FORTRAN, and some of knowledge-commands newly-introduced, with ordinary method-commands sampled. The functions of monitor is for the benefit not only of the users, but also of the cooperation of NISAN research members. The functions of knowledge-commands are emphasized to be effective on statistical study and data analysis in practice for the senior statistician and the beginner, because of reasonable evaluation and preventing misapplication of statistical methods.

NISAN system is still developing into fullness on the ordinary usefulness and with the originality of statistical research.

References

- [1] ASANO, CH., WAKIMOTO, K., SHOHOJI, T. ET AL.: *The statistical principle and methodology in NISAN system*, Proceedings in Computational Statistics, Physica-Verlag (1978).
- [2] GOTO, M., UESAKA, H. AND ASANO, CH.: *Methodology of data investigation in NISAN system; multivariate normality tests*, Res. Rep. No. 90, Res. Instit. Fund. Inform. Sc., Kyushu University (1978).
- [3] ASANO, CH., TANAKA, K., JOJIMA, K., NISHIGORI, N.: *Command Construction Manual for NISAN Research Members*, Res. Rep. No. 101, Res. Instit. Fund. Inform. Sc., Kyushu University, (1981).
- [4] ASANO, CH., SHOHOJI, T., WAKIMOTO, K. ET AL.: *System construction of new interactive statistical analysis (NISAN)*, Proceeding of Intern. Statist. Instit. (1979).
- [5] ASANO, CH., WATANABE, M. AND HOSOKAWA, K.: *NISAN System Users' Guide*, Res. Instit. Fund. Inform. Sc. (1981).
- [6] ASANO, CH.: *Estimation after preliminary test of significance and their applications to biometrical researches*, Bull. Math. Stat., **9**, (1960), 1-23.
- [7] KITAGAWA, T.: *Successive process of statistical inferences (5)*, Mem. Fac. of Sc., Kyushu University, Ser. A, **7** (1953), 81-106.
- [8] BANCROFT, T.A. AND HAN, C.P.: *Inference based on conditional specification: A note and a bibliography.*, International Statistical Review, **45** (1977) 117-127.